Have a SEAT on Stage: Restoring Trust with Spectator Experience Augmentation Techniques

Olivier Capra, Florent Berthaut, Laurent Grisoni

Have a SEAT on Stage : Restoring Trust with Spectator Experience Augmentation Techniques

Olivier Capra
CRIStAL, CNRS,
University of Lille, France
olivier.capra@univ-lille.fr

Florent Berthaut
CRIStAL, CNRS,
University of Lille, France
florent.berthaut@univ-lille.fr

Laurent Grisoni
CRIStAL, CNRS,
University of Lille, France
laurent.grisoni@univ-lille.fr

ABSTRACT

When the collaboration between humans and machines happens in public, the audience can face difficulties in distinguishing the actual human contribution from the contribution of autonomous processes. In music concerts involving digital interfaces doubts about the performer's contribution can drastically hinder the audience interest. The disappearing of the direct physical link between actions and effects is one of the reasons of this confusion. Consequently both artists and researchers have explored techniques to augment the experience of spectators. However their respective impact on the multiple aspects of audience experience has not yet been formally compared. In this controlled study, we compare two techniques : pre-performance explanations and visual augmentations. Despite contradictory results on comprehension tasks, we show that contrary to pre-performance explanations, visual augmentations improve the audience experience, increase their subjective comprehension and restore the trust in performers by reversing the doubt in their favour.

Author Keywords

interactions; spectator experience; visual augmentations ; design for public interactions; human machine collaboration;

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); Laboratory experiments; *Mixed / augmented reality;* **•Applied computing → Sound and music computing;**

INTRODUCTION

When the collaboration between humans and machines happens in public, the audience can face difficulties in distinguishing the actual human contribution from the contribution of autonomous processes. In artistic performances involving digital interfaces, and especially in music concerts, doubts about the performer's involvement can drastically hinder the audience interest [54, 56]. In the HCI community, the question of the audience experience during such performances is growing in interest as evidenced by the extensive related literature [49, 48, 6]. It has also grown in importance in specific domains such as musical interaction, where it is in particular used as an evaluation and design methodology [2, 40, 29]. Among the issues pointed out in this literature, two can be highlighted : the loss of the physical link between the gesture and the sound production and the doubt about the performer's contribution.

DIS2020, July 06–10, 2020, Eindhoven, NL

ISBN 978-1-4503-6974-9/20/07...$15.00

DOI: 10.1145/3357236.3395492

First, the continuous technological improvements in Digital Musical Instruments (DMIs) dissolved the physical constraints in the production of sound, offering musicians almost infinite creative possibilities. With the multiplication of these software instruments, musicians became their own instrument makers, amplifying the diversity of interfaces and gesture to sound mappings leading to better intimacy or expressivity for the musicians [36, 62]. Furthermore, the arrival of computers on stage relieved the artists of the physical requirements, allowing for expressive gestures with less amplitude, if not hidden [30, 14, 10]. From an audience perspective, these mutations of the intuitive link between actions and effects [48] have multiple consequences. A disturbed perception can lower the transmission of emotions [24], disrupt the anticipation and its related pleasure [32] and therefore hinder the experience of live music. Thus, less constrained by the laws of physics, musicians face new challenges to better expose their skills and intentions when on stage.

A second issue lies in the very interest of attending a live artistic performance : the singularity of seeing and hearing artists expressing themselves doing things we are not able to do. If a performance with digital devices gives the impression that everything could flow the same way whether or not the performer is involved, the audience may doubt their actual contribution. The risk is to make the performance as interesting as watching people "doing their taxes" (David Zicarelli, founder of Cycling '74, a major music software company, reported in [54]). Beyond the simple interest, the experience of flow, that is the total immersion in the performance [23], also requires to perceive the artistic demand of the performance and the skills of the artist [23, 38, 50].

Artists and researchers alike have explored solutions to the degradation of the audience experience. In particular, they have proposed techniques to augment performances without constraining the design of performance interfaces. These techniques range from pre-performance demos [9] to augmented-reality displays [8] and provide various types of information about the instrument, its mechanisms and the ongoing interactions of the musician. While they constitute a promising solution to the issue of audience experience with digital performances, these Spectator Experience Augmentation Techniques (SEATs) have not yet been formally compared. Therefore, their respective effect on the many aspects of the spectators experience remains unclear.

In this paper, we compare two of these SEATs, i.e. pre-performance demo and visual augmentations, through a controlled study, in the context of musical performances. We investigate their effect on multiple aspects of the audience experience, including subjective and objective comprehension, especially looking at the perception of the performer's contribution. From the results, we derive insights which can inform both the design of future techniques and the evaluation of the audience experience with digital performances.

Related work

Spectator experience augmentation techniques (SEATs)

A first strategy for improving the audience experience with digital performances is to design the performers interfaces, e.g. musical instruments, to maximise their propensity to make the observers intuitively understand their behaviour and potentialities, aka the transparency [27]. However, this may require the performers to restrain their expressivity with respect to all potential combinations of gestures, sensors, mappings and content. These restrictions can be lifted by audiovisual interfaces designed to expose the instrument capabilities and facilitate the comprehension of both musicians and audience. This approach shows interesting benefits in the case of DMIs [35, 21] but it adds constraints to instrument design outside the sole musician's expressive needs. A second strategy explored by both artists and researchers is to augment the performance with information that complements the audience experience, without modifying the performer's interface. This augmentation can be done for various goals regarding the audience experience, such as increasing the audience enjoyment, engagement or comprehension, e.g. turning *magical* interfaces into *expressive* ones [48]. It may also take many forms.

Visual feedback can be projected behind the performers [47], overlapped with the instrument using an AR display [8], or provided both on stage and on mobile devices [6]. In a precedent work, we proposed a pipeline based on the integration of physiological data to select relevant visual augmentation [15]. Haptic feedback can also be used to amplify musicians' gestures [1, 58]. Other approaches involve pre-concert explanations of the instrument [9, 29] or of the performer's intention [29]. Finally, common techniques developed by artists include tilting their interface towards the audience and projecting a close-up of their gestures or their graphical user interface.

Although these techniques may definitely benefit the spectators of digital performances, to our knowledge, their efficiency has not been formally compared and no insights have been provided about the components of the audience experience that they respectively affect.

Measuring the audience experience

In the study of audience experience, most research has relied on an "in the wild" approach [17, 5], utilising observations, post-performance questionnaires and guided interviews of spectators. Different approaches are proposed in terms of methodology (how they measure) and of the dimensions they address (what they measure).

Most contributions rely on pre/post-performance interviews and questionnaires to address components of the audience experience such as enjoyment [6, 11], comprehension [6], error perception [11], emotional and social aspects [12] and also to gather improvement suggestions [6]. Additionally, real time measurements can be used to provide feedback during the performance [11].

Similarly, augmentation techniques have mostly been evaluated "in the wild" by organising performances to study the impact of pre-concert explanations on error perception and enjoyment [9] or to better understand how different visual augmentations affect audience understanding [47, 20]. While this method in the field provides valuable information on some aspects of the audience experience, it also contains biases, such as the inherent differences between two runs of the same performance or the effect of the social context, which makes it difficult to precisely compare two techniques.

In contrast, few "in the lab" experiments have been conducted on audience experience. They all rely on videos of performances to address components of the experience. Questionnaires can be used to investigate the perceived causality and the attributed agency (the spectator's appreciation of the level of control of the musician) [7], to evaluate the error perception [28] or to rate the perceived tension [3]. Beyond such subjective reports, physiological measurement can provide objective insights like the level of engagement of spectators watching dance performances [41]. The integration of physiological sensors is a promising way to gather objective data about the audience experience [60, 15, 55]. However, the time resolution, the many sources of interference and the variety of interpretation of physiological data are still obstacles to robust applications.

Lab experiments are by definition far away from the concert hall. Thus the expression "audience experience" may be a bit misleading, especially when the experiment involves one participant at a time. The full experience of standing with dozens of people around, participating in the same event in synchronisation, is an important part of the experience that is de facto neglected in lab studies. We believe that these rich aspects are more grounded in the field of sociological studies, the interested reader is invited to read the related literature [46, 26, 12] on the subject.

Following these observations, we believe that the formal comparison of the effects of augmentation techniques on the audience experience can benefit from a controlled approach. We advocate for an "in-the-lab" methodology as it allows for a balanced evaluation of experimental conditions, precise measurements of the comprehension dimension and consequent refined analyses.

Contribution

In this work, we evaluate the impact of two augmentation techniques, preliminary explanations and visual augmenta-

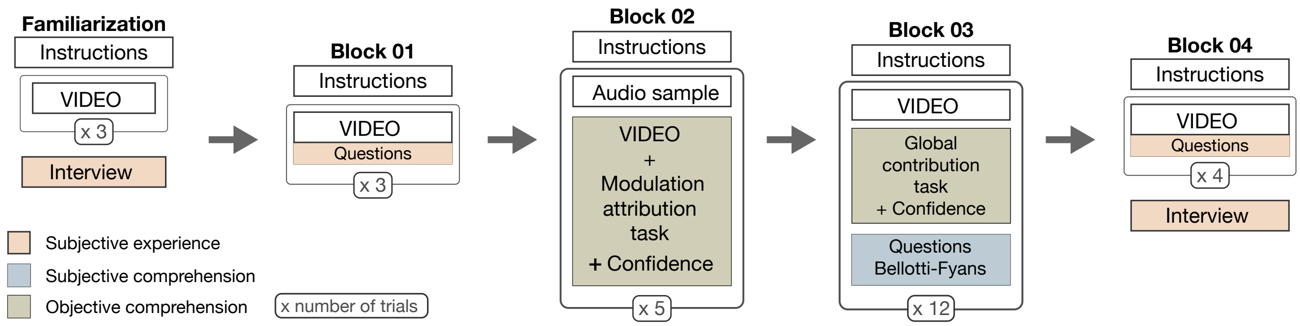

Figure 1. After a familiarisation phase, participants followed 4 successive blocks targeting the components of the experience : subjective experience in Blocks 01 and 04, objective comprehension in Blocks 02 and 03 and subjective comprehension in Block 03.

tions, on the experience of spectators watching a performance with a Digital Musical Instrument (DMI). Our approach draws from field studies and addresses the effects of what we define as Spectator Experience Augmentation Techniques (SEATs) on multiple aspects of audience experience, from subjective ratings to objective comprehension tasks. From the results, we provide insights on the design of SEATs and their effect on experience and comprehension. We also propose a new design challenge which can be used to evaluate the spectators experience.

STUDY : COMPARISON OF TWO AUGMENTATION TECHNIQUES FOR MUSIC PERFORMANCES

In this section, we describe the protocol used in this study. The video stimuli and the experimental conditions are presented, as well as the components of the experience we address. For the sake of clarity, the presentation of the tasks is structured by experience components rather than the chronological order. Following the general presentation of the protocol, methodological details of each component and the associated results are presented and briefly discussed one by one.

Overview

In this study we evaluate two Spectator Experience Augmentation Techniques (SEATs) : preliminary explanations of the instrument and visual augmentations. For the participants, it mostly consists in watching videos and answering subsequent questions. The experiment lasts one hour including consents signature, equipment setup and interviews. As shown on Figure 1, the experiment is composed of a familiarisation phase and four experimental blocks. Throughout the blocks, objective and subjective aspects of the spectator experience are addressed.

Hypotheses

Following the work of Berthaut [8, 7] on visual augmentations, and the conclusions of Bin [9] that even if explanations can familiarise the audience with a new digital instrument, they may not suffice to increase their experience, we hypothesised a greater impact of the visual augmentations on the spectator subjective experience and their objective comprehension. We also hypothesised that this impact is due to the provided information, not only to the increased visual richness of graphical material inclusion.

Decomposing the experience

To handle the elusive concept of experience, we split it into three measurable features, the *objective comprehension*, the *subjective experience* and *the subjective comprehension*. Each aspect is evaluated through one or more tasks and distributed across blocks of our protocol. This methodology maintains the participants attention with short and diversified tasks, and by addressing the experience under multiple aspects prevents from too narrow results [13].

Subjective experience

Experience is by definition a subjective phenomenon and is therefore difficult to assess in a quantitative manner. Moreover, qualitative user feedback represent useful insights that should be taken into consideration. In order to keep this information and allow for comparative analysis, the protocol addresses the subjective experience in 2 ways :

- A guided rating task where the participants are asked to independently rate constituents of the performance : the music, the virtuosity of the musician, the visual aspects and their overall appreciation of the performance. This task is tested twice and constitutes Block 01 and Block 04 (Figure 1). We placed the task at the beginning and at the end to let the time to the participants to integrate the SEATs' behaviour and control for a potentially too flat learning curve.
- Semi-structured interviews allow for more personal and detailed feedback and are crucial to fully understand what the participant experienced. For this purpose, with the consent of the participant, the experiment is fully audio taped. The interviews occurred after the familiarisation phase and at the end of the experiment (Figure 1).

Objective comprehension

The objective comprehension refers to the ability of the spectator to integrate factual elements of the interaction, like who from the musician or the machine is the author of a sound modulation, addressed in Block 02, or who is more contributing in a performance, addressed in Block 03 (Figure 1).

Subjective comprehension

Midway between subjective experience and objective comprehension, subjective comprehension relates to the inner feeling of being able to perceive an aspect of the interaction. This aspect of the experience is tested in Block 03 (Figure 1).

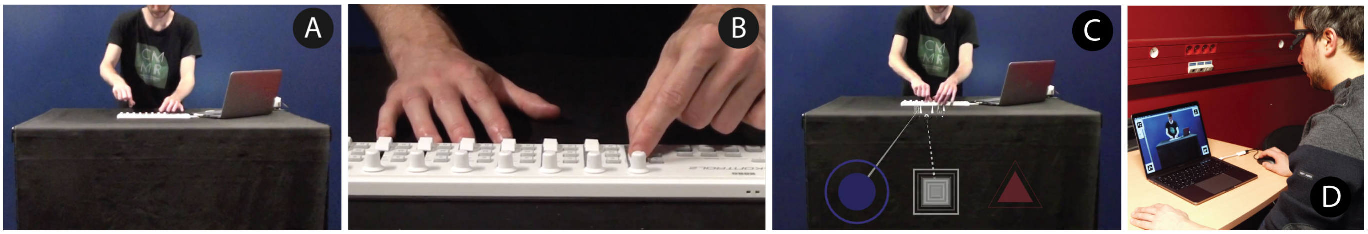

Figure 2. Conditions and *Spectator Experience Augmentation Techniques* evaluated in our study : A) *Control*, B) *Explain* (pre-performance explanations), C) *V-AUG* (visual augmentations), D) The experimental setup

Experimental conditions

To study the impact of the two SEATs, we presented our stimuli under four experimental conditions[1] :

Control condition

In the control condition, the performance is showed as is, without any cues or explanation. (See Figure 2-A)

Preliminary explanations (Explain)

In the *Explain* condition, a short video demo (65 seconds) is played before the performance. This video is a close-up on the musician hands and the instrument that will be used in the subsequent performance. The musician demonstrates each mapping, from sensors (button, slider, knob) to sound parameters, with verbal explanations (See Figure 2-B). The participant can hear each sound processes played individually as well as their modulations allowed by the instrument. The performance is played right after the demo.

Visual augmentations (V-AUG)

In the *V-AUG* condition, visual augmentations are overlaid on the performance (See Figure 2-C). Visual augmentations are graphical representations of the behaviour of the instrument updated in real time. Similar to Berthaut et al. [8] they represent:

- The sensors and their activity in order to make them more visible (Figure3, I)
- The mappings between sensors and sound processes, which appear when the musician manipulates them and disappear after. (Figure3, II)
- The three sound processes with three shapes whose appearance changes according to changes in their sound parameters (from either the musician or the computer) : colour hue represents pitch, size represents volume, rotation represents scrolling in the sound, contours represent a delay effect. (Figure3, III)

Disruptive visual augmentations (V-DIS)

In the *V-DIS* condition, visual augmentations are also overlaid on the performance but they are uncorrelated with the musician gestures. They were actually extracted from another performance. This condition aims at controlling that the effect of the visual augmentations on the experience are not only due to the increased richness of graphical material inclusion. The *V-DIS* condition is not introduced until Block 03 in order to let the participants integrate the correct behaviour of the visual augmentations. Note that the participants were not informed that incongruent visual augmentations could occur.

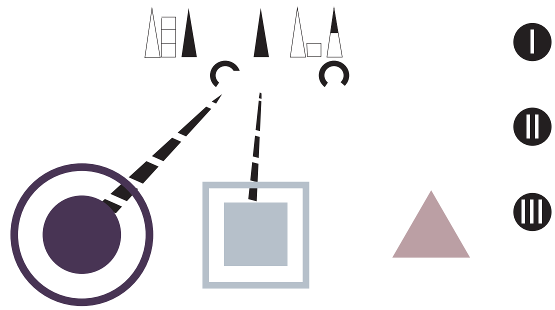

Figure 3. Visual augmentations are graphical representations updated in real time of the behaviour of the instrument. I) Sensors II) Mapping III) Processes. See Section *Visual augmentations (V-AUG)* for details.

Controlled stimuli

The stimuli used in this study are videos of short electronic music performances (about 30s) shot from an identical point of view[1]. Each video presents a musician behind a table on which a digital musical instrument (DMI) is placed next to a computer (see Figure 2-A). The video starts by a fade in with music already playing and the musician manipulating the DMI. The video ends with a fade out.

The control device used by the musician is a MIDI control surface (Korg NanoKontrol) composed of sensors : linear potentiometers (faders), rotating potentiometers (knobs) and buttons (see Figure 2-B). Linked to the device by USB, the computer is running a music software (Pure Data) allowing for the control of three sound processes (and their associated parameters) : a bass track (pitch, note_length, filter, volume), a beat/percussion track (trigger_note, echo, filter, volume) and a textural track (position_in_sample, filter, volume, echo). This setup (controller + computer) constitutes the common architecture of many DMIs.

The mapping

The mapping is the relation between the sensors of the control device and the sound parameters. It impacts multiple aspects of the performance [31]. In particular, the mapping influences the contribution of the musician by setting the ratio between the controls given to the musician and those given to autonomous processes. Thus the role of the mapping is crucial in the definition of an instrument [31]. In order to vary the contribution level of the musician, from very high to very low, we designed three instruments from the MIDI controller :

- In the first instrument, all sound parameters are associated with sensors, meaning that all changes in the sound of the

performance originate from the sensors manipulated by the musician. The musician's contribution is very high.
- In the second instrument, some of the parameters are automated while others are connected to sensors, the musician's contribution is lower.
- In the third instrument, most sound parameters are automated and some are shared, i.e. the musician can take control of them temporarily. The musician's contribution is therefore very low.

To limit the complexity in gesture perception by participants, all mappings were one-to-one : one sensor controls one parameter only.

Performance masking - Learning bias
The controlled evaluation of the augmentation techniques requires to display the same performances under each augmentation condition. Because of this repetitive exposition, the participant could memorise a performance and answer the questions related to a condition having in mind the cues provided by the same performance under a different condition. To avoid this learning bias, we created a number of sound banks, i.e. different bass, percussive and textural sounds, and used them in post production to generate multiple stimuli from original video footages. Thus, the resulting stimuli presented the very same gestures, performance structure and instrument mapping but with a slightly different sound output making it harder for a participant to notice the potential repetition of a performance. Besides this "performance masking", participants were informed before the experiment that the device used by the musician may have different settings throughout the videos and thus should be considered as a distinct instrument regardless of its identical physical aspect.

Participants
21 participants (16 males, 4 females, 1 deliberately not reported) aged of 34.9 $\pm$4.2 (22-57) were involved in the study. The protocol follows the ethic rules of the Helsinki Declaration. All participants were voluntary and signed an informed consent before getting equipped and start the experiment. The exclusion criteria included hearing and vision impairments. After control, no subject was excluded.

Procedure
After having read and signed the consent form, the participant was equipped with two watchbands measuring cardiac and electrodermal activity, and with a light eye-tracking device. The physiological aspects of the experiment are not mentioned in the present document and will be covered in a future study. During the whole experiment, the participant was seating at a desk in front of a laptop equipped with a mouse (see Figure 2 - D). She or he was informed that the experiment consisted in watching videos and answering the subsequent questions via a survey form. After each block, the participant was offered to take a short break.

The experiment started with a questionnaire to control the participant's hearing and sight abilities and to evaluate their expertise in three disciplines : music as a performer, electronic music (as a spectator in concerts) and recent technologies. Data treatment of the scores of expertise led to 3 expertise profiles : *musician* - *electro* - *techno*. These profiles were used to control the homogeneity of the groups in Block 02 that followed a between-groups design.

Data analysis
Data was recorded, anonymised and stored in real time during the experiment by a bespoke experiment software developed in Python. For technical reasons, one participant did not finish the second block and was excluded of the analysis for this block. Subjective reports were obtained via Likert scales and were analysed with parametric tools when the normality assumptions were met [57, 16]. All the analyses were conducted under the common frequentist paradigm and were combined to Bayesian statistics.

When the assumptions for parametric tests were met, we looked for a main effect through the analysis of variance (ANOVA or repeated measures ANOVA). We used a Mauchly's test to control the sphericity (with a Greenhouse-Geisser correction when required). In case of a significant main effect post-hoc T tests were conducted. When the assumptions for parametric tests were not met, the analysis was conducted with non parametric statistical treatments. For repeated measures, we used a Friedman test. In case of a significant main effect, we conducted pairwise comparisons with a Wilcoxon Signed-Rank Test. For dependent comparison, we used a Kruskall-Wallis test. In case of a significant main effect, we conducted pairwise comparisons with a Mann-Whitney Test. All post-hoc tests significance were adjusted with Bonferroni correction for multiple tests.

Bayesian statistics offer sounder analysis when dealing with relatively small samples like ours (n=21) [37, 33]. Besides, by clearly exposing the evidence in the data that support either the null or the alternative hypothesis, they also provide a more intuitive interpretation [39] and are increasingly considered as the researcher-centred design of statistics [37]. In this work, whenever possible, the ANOVA, the pairwise test [52, 44, 53, 51] and the linear regression [18, 19, 42] are expanded with their Bayesian version and a Bayes factor is reported as BF_{01} when data better support the null hypothesis and as BF_{10} when data support the alternative hypothesis (note that '01' becomes '10'). For example, the statement $BF_{10} = 2.4$ means that the data are 2.4:1 in favour of the alternative hypothesis, so 2.4 times more likely to occur under a model including the corresponding effect. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons[34, 63]. Analyses were performed with IBM SPSS v25, R studio 1.2 and JASP (Bayesian statistics) [34, 59].

TASKS AND RESULTS

Familiarisation
Before entering the first block of the protocol, the participants were asked to simply watch 3 short videos with no other instructions. The videos presented the same performance under three conditions : control, *V-AUG*, and *Explain*. After the trial, they were asked to freely comment on what they had just seen. Then they were informed that during the experiment they will see these three types of videos.

Subjective experience

The subjective experience is assessed in Block 01 and in Block 04 (See Figure 1) through a within subjects design (N = 21).

Protocol

In Block 01, the same short music performance is presented 3 times, once per condition, in random order : *V-AUG*, *Explain*, *Control*. In Block 04, another short music performance is presented 4 times, once per condition, in random order : *V-AUG*, *Explain*, *Control* and the disruptive condition *V-DIS* in which visual augmentations are not correlated with the performance. The participants were not informed that incongruent visual augmentations could occur.

Task

The task was the same for Block 01 and Block 04. The participants had to rate their global experience and three aspects that we hypothesise to support this experience, namely, the music, the visual aspect and the virtuosity of the musician. After each video, they had to answer 4 questions in randomised order on 7-point Likert scales :

- From 1 (min) to 7 (max), rate the global quality of the performance.
- Considering the visual aspect only, rate the performance from 1 (min) to 7 (max).
- Considering the virtuosity of the musician only, rate the performance from 1 (min) to 7 (max).
- Considering the music only, rate the performance from 1 (min) to 7 (max).

Results

The hypothesised decomposition of the subjective experience following the music, visual and virtuosity subjective ratings was strongly supported ($BF_{10} = 479$) by a linear regression model ($F_{(3,17)} = 14.7, p < 0.001$, R^2 adj : 0.67).

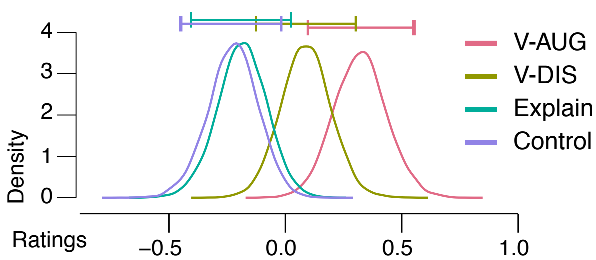

Figure 4. Bayesian posterior distributions of the global experience ratings in Block 04. Participants reported a better experience when the performance was augmented with visual augmentations. The *V-DIS* (uncongruent visual augmentations) and *Explain* conditions did not differ from the control condition.

A strong effect ($BF_{10} = 19.5$) of the SEATs was found on global experience ($\chi^2(3) = 17.40, p < 0.001$) only in Block 04. As illustrated on Figure 4, post-hoc tests showed that ratings in the *V-AUG* condition presented very strong evidence ($BF_{10} = 29.7$) of a superiority with the control condition and a substantial superiority ($BF_{10} = 3.3$) with the *Explain* condition. No difference ($BF_{01} = 1.9$) with the *V-DIS* condition was detected. However, the data showed evidence for the absence of difference between the control condition and the *V-DIS* condition ($BF_{01} = 2$) and the *Explain* condition ($BF_{01} = 3.6$).

An effect of the SEATs on the visual rating was found in both Block 01 and Block 04 ($F_{(3,60)} = 14.938, p < 0.001$ - $BF_{10} > 319$), obviously triggered by the graphical material inclusion of the visual augmentations.

Discussion

The visual augmentations led to a greater subjective experience in the last block but no effect of the augmentation techniques (SEATs) was measured in the first block. This can be explained by the learning curve associated to the visual augmentations. When we designed the experiment, we opted to deliver no prior information or comments about the conditions to the participants. So we hypothesised that the participants would require some expositions to the visual augmentations before integrating their behaviour and their purpose. Without a clear estimation of the learning curve, we placed the measurement of the subjective experience twice in the protocol, once at the beginning and once at the end. More formal information should be collected in ecological conditions, like a concert, but the visual augmentations used in the experiment are quite explicit and we advocate for a short learning curve of a few minutes. This estimation is supported by the emergence of better results in the *V-AUG* condition of Block 02 and a decisive contribution of the visual augmentations on the subjective comprehension assessed in Block 03.

Subjective comprehension

The subjective comprehension is assessed in Block 03 (See Figure 1) through a within-subjects design (N=21).

Protocol

Videos of short music performances were presented under four conditions : *V-AUG*, *V-DIS*, *Explain* and *Control*. Contrary to the *V-AUG* condition, the visual augmentations in the *V-DIS* condition are not correlated with the musician's gestures.

Task : Bellotti-Fyans challenges

This task is built upon 5 communication design issues _ namely Address, Alignment, Attention, Accident and Action _ introduced by Bellotti et al. [4] in a user-computer interaction context and transposed by Fyans et al. [28] to the perspective of spectators of interactions. In turn, we transpose these challenges from a design perspective to an evaluation perspective by assessing the perception of these design features by observers. We call them the Bellotti-Fyans challenges.

After each video, participants had to indicate on a 5-step likert scale their degree of accordance with statements built from the Bellotti-Fyans design challenges. Contrary to the *objective comprehension task* where participants must perform by giving good answers, in this task the participants were informed that there were no good or bad answer and they should report their inner feeling rather than a definitive answer. Only the extreme values of the scales had a label : "I totally disagree" and "I totally agree". The question asked was "To which extend do you agree with the following statement ?" :

- "In this video, I know when the musician is interacting with the instrument and when he is not."(Address)
- "In this video, I can see when the instrument is responding to the musician gesture and when it is not."

- "In this video, I can see if the musician is controlling the instrument or if he is not."(Action)
- "In this video, I can see when the instrument is correctly functioning and when it is not."(Alignment)
- "In this video, I can see if the musician or the instrument made a mistake."(Accident)

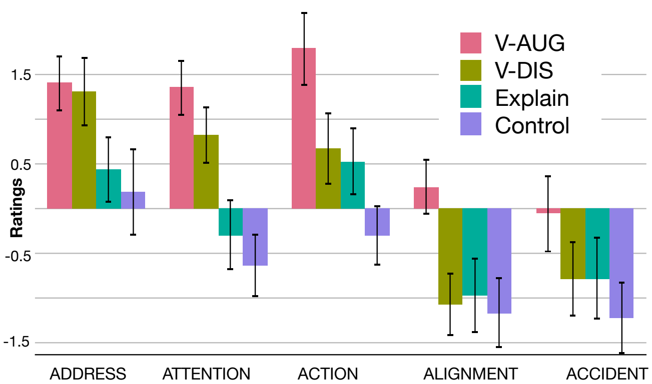

Figure 5. The effect of the Spectator Experience Augmentation Techniques (SEATs) on the Bellotti-Fyans challenges.

Results

The augmentation techniques (SEATs) often led to a distinct appreciation of a same interaction (See Figure 5). Despite rather subtle differences in the meaning of the questions, like in *Alignment* versus *Attention*, different patterns are clearly observed.

- Address : A substantial effect ($BF_{10} = 3.7$) of the SEATs was detected on the evaluation of the *Address* challenge ($\chi^2(3) = 10.07, p = 0.018$). Bayesian post-hoc tests showed weak to substantial evidence of a better evaluation for the *V-AUG* and the *V-DIS* conditions compared to the other conditions ($BF_{10} = 2$).
- Attention : A very strong effect ($BF_{10} = 160$) of the SEATs was detected on the evaluation of the *Attention* challenge ($\chi^2(3) = 12.95, p = 0.005$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the *Control* condition ($p = 0.02, BF_{10} = 99$) and to the *Explain* condition ($p = 0.051, BF_{10} = 13.8$) and weak evidence ($BF_{10} = 2.5$) for a better evaluation in the *V-DIS* condition compared to control.
- Action : A very strong effect ($BF_{10} = 161$) of the SEATs was detected on the evaluation of the *Action* challenge ($\chi^2(3) = 18.75, p < 0.001$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the control condition ($p = 0.03, BF_{10} = 112$), to the *Explain* condition ($p = 0.02, BF_{10} = 104$) and, with weak evidence, to the *V-DIS* condition ($BF_{10} = 2.7$).
- Alignment : A substantial effect ($BF_{10} = 5.12$) of SEATs was detected on the evaluation of the *Alignment* challenge ($\chi^2(3) = 10.99, p = 0.012$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the control condition ($p = 0.051, BF_{10} = 9$), the *Explain* condition ($BF_{10} = 2.8$) and quite interestingly a substantial difference ($BF_{10} = 4.4$) with the *V-DIS* condition.
- Accident : Data showed no effect ($BF_{01} = 2.4$) of the SEATs on the evaluation of the *Accident* challenge ($\chi^2(3) = 3.66, p = 0.301$).

Discussion

The assessment of the subjective comprehension revealed interesting contrasted results. By analysing the tendency of a spectator to project their ability in a fictive task, we are obviously not measuring their objective abilities but rather the confidence they have in the mental model they built up from the interaction [29]. The SEATs did not significantly influence the participants' subjective feeling of being able to detect errors from the musicians or the machine. From a cognitive point of view, this suggests that neither the visual augmentations nor the pre-performance explanations could reduce the spectator's difficulties in perceiving the performer's intention and predicting the result of an action [28].

However, visual augmentations (*V-AUG*) induce positive evaluations of the other Fyans-Bellotti challenges, where they overtake the *Control* and the *Explain* conditions, which suggests that they help spectators build and trust a mental model of the instrument and of the performer's interactions with it.

The difference between the congruent (*V-AUG*) and the disruptive augmentations (*V-DIS*) is not always measurable. The inclusion of incongruent visual augmentations led to the same feeling of being able to detect when the musician was interacting with the instrument (Address) and when the instrument was responding to the musician (Attention). Conversely the congruent condition *V-AUG* overtakes the *V-DIS* condition in the evaluation of the level of control of the musician (Action) and in the evaluation of the proper functioning of the instrument (Alignment). Since there is no global subjective rating of the stimuli in this task, these findings need refinements to clearly identify the link between the subjective comprehension and the global experience. Still, the design of SEATs and their evaluation can already benefit from these first insights of the approach by components.

Objective comprehension

The objective comprehension was assessed by two tasks, a *modulation attribution task* in Block 02 and a *Global Contribution Task* in Block 03 (See Figure 1).

Modulation attribution task

The task was to detect a given change of one of the processes parameters in a short music performance and correctly decide who from the musician or the computer was the author. The participants answered in real time, while the video was playing, by pressing a key associated with their answer (the musician or the computer) on the keyboard.

After each video, participants evaluated on Likert scales how difficult they found the task (from 'really easy' to 'really difficult') and what was their trust in their performance on 5-steps scale (1 : "I answered randomly", 5: "I'm very confident in my answer").

In order to limit the duration of the experiment, this task followed a between-groups design. The videos were displayed under a condition depending on the group to which the participant was randomly assigned. Groups: *V-AUG* (n=9), *Explain* (n=5 + 1 excluded), *Control* (n=6). The expertise of participants within each group was controlled, the 3 groups were homogeneous ($\chi^2_{(12)} = 25.48, p = 1$).

Results

The score of the *modulation attribution task* was computed by measuring the percentage of time during which the key pressed by the participant matched the actual origin of the modulation.

Participants reported that the task was quite difficult. The data confirmed that impression with an average score of 42.3% globally (see Table 1 for details).

Score [% of success]	V-AUG	Explain	Control
Mean	44.9	42.3	38.4
Stdev	4.0	4.3	6.3
Min	39.8	35.6	28.0
Max	53.7	47.0	45.0
Difficulty [-3,3]			
Median	0.000	1.000	0.500
Min	-0.667	-0.667	-0.667
Max	2.000	1.500	2.000
Confidence [-3,3]			
Mean	0.213	-0.700	-0.111
Stdev.	0.498	1.083	0.455
Min	-0.667	-2.000	-0.667
Max	0.667	0.667	0.667

Table 1. Modulation attribution task - Descriptive statistics

Despite concomitant results in favour of the visual augmentations (Table 1), the task did not permit to expose significant contrasts of the SEATs' influence on the participants' scores ($F_{(2,17)} = 3.215, p = 0.184$). This lack of power is mainly due to the rather small size of the groups and the difficulty of the task. The Bayesian analysis showed weak evidence in favour of an effect of the SEATs ($BF_{10} = 1.6$) and weak superiority of the scores in the *V-AUG* condition. No influence of the SEATs on the reported difficulty ($F_{(2,17)} = 0.257, p = 0.575$, $BF_{01} = 3.1$) was found, neither on the reported confidence ($F_{(2,17)} = 2.952, p = 0.220$ - $BF_{10} = 1.3$).

Global Contribution Task

This task occurred in Block 03 and followed a within-subjects design (N=21) with 2 factors : Augmentation technique : *V-AUG*, *V-DIS*, *Explain* and *Control* — Contribution level of the musician : *contrib_low*, *contrib_med*, *contrib_high*

Contrary to the *modulation attribution task* where the authorship of a specific event was targeted, in this task, the participants had to determine who from the musician or the computer was *globally* the most contributing over a whole performance.

As explained in Section *The mapping*, we designed three instruments which allowed us to vary the musician's and computer's respective contributions. These mappings resulted in, respectively, a large contribution of the computer, a large contribution of the musician and a rather balanced ratio still in favour of a greater contribution of the musician. Thus, the musician contributed the most in 66% (2/3) of the stimuli.

This two-factors design lead to 4 SEATs x 3 contribution levels = 12 videos. The stimuli implying the same contribution level of the musician were generated from the same original video footages. To control for a potential learning bias, these stimuli were post produced with different audio outputs following the *performance masking* technique detailed in Section *Controlled stimuli*.

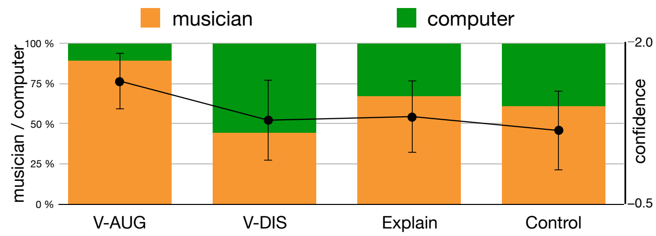

Figure 6. Global Contribution Task : the participants largely promoted the greater contribution of the musician (90%, *orange*) and were more confident in their answers *(black dots)* when the performances were augmented with visual augmentations (*V-AUG*). The performances augmented with incongruent visual augmentations (*V-DIS*) led to an under estimation of the musician contribution.

After each video, the participants indicated who from the musician or the computer was the most contributing and reported their level of confidence in their answer on a 6-step likert scale with labelled extreme values; 0: "I answered randomly" - 5: "I'm very confident".

Results

The average success score was $60.6\% \pm 12.5$ and 65.1% of the reported answers indicated that the musician contributed the most.

Considering the score of the participants to this task, the data showed a strong effect ($F_{(2,40)} = 7.18, p = 0.002$ - $BF_{10} = 600$) of the contribution level taken alone and a very strong effect ($F_{(6,120)} = 5.98, p < 0.001$ - $BF_{10} > 100000$) of the interaction between the contribution levels and the augmentation techniques. A second analysis was run to understand the details of the interaction. It exposed a very strong link ($BF_{10} > 7000$) between the choice of the participants for the most contributing and the SEATs ($\chi^2_{(4)} = 28.432, p < 0.001$). As illustrated in Figure 6, post-hoc tests showed that 90% of the stimuli presented in the *V-AUG* condition were associated to a greater reported contribution of the musician, 67% in the *Explain* condition, 60% for the control condition. Interestingly, the lowest count of reports for the contribution of the musician was measured in the *V-DIS* condition where 56% of the stimuli were associated with a greater contribution of the computer.

A moderate ($BF_{10} = 2.9$) effect of the SEATs was found on the confidence participants had in their answer ($F_{(3,60)} = 3.683, p = 0.017$). Following the overestimation of the musician contribution in the *V-AUG* condition, post-hoc tests showed that the confidence of the participants in their answer was significantly higher ($p = 0.049, BF_{10} = 4.19$) in the *V-AUG* condition than the control condition and ($BF_{10} = 2$) the Explain condition.

Discussion

The impact of augmentation techniques (SEATs) on the objective tasks is contrasted. The lack of power did not permit to clearly expose their role with weak evidence in favour of the visual augmentations in the *Modulation Attribution Task*. Conversely, the *Global Contribution Task* led to significant results and revealed a misleading effect of the visual augmentations that conducted the participants to over estimate the musician contribution. These findings are discussed in the general discussion alongside with the insights of the other blocks.

DISCUSSION

In this section, we combine the findings discussed in the previous sections and provide insights and perspectives on Spectator Experience Augmentation Techniques (SEATs), and in particular the use of visual augmentations. We also propose to add the *Association* as a new item in the Bellotti-Fyans challenges.

Visual augmentations offer a better experience

The findings of this study expose the positive role of the visual augmentations over multiple aspects of the experience of spectators of performances with digital instruments. They also confirm the weak contribution of the preliminary explanations reported in the literature [9].

The low efficiency of preliminary explanations can be interpreted as a difficulty for the audience to memorise and manipulate the large amount of information necessary to integrate complex interactions. The real time support of visual augmentations as well as their ability to abstract interactions mechanisms probably made a strong difference. Yet, with this sole first study, we can not strengthen these interpretations because we are missing formal descriptions of the two techniques. A taxonomy of the SEATs could help to refine the study of their impact by allowing for an approach by components rather than the holistic comparison we present here.

The contribution of the visual augmentations was not measurable on the ratings of the music, or the virtuosity of the musician, but the data do support a significant improvement of the overall experience of the spectator. Importantly, the effective impact of the visual augmentations is not due to the only effect of the additional graphic material of this technique. When the augmentations were not correlated to the interactions, participants gave a better rating to the visual aspect but no difference with the control condition was detected in terms of global experience.

The interest for the visual augmentations was also reported during short interviews when we asked the participants what augmentation technique they would choose if they were to attend an electronic music performance. A large majority (80%) opted for the visual augmentations but we also had some participants advocating for no augmentations at all, neither explanations nor visual augmentations, to preserve "their natural experience of music". Depending on the expertise of the participants and their tolerance to graphical inclusions in terms of cognitive interference, the SEATs should permit to vary the level of details and even let the user turn it off.

Restoring the trust in musicians

Even if the visual augmentations did not increase the objective comprehension of the spectator as we hypothesised, they led to an interesting side effect. When the participants had to distinguish the contribution of the musician from the one of the computer, they over estimated the contribution of the musician and at the same time were more confident in their evaluation when the performances were displayed with visual augmentations (*V-AUG*). From an objective perspective, this observation lowers the efficiency of the visual augmentations used in the study to clearly expose the reality. But it also represents a positive bias that can compensate the doubt of uncomprehending spectators and restore their trust in the musician. Conversely, the disruptive augmentations (*V-DIS*), extracted from another performance, led to a sub estimation of the contribution of the musician, a reduced confidence of the participants in their answers, and the misperception of a greater contribution of the computer.

These observations could highlight the presence in the audience experience of a subtle balance between the continuous objective cues emitted by the performer and an inner, potentially discreet, validation by the observer that everything is going as expected, with a clear contribution of the musician and a moderate support of the machine. As a design guideline, this suggests that the more a performance is demanding in term of artistic virtuosity, the more the audience needs to be able to perceive the ratio of the human/machine contribution, in order to preserve their trust.

The unexpected benefits of disruptive cues

The insertion of uncorrelated visual augmentations led the participants to evaluate differently some aspects of the interactions but not all. The confusion is more visible in the results of the Bellotti-Fyans challenges (See the discussion of the *Subjective comprehension* section). These contrasts in the results reflect the composite nature of the mental models at the origin of the spectator's experience and expectations [32, 25]. They also highlight the inconsistent effect of the augmentations used in the study. Now that the potential efficiency of the visual augmentations is formally supported, design research should compare different types of visual augmentations, graphical material or levels of detail and explore their influence on identified observer's mental models like the one from the Bellotti-Fyans challenges or the components of the experience we addressed in this work.

Controlling the level of magic

We saw that disruptive visual augmentations can lead to the perception of a lower contribution of the musician. They also tend to blur the mental models of the observers. This property could be manipulated by the musicians. By designing the SEATs to increase the amount of disruptive cues in the augmentations, musicians could decide to veil their manipulations at the same time of a music climax and let the audience think that something "magical" is happening. In a sense, if too much doubt about the musician's contribution on the music can reduce the interest of the audience, a parsimonious approach can lead to interesting strategies in the artistic performance, like a navigation in the 2-dimensional space proposed by Reeves et al.[49] (See Figure 7).

Call for a new design challenge : Association

The Bellotti-Fyans challenges provided insightful data for the evaluation of the subjective experience of the spectators. However, it is not clear if these challenges are sufficient to characterise the ability of a device to expose the contributions of both the user and the system to the audience. To this end, we propose to add another "A" to the five already present in the Bellotti-Fyans challenges with the "Association" challenge. The term *Association* relates to the design decisions that enable a device to expose to the spectators the respective and shared contributions of its user and the system itself. This challenge also relates to the exclusivity dimension of agency defined by Wegner et al. [61], transposed to the audience.

Interactions with simple devices do not need such augmentations as the integration by an observer of the inherent cues is obvious to evaluate the contribution of the user, like when observing a seller with a cash machine. But when the interactions become more complex, when the system they address is capable of producing actions that mimic the one a user could make, like a musician with a digital instrument, or a worker with an exoskeleton, the contributions of the user and the machine are mixed from a spectator perspective. An observer of such interaction implying a device with a high *Association* score should have a clear appreciation of the actual part of the user's actions in the intended production, without the difficulty to integrate the multiple cues perceived from the interaction. The shared contribution of both the system and the user should also be appreciable. The decisions that lead to a good *Association* score should be part of the design process or ensured by dedicated augmentations.

Limitations

In this work, we aimed at investigating refined aspects of the experience and minimising the biases inherent to the study of human behaviour by conducting a controlled study. Although the results of the experiment did provide genuine insights on the effect of SEATs, a number of limitations remain. A first improvement would be a validation of the questionnaires, in particular regarding subjective comprehension.

As other "in the lab" studies, our work relies on the data analysis methods for controlled experimentation that can be considered as a risk for misleading or overly narrow results [45]. In consequence, we chose to include interviews and to mix objective and subjective trials to limit the bias often met in field studies [13]. Besides, we believe that the extension of the data analysis with Bayesian statistics is particularly helpful in this context of medium sample size (21 participants).

Finally, another limitation of our work naturally lies in the evaluation of a live phenomena in a controlled environment. In this context, important aspects of the experience of live music are missing, from the social environment to the ceremonial of the concert hall. Aware of these limitations, we do not consider this work as an exhaustive approach, much to the contrary as we are very open to the contributions of other disciplines in the study of the relations between humans and artificial systems. Joining the "movement" of mixed-methods [22], we believe that there should be more dialog between the "in the lab" and "in the wild" paradigms. This enthusiasm is shared by recent discussions in the HCI community [43], which highlights the importance of identifying insights and biases of each paradigm.

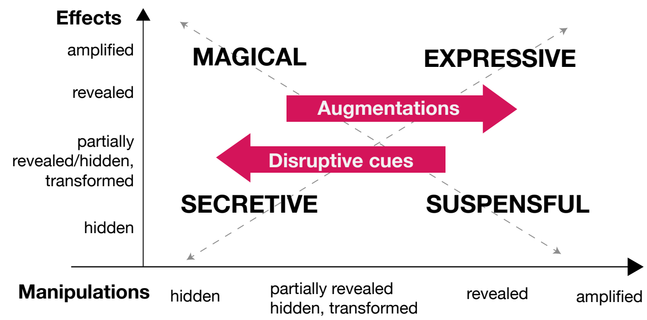

Figure 7. When controlled by the musician, the nature and the amount of visual augmentations allow to navigate in the 2-dimensional space proposed by Reeves et al.[49] (Illustration adapted from [49])

CONCLUSION

This work presents a controlled study to evaluate the effect of Spectator Experience Augmentation Techniques on the audience experience. We showed that, contrary to pre-concert explanations, visual augmentations increase the global subjective experience, support the observer's confidence in their mental representations of the interactions, and induce an over estimation of the musician's contribution compared to the machine's, thus restoring spectators' trust in electronic musicians. We proposed to extend the Bellotti-Fyans challenges with the *Association* challenge, as the ability of a device to expose the contributions of both the user and the system to the audience.

We showed strong evidence of differences in the impact of the preliminary explanations and the visual augmentations. However, the explanation of these contrasts would benefit from formal descriptions of the two techniques. To this end, future work could propose a taxonomy of SEATs to facilitate their formal comparisons and allow for grounded design guidelines. We believe that the evaluation of Spectator Experience Augmentation Techniques on specific components of the experience can provide robust insights for applications in many other domains where the collaboration of humans and machines occur in public, like performance interfaces or pedagogical devices.

[1]All stimuli, illustration videos of the conditions, anonymised raw results and statistical analyses can be found here : http://o0c.eu/0DA

REFERENCES

[1] Joanne Armitage. 2016. Revealing timelines: Live coding and its gestures. *Proceedings of ICLC* (2016).

[2] Jerônimo Barbosa, Filipe Calegario, Veronica Teichrieb, Geber Ramalho, and Patrick McGlynn. 2012. Considering Audience's View Towards an Evaluation Methodology for Digital Musical Instruments.. In *Proceedings of New Interfaces for Musical Expression*.

[3] Mehmet Aydin Baytas, Tilbe Göksun, and Oğuzhan Özcan. 2016. The Perception of Live-sequenced Electronic Music via Hearing and Sight. In *Proceedings of New Interfaces for Musical Expression*.

[4] Victoria Bellotti, Maribeth Back, W Keith Edwards, Rebecca E Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 415–422.

[5] Steve Benford, Chris Greenhalgh, Andy Crabtree, Martin Flintham, Brendan Walker, Joe Marshall, Boriana Koleva, Stefan Rennick Egglestone, Gabriella Giannachi, Matt Adams, and others. 2013. Performance-led research in the wild. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 3 (2013), 14.

[6] Steve Benford, Chris Greenhalgh, Adrian Hazzard, Alan Chamberlain, Maria Kallionpää, David M Weigl, Kevin R Page, and Mengdie Lin. 2018. Designing the audience journey through repeated experiences. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 568.

[7] Florent Berthaut, David Coyle, James W Moore, and Hannah Limerick. 2015. Liveness through the lens of agency and causality. In *Proceedings of New Interfaces for Musical Expression*.

[8] Florent Berthaut, Mark Marshall, Sriram Subramanian, and Martin Hachet. 2013. Rouages: Revealing the mechanisms of digital musical instruments to the audience. In *Proceedings of New Interfaces for Musical Expression*.

[9] SMA Bin, AP MCPHERSON, N BRYAN-KINNS, and others. 2016. Skip the pre-concert demo: How technical familiarity and musical style affect audience response. (2016).

[10] SM Astrid Bin, Nick Bryan-Kinns, A McPherson, and others. 2017. Hands where we can see them! investigating the impact of gesture size on audience perception. (2017).

[11] SM Astrid Bin, Fabio Morreale, Nick Bryan-Kinns, and Andrew P McPherson. 2017. In-the-moment and beyond: Combining post-hoc and real-time data for the study of audience perception of electronic music performance. In *IFIP Conference on Human-Computer Interaction*. Springer, 263–281.

[12] Alan S Brown and Jennifer L Novak. 2007. *Assessing the intrinsic impacts of a live performance*. WolfBrown San Francisco, CA.

[13] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1657–1666.

[14] Claude Cadoz and Marcelo M Wanderley. 2000. Gesture-music. (2000).

[15] Olivier Capra, Florent Berthaut, and Laurent Grisoni. 2017. Toward Augmented Familiarity of the Audience with Digital Musical Instruments. In *International Symposium on Computer Music Multidisciplinary Research*. Springer, 558–573.

[16] James Carifio and Rocco Perla. 2008. Resolving the 50-year debate around using and misusing Likert scales. *Medical education* 42, 12 (2008), 1150–1152.

[17] Alan Chamberlain, Andy Crabtree, Tom Rodden, Matt Jones, and Yvonne Rogers. 2012. Research in the wild: understanding'in the wild'approaches to design and development. In *Proceedings of the Designing Interactive Systems Conference*. 795–796.

[18] Merlise A Clyde, Joyee Ghosh, and Michael L Littman. 2011. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics* 20, 1 (2011), 80–101.

[19] Guido Consonni, Dimitris Fouskakis, Brunero Liseo, Ioannis Ntzoufras, and others. 2018. Prior distributions for objective Bayesian analysis. *Bayesian Analysis* 13, 2 (2018), 627–679.

[20] Nuno N Correia, Deborah Castro, and Atau Tanaka. 2017. The role of live visuals in audience understanding of electronic music Performances. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*. ACM, 29.

[21] Nuno N Correia and Atau Tanaka. 2017. AVUI: Designing a toolkit for audiovisual interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1093–1104.

[22] John W Creswell and Amanda L Garrett. 2008. The "movement" of mixed methods research and the role of educators. *South African journal of education* 28, 3 (2008), 321–333.

[23] M. Csikszentmihalyi. 2009. *Flow: The Psychology of Optimal Experience*. HarperCollins e-books. `https://books.google.fr/books?id=epmhVuaaoK0C`

[24] Sofia Dahl and Anders Friberg. 2007. Visual perception of expressiveness in musicians' body movements. *Music Perception: An Interdisciplinary Journal* 24, 5 (2007), 433–454.

[25] James K Doyle and David N Ford. 1998. Mental models concepts for system dynamics research. *System dynamics review: the journal of the System Dynamics Society* 14, 1 (1998), 3–29.

[26] Mark Duffett. 2012. Applying durkheim to elvis. *Transatlantica. Revue d'études américaines. American Studies Journal* 2 (2012).

[27] Sidney Fels, Ashley Gadd, and Axel Mulder. 2002. Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound* 7, 2 (2002), 109–126.

[28] A Cavan Fyans, Michael Gurevich, and Paul Stapleton. 2009. Spectator understanding of error in performance. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 3955–3960.

[29] A Cavan Fyans, Michael Gurevich, and Paul Stapleton. 2010. Examining the Spectator Experience.. In *Proceedings of New Interfaces for Musical Expression*, Vol. 10. 1–4.

[30] Michael Gurevich and A Cavan Fyans. 2011. Digital musical interactions: Performer–system relationships and their perception by spectators. *Organised Sound* 16, 2 (2011), 166–175.

[31] Andy Hunt, Marcelo M Wanderley, and Matthew Paradis. 2003. The importance of parameter mapping in electronic instrument design. *Journal of New Music Research* 32, 4 (2003), 429–440.

[32] David Brian Huron. 2006. *Sweet anticipation: Music and the psychology of expectation*. MIT press.

[33] John PA Ioannidis. 2005. Why most published research findings are false. *PLoS medicine* 2, 8 (2005), e124.

[34] JASP Team. 2019. JASP (Version 0.11.1)[Computer software]. (2019). `https://jasp-stats.org/`

[35] Sergi Jordà. 2003. Interactive music systems for everyone: exploring visual feedback as a way for creating more intuitive, efficient and learnable instruments. In *En Proceedings of the Stockholm Music Acoustics Conference, August*. 6–9.

[36] Sergi Jordà. 2004. Digital instruments and players: part I—efficiency and apprenticeship. In *Proceedings of the 2004 conference on New interfaces for musical expression*. 59–63.

[37] Matthew Kay, Gregory L Nelson, and Eric B Hekler. 2016. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4521–4532.

[38] Patrick Kennedy, David B Miele, and Janet Metcalfe. 2014. The cognitive antecedents and motivational consequences of the feeling of being in the zone. *Consciousness and cognition* 30 (2014), 48–61.

[39] John K Kruschke. 2010. What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences* 14, 7 (2010), 293–300.

[40] Chi-Hsia Lai and Till Bovermann. 2013. Audience Experience in Sound Performance.. In *Proceedings of New Interfaces for Musical Expression*. 170–173.

[41] 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*. ACM, 1845–1854.

[42] Feng Liang, Rui Paulo, German Molina, Merlise A Clyde, and Jim O Berger. 2008. Mixtures of g priors for Bayesian variable selection. *J. Amer. Statist. Assoc.* 103, 481 (2008), 410–423.

[43] Can Liu, Hamed S Alavi, Enrico Costanza, Shumin Zhai, Wendy E Mackay, and Wendy Moncur. 2019. Rigor, Relevance and Impact: The Tensions and Trade-Offs Between Research in the Lab and in the Wild.. In *CHI Extended Abstracts*.

[44] RD Morey and JN Rouder. 2015. BayesFactor (version 0.9. 10-2). *Computer software* (2015).

[45] Jakob Nielsen. 2004. Risks of quantitative studies. *Jakob Nielsen's Alertbox* (2004).

[46] Dario Páez, Bernard Rimé, N Basabe, A Wlodarczyc, and L Zumeta. 2012. Collective emotional gatherings. Five studies of identity fusion, shared beliefs and social integration. *Manuscript in preparation* (2012).

[47] Olivier Perrotin and Christophe d'Alessandro. 2014. Visualizing Gestures in the Control of a Digital Musical Instrument. In *Proceedings of New Interfaces for Musical Expression*. Goldsmiths, University of London, London, United Kingdom, 605–608. `http://www.nime.org/proceedings/2014/nime2014_406.pdf`

[48] Stuart Reeves. 2011. *Designing interfaces in public settings: Understanding the role of the spectator in Human-Computer Interaction*. Springer Science & Business Media.

[49] Stuart Reeves, Steve Benford, Claire O'Malley, and Mike Fraser. 2005. Designing the spectator experience. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 741–750.

[50] F Rheinberg and R Vollmeyer. 2003. Flow experience in a computer game under experimentally controlled conditions. *Zeitschrift fur Psychologie* 211, 4 (2003), 161–170.

[51] Jeffrey N Rouder, Christopher R Engelhardt, Simon McCabe, and Richard D Morey. 2016. Model comparison in ANOVA. *Psychonomic bulletin & review* 23, 6 (2016), 1779–1786.

[52] Jeffrey N Rouder, Richard D Morey, Paul L Speckman, and Jordan M Province. 2012. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology* 56, 5 (2012), 356–374.

[53] Jeffrey N Rouder, Richard D Morey, Josine Verhagen, April R Swagman, and Eric-Jan Wagenmakers. 2017. Bayesian analysis of factorial designs. *Psychological Methods* 22, 2 (2017), 304.

[54] W Andrew Schloss. Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research* 32, 3 (????).

[55] Catherine Stevens, Renee Glass, Emery Schubert, Johnson Chen, and Heather Winskel. 2007. Methods for measuring audience reactions. In *Proceedings of the inaugeral International Conference on Music Communication Science*. Citeseer, 155.

[56] Caleb Stuart. 2003. The object of performance: Aural performativity in contemporary laptop music. *Contemporary Music Review* 22, 4 (2003), 59–65.

[57] Gail M Sullivan and Anthony R Artino Jr. 2013. Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education* 5, 4 (2013), 541–542.

[58] Luca Turchet, Travis West, and Marcelo M Wanderley. 2019. Smart Mandolin and Musical Haptic Gilet: effects of vibro-tactile stimuli during live music performance. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound on ZZZ*. 168–175.

[59] Eric-Jan Wagenmakers, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Jonathon Love, Ravi Selker, Quentin F Gronau, Martin Šmíra, Sacha Epskamp, and others. 2018. Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review* 25, 1 (2018), 35–57.

[60] Chen Wang and Pablo Cesar. 2017. The Play Is a Hit: But How Can You Tell?. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. ACM, 336–347.

[61] Daniel M Wegner and Thalia Wheatley. 1999. Apparent mental causation: Sources of the experience of will. *American psychologist* 54, 7 (1999), 480.

[62] David Wessel and Matthew Wright. 2002. Problems and prospects for intimate musical control of computers. *Computer music journal* 26, 3 (2002), 11–22.

[63] Peter H Westfall, Wesley O Johnson, and Jessica M Utts. 1997. A Bayesian perspective on the Bonferroni adjustment. *Biometrika* 84, 2 (1997), 419–427.