

Levels of Detail in Visual Augmentations for the Novice and Expert Audiences

Olivier Capra, Florent Berthaut, Laurent Grisoni

▶ **To cite this version:**

Olivier Capra, Florent Berthaut, Laurent Grisoni. Levels of Detail in Visual Augmentations for the Novice and Expert Audiences. Computer Music Journal, In press. hal-03106626

HAL Id: hal-03106626
https://hal.science/hal-03106626

Submitted on 11 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Levels of Detail in Visual Augmentations for the Novice and Expert Audiences

Olivier Capra, Florent Berthaut, and Laurent Grisoni

Olivier Capra
ETHICS EA 7446 - Anthropo-Lab,
14 boulevard Vauban,
CS 40109,
F-59016 Lille cedex,
France
olivier.capra@univ-catholille.fr

Florent Berthaut
Univ. Lille, CNRS, Centrale Lille,
UMR 9189 CRIStAL,
F-59000 Lille,
France
florent.berthaut@univ-lille.fr

Laurent Grisoni
Univ. Lille, CNRS, Centrale Lille,
UMR 9189 CRIStAL,
F-59000 Lille,
France
laurent.grisoni@univ-lille.fr

Abstract

Digital Musical Instruments offer countless opportunities for musical expression as they allow artists to produce sound without the physical constraints of analogical instruments. However, by breaking the intuitive link between gestures and sound, they may hinder the audience experience, making the musician's contribution and expressiveness difficult to perceive. In order to cope with this issue without altering the instruments, researchers and artists have designed techniques to augment their performances with additional information, through audio, haptic or visual modalities. These techniques have, however, only been designed to offer a fixed level of information, without taking into account the variety of spectators' expertise and preferences. In this paper, we introduce the concept of controllable *Level of Details* (LOD) for visual augmentations. We investigate their design, implementation and effect on objective and subjective aspects of audience experience. We conduct a controlled experiment with 18 participants, including novices and experts in electronic music. Our results expose the subjective nature of expertise and its biases. We analyse quantitative and qualitative data to reveal contrasts in the impact of LOD on experience and comprehension for experts and novices. Finally, we highlight the diversity of usage of LODs in visual augmentations by spectators and propose a new role on stage, the augmenter.

Introduction

Digital Musical Instruments offer countless opportunities for musical expression as they allow artists to produce sound without the physical constraints of acoustic instruments. This *control dislocation* (Miranda and Wanderley 2006), by breaking the intuitive link between gestures and sound, may however hinder the audience experience making the musician's contribution and expressiveness difficult to perceive. Consequently, it also contributes to degrading spectator's attributed agency (Berthaut et al. 2015), that is

the level of control they perceive from the musician. Such difficulties in the integration of the musician's gestures can lower the interest of observers (Schloss 2003) and lead them to doubt the genuine contribution of the artist compared to the one of autonomous processes such as prerecorded audio samples or computer controlled sequences. Furthermore, the diversity and complexity of DMIs makes it difficult for the audience to build a familiarity with every instrument.

Thus, audience experience progressively became an important aspect in the creation of Digital Musical Instruments (DMIs), either as an evaluation method (Barbosa et al. 2012) or as a dimension which should be addressed at the design (Fels et al. 2002; Jordà 2003; Correia and Tanaka 2017) or performance stages (Reeves et al. 2005; Benford et al. 2018). Artists and researchers alike have designed techniques which augment the instruments with additional information to improve the audience experience and restore the trust of spectators in the musician's involvement. While these techniques explore different modalities (visual, haptic, auditory) and address different aspects of the performance (technical, gestural, intentional), they mostly offer a fixed level of information to all spectators.

However, we think that augmenting the audience experience can be more effective when considering spectators from an individual perspective. The information needed by each spectator can differ depending on their personal sensitivity and expertise. In order to ensure an optimal experience for spectators, we propose to allow the audience to dynamically change this level of information using visual augmentations with variable levels of detail (LODs).

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

Augmenting the audience experience

A number of augmentation techniques for spectator experience have been designed. Here we only provide a few examples, a more detailed analysis can be found in the taxonomy that we proposed in Capra et al. (2020b).

Perhaps the simplest is the organisation of pre-concert demonstrations, such as described by Bin et al. (2016). More common is the use of visual projections that represent the instrument structure and parameters or musician's gestures. Examples can be found in many electronic performances, with accompanying visuals displaying changes in sound processes as abstract or figurative elements. Perrotin and d'Alessandro (2014) have proposed to display the musical controls of musicians in an orchestra with a video projection, to help the audience perceive the actions of each orchestra member by representing both gestures and musical parameters. Similarly, Correia et al. (2017) discuss the role of visuals in live performances and insist on the importance of showing both the gestures (interface) and parameters to the audience. Berthaut et al. (2013) describe an augmented reality which can be used to reveal the mechanisms of DMIs. Haptic augmentations can also be created to increase the audience's engagement, as proposed by Turchet and Barthet (2019). All these augmentation techniques however only offer a fixed set of information for the whole audience.

Benford et al. (2018) go beyond fixed information by combining projected visual augmentations during the performance and visual/textual augmentations on a mobile app after the performance, thus allowing spectators to access two different levels of representation. Finally, Capra et al. (2018) propose adaptive augmentations as part of a pipeline for augmented familiarity, but they do not provide an implementation or evaluate the impact of the described levels. In contrast, we propose to adapt the amount and the type of information provided by visual augmentations using a Level Of Detail approach.

Contribution

In this paper, we first introduce the concept of *Level of Details* (LOD) for visual augmentations, gathering LOD approaches from research fields, other than those in music. Second we describe the design and implementation of dynamic and controllable LODs for the audience of digital musical performances. Third, through a controlled experiment based on a protocol that we proposed in Capra et al. (2020a), we study the effect of LODs on the experience of novice and expert spectators, and investigate how they could be used in performance settings.

Levels of Detail for the Visual Augmentation of DMIs

The concept of *Level of Detail* (LOD for short) originates from the field of computer graphics (Luebke et al. 2003) where 3D models and scenes complexity are adapted in order to reduce rendering load. It takes inspiration from existing signal analysis tools, such as wavelets (Stollnitz et al. 1996) or more basic simplifications systems such as down-sampling. LODs are meant to bring some flexibility in terms of calculus cost in all possible aspects of 3D representations (geometric models, textures, collision detection, etc.) by allowing one to adapt the representation to the context of use. Such adaptations are usually made thanks to the context of visualisation (expectations of users, hardware possibilities, etc.).

In the Human Computer Interaction (HCI) literature, LODs allow users to access different levels of complexity in the interface, such as with Zoomable User Interfaces (Bederson and Hollan 1994), or in a musical context to build and manipulate complex musical structures (Barbosa et al. 2013). Finally, LODs have been also used in augmented reality (Sung et al. 2014) to provide access to more or less detailed information on physical objects and in the field of information visualisation to adapt quantity of information in order to limit visual overload (Holten 2006; Wang et al. 2006).

Practical existing applications fields where LODs are used involve two classes of techniques, depending on the nature of data to handle. Such data may be *discrete* (i.e. sampling from a continuous phenomenon, whatever its nature), in which case most of the existing digital techniques can be adapted from interactive graphics and signal analysis. Such data may also be *symbolic* (i.e. referring to a dimension of information that is mostly conceptual and can not be directly represented by any sampling set), in which case LODs may be achieved by drawing inspiration from other communities, such as data visualisation and HCI.

LODs applied to augmentations for the audience

In this paper, we propose to apply the LODs approach to the design of augmentations for the audience of digital musical performances. As discussed above, digital musical interactions can prove highly difficult to perceive and understand, due to potentially small or hidden sensors and gestures, potentially complex mappings between sound and gestures, and complex and partly autonomous (pre-defined or automated) sound generating processes. Augmentations proposed in the literature aim at compensating this by providing the audience with information to enrich their experience.

Technically, there are no major obstacles for these augmentations to provide access to all information from the instrument : the exact sensor values, the audio graph that results in the sound with all the used processes and their corresponding parameters, the list of mappings between all the sensors and all the sound parameters.

However, this volume of information might not benefit the audience due to various reasons, such as:

- too much information at once;
- information requiring some expertise on DMIs to be understood;

- inconsistency with the audience preference when attending a performance, which might range from trying to understand the musician's actions to only focusing on the music.

Therefore, we believe that it is essential to provide a mechanism for the spectator to select the level of detail provided by these augmentations. The LOD approach can help adapt augmentation techniques, in our case visual augmentations, to the variety of expertise and preferences of spectators. In the idea of a better adaptation to personal needs, LODs on augmentations could be chosen dynamically by the spectators during the performance, either individually or as a group.

In the following sections, we describe how these LODs can be applied to visual augmentations, and how they can be implemented.

Levels of Detail in Visual Augmentation

In this paper, we apply our LOD approach more specifically to *visual augmentations* for the audience. These augmentations are graphical representations of the controls and mechanisms of a DMI, which are superimposed on the physical performance with the help of an augmented reality display. The purpose of visual augmentations is to reveal aspects of DMIs that are not easily perceived by the audience due to their lack of familiarity with them and the absence of physical link between gesture and sound. This includes subtle and/or hidden gestures sensed by the interface, complex or unusual mappings between the gestures and the various controllable parameters and the dynamic behaviour, potential range of output and internal structure of a DMI.

Following what Berthaut et al. (2013) proposed , our visual augmentations represent the three main sections of the instrument :

- the physical *interface* composed of sensors (e.g. a MIDI control surface);

- the *mappings*, i.e. the connections between sensors and musical parameters (e.g. the first fader controls the volume of the first audio track);
- the *processes* (e.g. tracks, loops, patterns) that generate the sound.

An important aspect of visual augmentations is that they do not restrain the design of DMIs. Instrument designers and musicians are free to choose their interfaces, mappings and processes with expressiveness in mind, without worrying about the transparency (Fels et al. 2002) of the musicians' actions or the familiarity (Gurevich and Fyans 2011) of the audience with the instrument, since these aspects are handled by the augmentations.

However, the potential complexity of DMIs implies that visual augmentations may become too detailed if one aims at representing all their events and components, which might in turn degrade the spectator experience that we are trying to enhance (Leman et al. 2008). Spectators might also prefer more or less detailed information for aesthetic reasons and at various times in the performance. Finally, musicians or accompanying visual artists might want to modify the level of information provided in order to alter the audience experience during the performance, e.g. to change from expressive to magical interfaces Reeves et al. (2005).

We propose to implement LODs in visual augmentations by defining dedicated levels of detail for each section (Interface, Mappings, Processes) of the visual augmentations. These *local LODs* can be chosen independently or combined as *global* LODs such as the ones we describe in section .

Local LODs

As illustrated in Figure 1, we propose 4 levels of detail for the Interface section, 3 levels for the Mappings section, and 5 levels for the Processes section. Each *local LOD* features a level 0 in which the section is not augmented. If all three sections are at level 0,

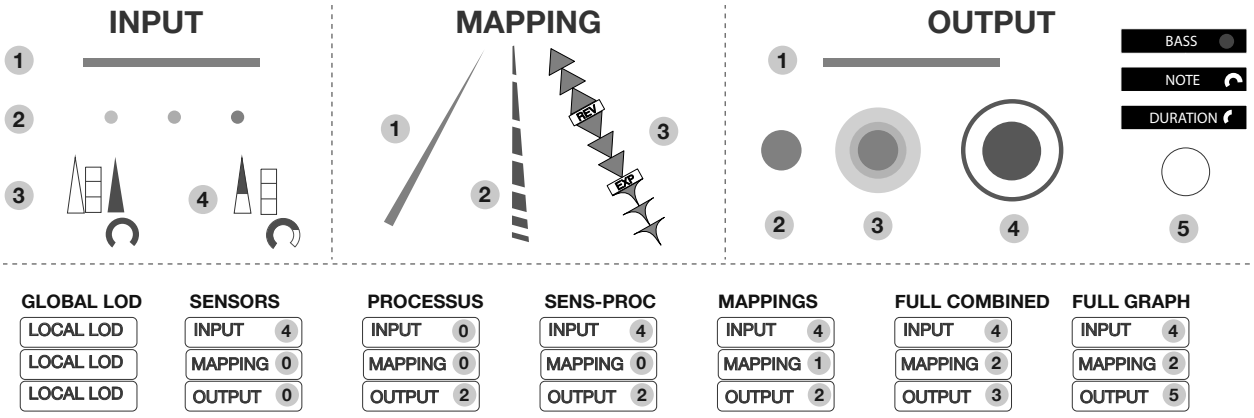


Figure 1. (top) Local LODs (Input, Mapping, Output) refer to a specific dimension of the interaction with a DMI. We also provide examples for each level in these Local LODs. (bottom) The global LODs used in the experiment and their respective sets of local LODs.

no information is added to the performance. One should note that the information provided by each level can be displayed in different ways, the representations proposed in our implementation are only one of the many possibilities that artists can explore.

Interface section

Level 1 only indicates the global activity, e.g. when the musician performs a gesture sensed by the system.

Level 2 represents the activity of each sensor of the physical interface, allowing one to perceive fast and complex gestures such as bi-manual or multi-finger interactions.

Level 3 describes both the activity and the type of each sensor (discrete/continuous, shape of sensor ...).

Level 4 adds a representation of their values and range.

Mappings section

Level 1 only describes to which processes the sensors are connected.

Level 2 refines the connection to the parameter level by representing on the links the multiple parameters modified through this mapping. For example the link can change colour to show its impact on the pitch of the associated process, while changing texture to

show that it also impacts the timbre of the sound.

Level 3 adds a representation of the operation or series of operations which transform sensor values into parameters values (Fels et al. 2002), e.g. scaling, inverting, combining and so on.

Processes section

Level 1 visualises the output of the system as a whole, merging the activity of all sound processes.

Level 2 provides a detailed activity for each process of the system, e.g. a distinct shape whose size indicates the volume of the corresponding sound process.

Level 3 adds a dynamic representation of parameters (i.e. inputs) that can be controlled on the processes.

Level 4 adds parameters names, types and values range, i.e. as performers would see them when performing with a GUI.

Level 5 provides a detailed representation of the complete internal graph of audio synthesis and effects that generate the sound of each process. It corresponds to what the musician would access when designing their instrument, and is potentially similar to the mental model they have when performing.

While the LOD could be chosen by spectators independently for each section, we believe a simpler solution is to define a number of *global* levels, i.e. presets of *local LODs*, so that spectators are provided during the performance with a unique control.

Global LODs

Global LODs are a combination of *local LODs*. They provide the spectators with a convenient way to control the level of detail by modifying several sections at a time : Interface (I), Mappings (M) and Processes (P). For instance, "*SENSORS* (I4-M0-PO)" is a *global LOD* called "SENSORS" and uses Level 4 for the Interface section and Level 0 for the

others.

In the following study, we use 7 global LODs with increasing quantity of information (See Figure 2).

NONE (I0-M0-P0) provides no information at all. The performance remains unchanged.

SENSORS (I4-M0-PO) amplifies the gestures performed by displaying representations of the types and values for all sensors of the interface. It is therefore similar to the level of details provided by Turchet and Barthet (2019) with haptics, and Perrotin and d'Alessandro (2014) for visuals. In the case of our study, faders, knobs and buttons of a MIDI controller are displayed.

PROC (I0-M0-P2) displays the sound processes of the instrument as separate shapes with graphical parameters associated to extracted audio features (loudness with size, pitch with colour hue, brightness with colour luminance), allowing spectators to identify the broad structure of the instrument and the activity of processes. This LOD corresponds to the representations traditionally used to illustrate electronic music performances (e.g VJiing) and defined as audiovisual entities by Correia et al. (2017).

SENS_PROC (I4-M0-P2) shows both amplified gestures and the activity of separate processes. It provides information on both the interface and processes of the instrument, without detailing its internal structure or behaviour.

MAPPINGS (I4-M1-P2) adds information pertaining to how sensors are mapped to the sound processes. It shows when a sensed gesture has an effect on a sound process but not what effect it has, i.e. not what is exactly controlled by each sensor. In our implementation, mappings are displayed as lines between sensors and processes, which appear when a control is performed and then fade out. It is similar to the level of information proposed in the Rouages project (Berthaut et al. 2013).

FULL_COMBINED (I4-M2-P3) refines both the Mappings and Processes sections. It shows which parameters are controlled by each sensor and displays both the parameters and activity of the processes. In our implementation, each process is represented by a composite shape with an outer ring displaying the input parameters (i.e. gain with size, filter cutoff with color luminance, position in sample with rotation, delay feedback with shape repetition, pitch with color hue), while the activity is shown by an inner graphical element. This level is similar to the augmentations described by Berthaut et al. (2015). As suggested by them, this LOD should improve the exclusivity dimension of attributed agency, by showing when a change in the sound actually comes from the musician and when it is automated.

FULL_GRAPH (I4-M2-P5) provides a complete overview of the instrument with parameters names and value range, processes names and mappings between each sensor and the parameters. It corresponds to the mental model musicians might have of their instrument, with the exact structure, mappings and range of sonic possibilities. In our implementation, each process is labelled and displayed as a group of graphical sliders and buttons representing each parameter, with their names, value and range of values, and another slider serves as a VU-meter. Although this *global LOD* uses the maximum of each *local LODs*, we chose to limit the Mappings section to level M2 so that the amount of information remains reasonable. Similarly, the structure of the instrument used in our study is essentially a stack of samplers and effects with one parameter each, so that level P5 adds very little information compared to level P4. This structure was chosen in order to reduce the gap in quantity of information from the previous *global LOD*, i.e. we do not add a complex audio graph in addition to the details on parameters when going from *FULL_COMB* to *FULL_GRAPH*. FULL_GRAPH can be seen as similar to approaches where the full complexity of the instrument is shown such as in live-coding performances.

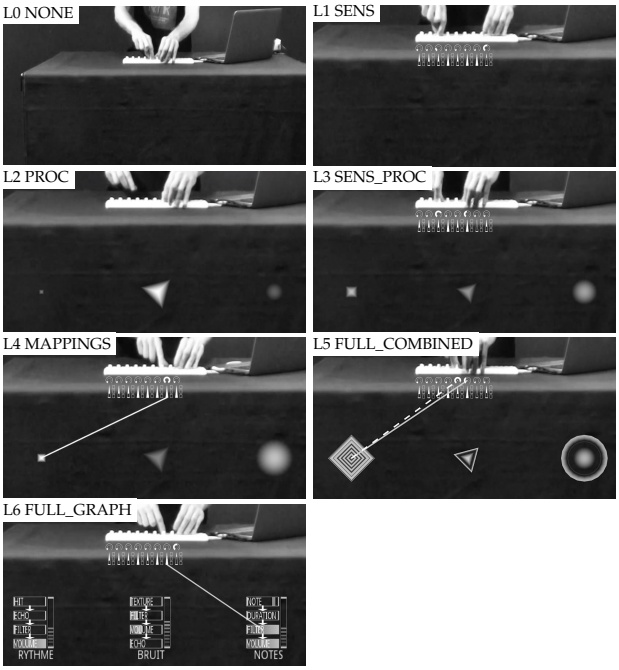


Figure 2. The 7 global *levels of detail (LODs) used in our experiment, as seen by the participants. Each is built as a combination of* local *LODs for the Interface, Mappings and Processes sections (details in and).*

Implementation

In order to provide such dynamic LODs to spectators, one needs to access internal parameters of DMIs and to adapt to various display strategies.

Accessing LODs

Depending on the chosen LODs, the granularity of information required by the system can increase rapidly as well as the real time processing of extracted data. For low LODs (i.e. less detail), data for both the *interface* and *processes* sections can easily be gathered by directly accessing messages (e.g. MIDI/OpenSoundControl) sent by the interface and by extracting audio features from the audio output of the instrument. In higher LODs where the mappings and internal structure of the *processes* need to be displayed, one must gain access to internal events and data of the software part of the

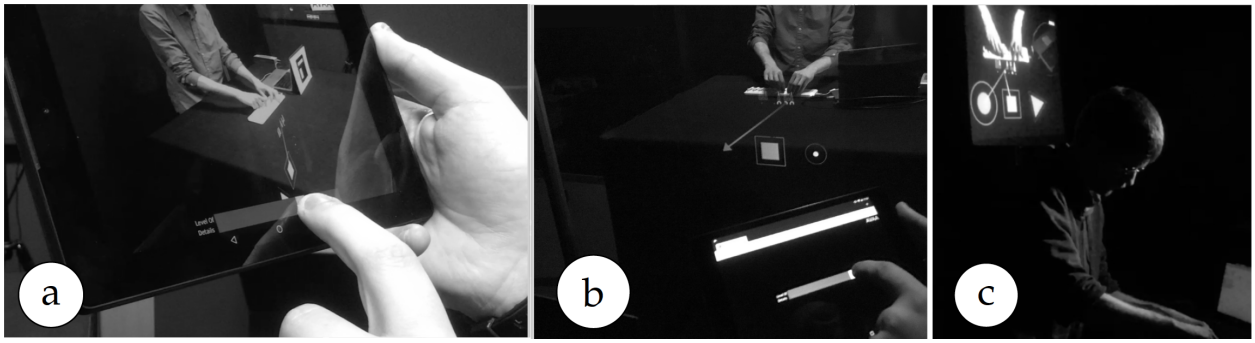


Figure 3. Possible implementations[1] of visual augmentations with LODs : a) Mobile AR with individual LOD control, b) Spatial AR with an optical combiner shared between all spectators and mobile control to vote for the LOD, c) Shared close-up with Video AR projected behind the musician and artist defined LOD.

system. The case of patch-based instruments or open-source software is the most convenient as it offers a deep access to all of the software components. The instrument used in this study is such a patch-based instrument. Digital Audio Workstations such as Ableton Live, which are used by many electronic musicians, might offer access to their control data through plugins, or in the case of Ableton live a dedicated API. They, however, do not guarantee a full access to every settings of the instrument, e.g. the set of mappings. In general, the use of visual augmentations has implications on the design of DMIs, which need to integrate a protocol for querying their structure and state and for listening to internal events.

Displaying LODs

Once the information on the instrument's structure, state and activity is captured and translated to the visual representations proposed in the previous section, it needs to be displayed for the audience in the form of visual augmentations overlapping the performance and instrument. We envision multiple possibilities for implementing visual augmentations with LODs in a performance setting. A first one relies on individual views of the augmentations, in order to allow each spectator to choose their LOD freely. This can be implemented with a mixed-reality headset or a mobile device as shown in Figure 3.a. In

our case, spectators access a web page with their mobile devices. Based on OpenCV and WebGL, it uses printed markers placed around the instrument to superimpose the augmentations on the mobile camera image in real time. Updates to the augmentations are received via WebSockets and a slider allows spectators to quickly explore the LODs.

To avoid forcing the audience to wear or hold devices which may impair their experience, another possibility is to use a single spatial AR display, either projection mapping or an optical combiner (e.g. Pepper's ghost display), such as depicted in Figure 3.b, in which case viewers all perceive the augmentations spatially aligned with the physical instrument. Another possibility is to film and re-project a close-up view of the interface integrating the augmentations, as shown in Figure 3.c. This solution however moves the focus away from the physical performer. In these scenarios, all the spectators share the same LOD. LOD control may be performed by musicians or accompanying visual artists, so that they can modulate the audience experience during the performance. But the shared LOD can also be chosen by spectators. Voting system such as the one used in the Open Symphony project (Wu et al. 2017) may be used, in the form of a web interface accessible from their mobile devices, as depicted in Figure 3.b. In this case the displayed LOD reflects either the majority or the average vote.

Finally, an intermediary solution is to provide multiple views of the augmentations for groups of spectators, using video (i.e. multiple or multiscopic screens such as the one proposed by Karnik et al. (2012)) or optical AR (with mirrors at multiple angles). For each group, the LOD can be fixed at a different value, so that spectators can move towards or look at the display they prefer. A voting system may also be setup separately for each group.

Usage and effects of LODs

In this section, we present an experiment that aims at evaluating the impact of LODs on audience experience and understanding, and studies the use of controllable LODs by spectators with different expertise.

In order to retrieve accurate and individual data on spectator experience we chose to conduct a controlled experiment in the lab. We discuss the advantages and limitations of such 'in the lab' studies in more details in (Capra et al. 2020a) and plan to address social and environmental aspects of public performances in a future work.

Hypotheses

From the literature analysis, we hypothesise that the different LODs, with their various amount and type of information, will effect the experience of participants, improving their understanding and experience up to a certain level but differently for novices and experts. We also hypothesise that, if given the choice, participants will select the LOD depending on their expertise with DMIs.

Procedure



Figure 4. During the experiment, participants watched videos of short performances with digital musical instruments, equipped with headphones and a lightweight eye tacking device.

18 participants (16 M, 2 F) took part in the experiment, aged of mean 29 (± 7.3 , min=20, max=43). As illustrated in Figure 4, before the beginning of the experiment, participants were presented with the details of the experiment and signed a consent form. Participants sat in front of a 24" screen, equipped with headphones and a Pupil-labs Core eye-tracking device (the details of the eye tracking are addressed in a forthcoming study). We measured their expertise with the instrument presented in the study using questions regarding their practice of DMIs, their use of graphical user interfaces similar to the one in Figure 2 and their use of control surfaces. We also asked how often they attended electronic music performances. This allowed us to compute an expertise score, and we used it to separate them into two groups : 9 experts and 9 novices. The experts had a music practice of 17.3 years (± 6.4, min=10, max=30) and an electronic music practice of 10.7 years (± 7.3, min=2, max=25) against 1.6 years (± 2.6, min=0, max=7) of music practice and no electronic music practice for the novices. Experts had all used both graphical interfaces for music and control surfaces such as the ones presented in the experiment. Per year, experts claimed going to 12.8 (± 8.3, min=2, max=30) electronic music performances, while for novices the average was 0.6 (± 1.5, min=0, max=5).

Dynamic stimuli

The stimuli were videos of short performances of a male musician playing with a Digital Musical Instrument. The DMI was composed of a Korg NanoKontrol controlling a set of Pure Data patches with three sound processes (melodic, rhythm, granular texture) each with multiple parameters (See Figure 2). We designed 3 sets of mappings between the interface sensors (knobs, faders, buttons) and the parameters. Each set was intended to target a different level of contribution of the musician, i.e how much of the changes in the sound are due to them vs automated. The first set is completely manual so no changes happen without a gesture. It corresponds to the maximum contribution level. The second features automations for half the parameters, the rest being manipulated by the musician.

In the third set of mappings, most parameters are automated and the musician is able to take control of some of them temporarily, giving the highest contribution to the computer.

In order to play the videos with dynamic overlapping visual augmentations, we designed the experiment in the Godot game engine. Videos were played synchronised with the playback of control data recorded in Pure Data, so that the sound and the visual augmentations were generated dynamically during the playback. This technical setup gave us the flexibility to play the video footage of a performance and to accompany it with arbitrary audio processes and visual augmentations in real time. The experiment lasted around 45mn and was composed of 2 blocks.

Block 1 : fixed LODs

In the first block, participants watched 7 LODs x 3 contribution levels = 21 videos of short performances (20s). Each video was followed by a questionnaire of 9 order-randomised questions to evaluate their experience and comprehension. The survey included only one objective question. We evaluated the ability of the participants to correctly detect the contribution levels that we induced by the mappings by answering the question "Who from the musician or the computer contributed the most to the performance ?". They also could choose 'both equally'.(The questions were all posed in French but appear in translation here.)

The other questions evaluated the participants' subjective comprehension. They were based on 5 communication design issues introduced by Bellotti et al. (2002) and transposed to the spectator perspective by Gurevich and Fyans (2011). We complemented them with *Association* that targets the capacity to expose to spectators the respective and shared contributions of the user (musician) and the system (DMI) (Capra et al. 2020a). These design challenges are well adapted to the evaluation of NIMEs as they allow for an assessment by components of the subjective experience of spectators.

By evaluating subjective comprehension, we do not target the objective ability to detect a parameter of the interactions like we do with the first question of the block. Instead, these questions aim at providing insights into the confidence spectators have in the inner representation of the interactions they build up along the performance. Participants answered on 7-step scales to the question *"To which extent do you agree with the following statement ?"*. Only the extreme values of the scales had a label : "I totally disagree" and "I totally agree".

"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's gesture and when it is not." (Attention)

"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 properly functioning and when it is not."(Alignment)

"In this video, I can see if either the musician or the instrument made a mistake."(Accident)

"In this video, I can see the contribution of the musician and the one of the computer."(Association)

Finally, the participants had to report their personal rating of the performer's virtuosity and the overall performance on a 7-point scale.

Block 2 : dynamic LODs

In the second block, participants could change with the scroll wheel the LOD of the augmentations as the video was playing. In a first task, they watched 3 short (60s) performances and were asked to select the LOD that gave them the best experience, i.e. that they preferred. In a second task, they watched the same performances and were

asked instead to choose the LOD that allowed them to understand best what the musician was doing.

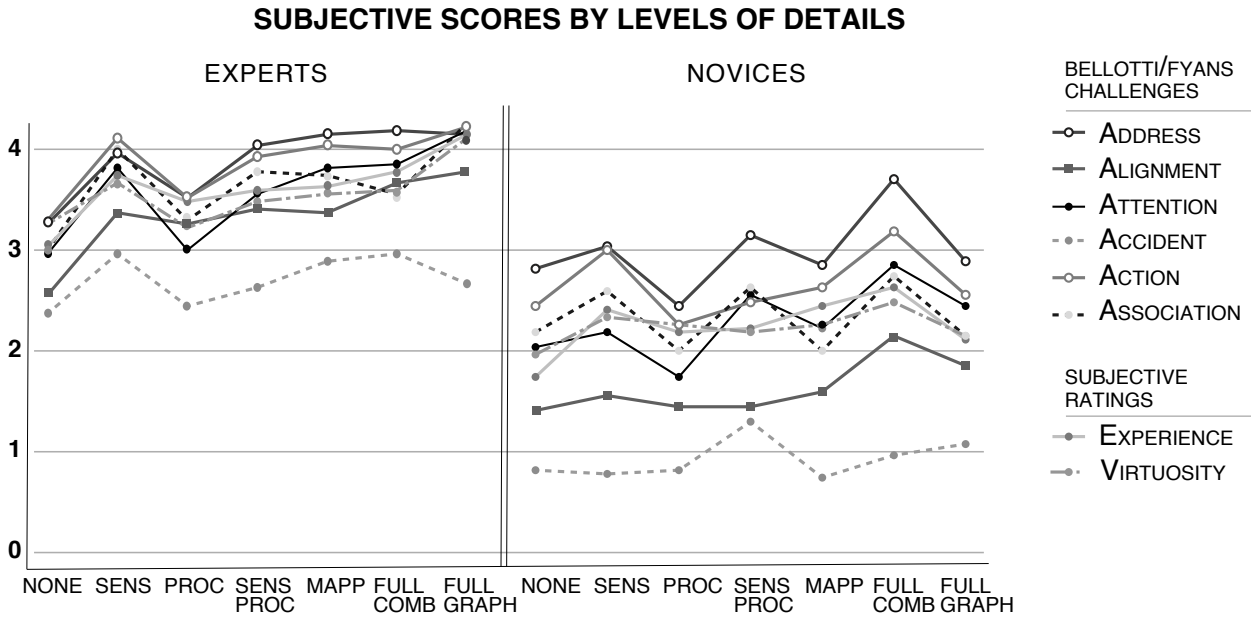


Figure 5. The levels of detail (LODs) of the visual augmentations did not impact equally the subjective perception of the interactions. Moreover, compared to novices, experts reported higher evaluations of the Bellotti/Fyans challenges (subjective comprehension) and higher ratings of their experience and the virtuosity of the musician.

Results

Data was recorded, anonymised and stored in real time during the experiment by a bespoke experiment software developed in the Godot game engine. Subjective reports were obtained via likert scales and were analysed with parametric tools when the normality assumptions were met.

Data analysis

The analyses were conducted under the common frequentist paradigm and were combined to Bayesian statistics (Kay et al. 2016). 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

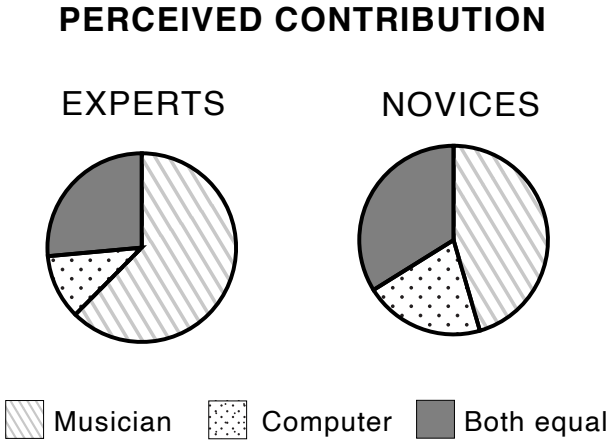


Figure 6. Regardless of the LOD, experts perceived a higher contribution of the musician than Novices.

that the data are 2.4 times more likely to occur under a model including the corresponding effect compared to the one implying no effect (H0). 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. Analyses were performed with SPSS v25, R studio 1.2 and JASP (JASP Team 2019).

Block 1 : fixed LODs

Contrary to our hypothesis on objective tasks, analysis did not reveal any group effect and any effect of the levels of detail (LODs) on the objective task. Overall, the evaluation of the factual contribution ratio between the musician and the computer proved difficult.

Our hypothesis on the Bellotti/Fyans challenges was confirmed. Thus, from a subjective perspective, an interesting group effect ($\chi^2 = 12, p = 0.002, BF_{10} = 11$) showed that experts considered the musician contributed more than the computer in 62% of the stimuli compared to 45.5% for Novices (Figure 6). As depicted in Figure 7, experts reported higher evaluations of the subjective questions but did not perform better than novices in the objective task of evaluating the most contributive to the performance from

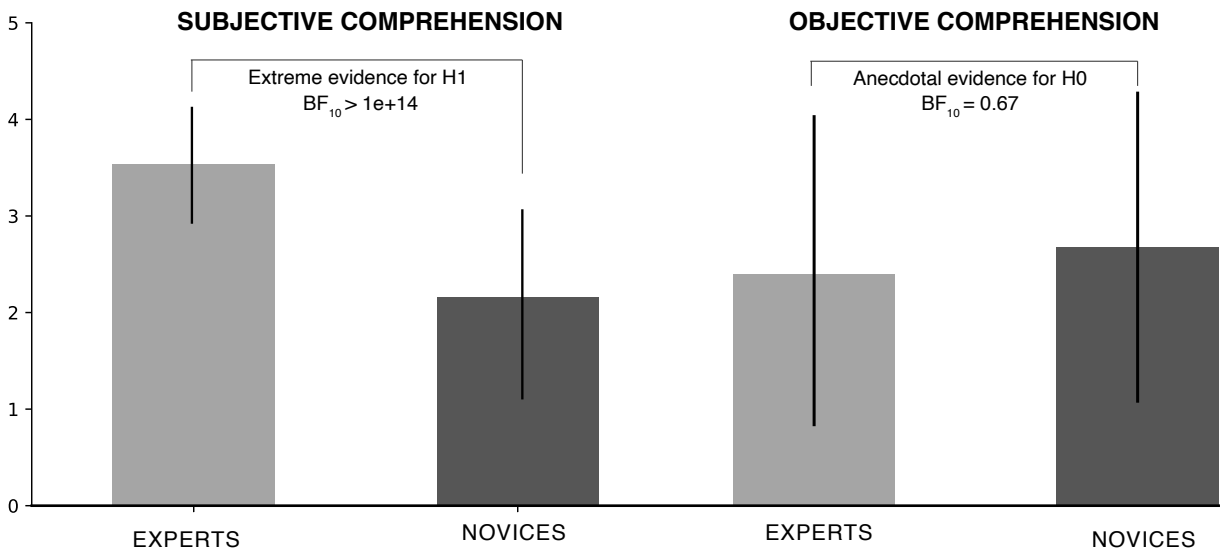


Figure 7. Even if experts have more trust in what they think they can perceive from the interactions (subjective comprehension), they do not perform over Novices in the evaluation of the objective contribution of the musician (objective comprehension).

the musician or the computer.

When detailing the Bellotti/Fyans challenges, regardless of the group, the *Accident* was the least rated, meaning that participants were not so confident in their capacity to detect errors. The effect of the LOD was revealed on most of the subjective questions ($all\ p-values < 0.027, all\ BF_{10} > 6$), with the exception of Accident and Virtuosity ($all\ p-values > 0.22, all\ BF_{01} > 4$). Two LODs were particularly effective, *SENS* and *FULL_COMB*.

Reading the experts graph (Figure 5 - left) from left to right, compared to *NONE*, the control condition, *SENS*, the level of detail exposing the sole sensors activity, presents a significant boost in all dimensions, then *PROC* exposes an equivalent score to *NONE*. From *SENS_PROC* to *FULL_COMB* a rather linear progression is observed and extends to *FULL_GRAPH*. On Figure 5 (right), in a much more volatile distribution, the results for the Novices group nevertheless present *FULL_COMB* as the most effective.

The efficiency of *FULL_COMB* for Novices is also supported by an analysis of the

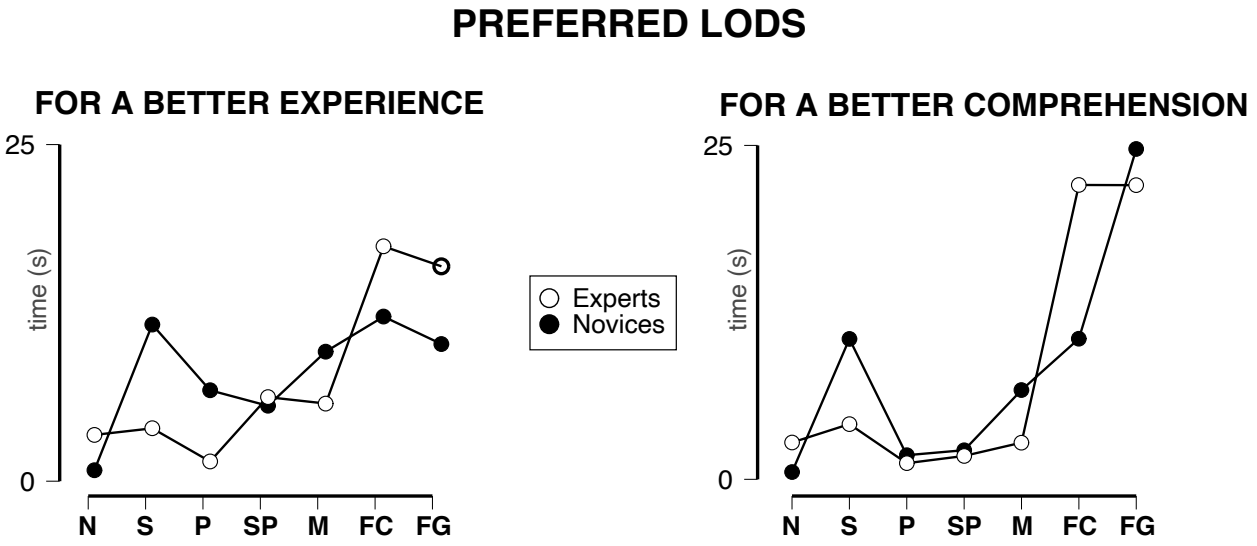


Figure 8. When participants could choose their favourite LODs in real time (they were asked to select it and remain on it), strategies emerged as illustrated. They confirmed and discussed their choices in the interviews.

difference with the experts' scores. For 6 (out of 9) dimensions, the smallest difference is measured when visual augmentations are presented with *FULL_COMB*. This result is a good illustration of the expected role of visual augmentations, compensate the lack of expertise of novices for a better experience.

Block 2 : dynamic LODs

The score for these tasks was calculated by accumulating the time participants spent using each LOD. Both tasks, experience and comprehension, show comparable evolution characterised by a minimum for the control condition *NONE* and a maximum for the higher LODs (Figure 8). A decisive effect of LODs was found ($F_{(6,90)} = 9.94, p < .001, BF_{10} > 10000$) but with no difference between the groups ($BF_{01} = 4$). Novices favoured *FULL_COMB* and *SENS* for experience and *FULL_GRAPH* for comprehension. Experts chose the highest LODs for experience and *FULL_COMB* and *FULL_GRAPH* for comprehension.

Discussion

In this section, we complete the quantitative data with subjective insights from interviews and we discuss our results.

LODs affect subjective comprehension

The interviews confirmed and extended the quantitative analyses. Despite the absence of effect of LODs on the ability of spectators to perceive the musician objective contribution (objective comprehension), participants favoured levels *FULL_GRAPH* and *SENSORS* for understanding the performance, especially when the music got more complex with many fast changes in the sound. This indicates that LODs influence the subjective comprehension of spectators in the sense that spectators feel more confident in what they perceive from the interactions even if their factual understanding is not improved. It also suggests that amplifying the gestures (*SENSORS* level) might be more informative than displaying the activity of processes alone (*PROC* level).

The role of expertise

Our study reveals interesting insights into the nature of expertise in DMI spectators. Results of Block 1 showed that experts perceive a higher contribution of the musician when novices perceive a higher contribution of the computer. Also, experts put more trust in their personal representation of the interactions as proven by their higher evaluation of the Bellotti-Fyans challenges (subjective comprehension). This contrast is confirmed in Block 2 where only novices favoured the *SENSORS* LOD over no augmentations for a better comprehension and experience (Fig. 8), as if experts already had an internal representation of the interactions with the sensors and therefore did not need that LOD. Apart from *SENSORS*, both experts and novices mostly utilised *FULL_COMB* when they could choose their favourite LOD. But when they had to choose a LOD in order to better

understand the interactions, experts equally used *FULL_COMB* and *FULL_GRAPH* when novices massively favoured *FULL_GRAPH*. As both groups scored poorly in the objective task in Block 1, whatever the LOD, these preferences in LOD are to be taken as subjective beliefs in a facilitation of understanding rather than a factual help.

In a previous study (Capra et al. 2020a), we already underlined the selective impact of visual augmentations on the sole subjective comprehension. Additionally, we showed that when participants watch DMI performances with visual augmentations, they overestimate the musician's contribution from the computer's, as the experts do when compared to novices in this study. Thus, multiple evidence support the idea of a rather subjective nature of expertise. The way we evaluate objective comprehension surely holds limitations. However, increasing data supports the hypothesis that, from a spectator perspective, experts are experts because they *feel* like they are, not due to a better perception compared to novices. In our results, experts do not show a superior ability in the understanding of interaction, on the contrary, we saw that their perception is biased towards a greater contribution of the musicians over automated processes.

Thus, by biasing the spectator perception towards a greater contribution of the musician (Capra et al. 2020a), and by comforting the confidence of spectators in their representation of the interactions (subjective comprehension), visual augmentations level up novices toward the rank of experts, especially when novices can select their favourite LODs.

Errors and virtuosity

The absence of effect of LODs on both the *Accident* dimension (i.e. the feeling of being able to perceive a potential error) and the virtuosity ratings underlines the crucial role of error perception in the emergence of a judgement of virtuosity (Gurevich and Fyans 2011). A solution to this issue could be inspired by music video games where the virtuosity is materialised by screen indications of combinations (combos) of successful moves. Such

informative contents are efficient and spectacular but imply the restriction of any improvisation or non-expected techniques. Another solution would be to design LODs that inform virtuosity, such as visualisations of input complexity or extra-ordinary values for controls and musical parameters.

LOD choice strategies

Strong differences in the choice of favoured LODs at the individual level were revealed by the data and refined by the interviews. When analysing the answers of participants regarding how they would use the LODs in public performance, we can distinguish 3 clear strategies: **all or (almost) nothing:** 4 participants claimed they would alternate between the maximum LOD (or just start with it) in order to form a mental image of how the instrument works (i.e. its capabilities) and then go back to no augmentations or to the *SENSORS* level, in order to focus on the musician's gestures. **adapting to complexity / performance:** 4 participants claimed they would use LODs as a way to adapt to the complexity of the instrument or music, or change it depending on the musician playing; **progression:** 2 participants mentioned that their appreciation of LODs evolved over time, the more complex ones becoming more enjoyable and accessible, so that they would end up not going back to the lower LODs. One must note than even within these strategies there are interpersonal variations, again highlighting the utility of a controllable LOD on visual augmentations.

Mediation through LODs : The role of the Augmenter

A part of this work is dedicated to finding solutions to make the audience feel more aware of what is going on on stage during digital music performances. The extra mediation of visual augmentations to make more transparent interactions already mediated by technology may seem redundant. One may ask, why not explore a mediation that could suit both the musicians and the audience ? This question may find an answer

thanks to cumulative data gathered about the subjective comprehension. With this study, the idea that spectators are more influenced by their inner representations than the actual objective reality of an interaction was strengthened. While this potential mismatch between perception and reality is a common phenomenon well known to illusionists and neuroscientists, here we have the possibility to infer the role of the augmentations in the constitution of more reliable inner representations.

In the diversity of potential representations of a digital musical interaction, the ones expert observers can build should have a greater similarity with the ones from the musicians. As we saw it, from our data, it seems that they do not. In fact, the role of visual augmentations, and Spectator Experience Augmentation Techniques (SEATs) in general Capra et al. (2020a), may not be only to make the interactions more objectively understandable. They should not leave behind this role of a facilitation of the objective comprehension, but they should especially embed cues that contribute to the subjective comprehension, even if these cues are contradictory with objective cues. Besides, LODs are an effective way to offer rather balanced and customised information to spectators, preserving them from potential cognitive overload such as in fully descriptive visual augmentations.

To sum up, techniques to augment the experience of spectators should deliver a subtle ratio of objective and subjective cues and should also consider the audience's direct reactions. Such a sensitive role is not a purely technical role any more. It requires integrating a lot of information and to "feel" what should be the proper way to represent the ongoing interactions. For these reasons we think there is place on stage for one more artist, *the augmenter.*

The augmenter could act as an augmentations conductor and compose with the direct inputs from the musicians' instrument and connect them to visual augmentations, while selecting LODs to emphasise parts of the interactions. On the contrary the augmenter

could leave the mystery on some parts, or even disturb the audience perception on purpose with disruptive augmentations. As many artistic activities, the augmenter would require training to reach the level of precision and virtuosity to personify (per-sonify ?) the artistic intentions of the musicians. Compared to the VJs whose role is to illustrate the music with rather exclusively graphical considerations, the augmenter would perform as a human mediation between digital systems and human agents to reveal the virtuosity of the musicians and the expressiveness of instruments.

Conclusion

In this paper, we introduce the concept of Levels of Detail (LODs) in visual augmentations for the audience of Digital Musical Performances. We designed and implemented these LODs and we investigated their impact on expert and novice spectators.

The data we obtained from a controlled experiment show that whatever the LOD we used, the objective ability of spectators to perceive components of the interactions of musicians with their Digital Musical Instrument (DMI) remains relatively low, with no measurable difference between novices and experts. In particular, we found that the latter overestimate the contribution of the musician compared to the one of automated processes. Beside this newly identified bias in favour of the musician's involvement, experts and novices are only distinguished by their subjective comprehension of the interactions, i.e. what they think they understand of the interactions rather than what they do (objectively) understand. These results lead us to hypothesise a rather subjective nature of the expertise, from a spectator perspective.

Regarding the Levels of Detail in visual augmentations, our study revealed their impact, once again on the sole subjective aspects of the spectator experience. From quantitative data, we identified the most effective LODs with respect to the expertise of

observers and analysed their respective strategies during guided interviews. Our experimental approach suggests that by comforting the confidence of spectators in their representation of the interactions (subjective comprehension), visual augmentations are a particularly effective way to level up novices toward the rank of experts, especially when novices can select their favourite LODs.

Finally, in order to cope with the many challenges of the mediation between musicians and audience, we propose a new role in the digital musical performance ecosystem, the *augmenter*, who can manipulate the augmentations and their LODs during performances.

While our results provide useful insights, we believe the controlled experiment approach that we took should be combined with *in the wild* study of performances. As future work, we think that augmentations with LODs should be extended to other interfaces beyond control surfaces, e.g. gestural controllers or graphical interfaces such as live-coding, and that the effect of aesthetic choices on the design of augmentations should be investigated.

References

Barbosa, J., F. Calegario, V. Teichrieb, G. Ramalho, and G. Cabral. 2013. "Illusio: A drawing-based digital music instrument." In *Proceedings of NIME*.

Barbosa, J., F. Calegario, V. Teichrieb, G. Ramalho, and P. McGlynn. 2012. "Considering Audience's View Towards an Evaluation Methodology for Digital Musical Instruments." In *Proceedings of the International Conference on New Interfaces for Musical Expression*. URL http://www.nime.org/proceedings/2012/nime2012_174.pdf.

Bederson, B. B., and J. D. Hollan. 1994. "Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics." In *Proceedings of the 7th Annual ACM Symposium*

on User Interface Software and Technology, UIST '94, pp. 17–26. URL https://doi.org/10.1145/192426.192435.

Bellotti, V., M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. 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*, CHI '02, pp. 415–422. URL https://doi.org/10.1145/503376.503450.

Benford, S., C. Greenhalgh, A. Hazzard, A. Chamberlain, M. Kallionpää, D. M. Weigl, K. R. Page, and M. Lin. 2018. "Designing the Audience Journey through Repeated Experiences." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–12. URL https://doi.org/10.1145/3173574.3174142.

Berthaut, F., D. Coyle, J. Moore, and H. Limerick. 2015. "Liveness Through the Lens of Agency and Causality." In E. Berdahl and J. Allison, eds. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 382–386. URL http://www.nime.org/proceedings/2015/nime2015_272.pdf.

Berthaut, F., M. T. Marshall, S. Subramanian, and M. Hachet. 2013. "Rouages: Revealing the Mechanisms of Digital Musical Instruments to the Audience." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 164–169. URL http://nime.org/proceedings/2013/nime2013_51.pdf.

Bin, S. A., N. Bryan-Kinns, and A. P. McPherson. 2016. "Skip the Pre-Concert Demo: How Technical Familiarity and Musical Style Affect Audience Response." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, vol. 16, pp. 200–205. URL http://www.nime.org/proceedings/2016/nime2016_paper0041.pdf.

Capra, O., F. Berthaut, and L. Grisoni. 2018. "Toward Augmented Familiarity of the Audience with Digital Musical Instruments." In M. Aramaki, M. E. P. Davies, R. Kronland-Martinet, and S. Ystad, eds. *Music Technology with Swing*, pp. 558–573.

Capra, O., F. Berthaut, and L. Grisoni. 2020a. "Have a SEAT on Stage: Restoring Trust with Spectator Experience Augmentation Techniques." In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, DIS '20, pp. 695–707. URL https://doi.org/10.1145/3357236.3395492.

Capra, O., F. Berthaut, and L. Grisoni. 2020b. "A Taxonomy of Spectator Experience Augmentation Techniques." In R. Michon and F. Schroeder, eds. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 327–330. URL https://www.nime.org/proceedings/2020/nime2020_paper63.pdf.

Correia, N. N., D. Castro, and A. 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*, AM '17. URL https://doi.org/10.1145/3123514.3123555.

Correia, N. N., and A. Tanaka. 2017. "AVUI: Designing a Toolkit for Audiovisual Interfaces." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 1093–1104. URL https://doi.org/10.1145/3025453.3026042.

Fels, S., A. Gadd, and A. Mulder. 2002. "Mapping transparency through metaphor: towards more expressive musical instruments." *Organised Sound* 7(2):109–126.

Gurevich, M., and A. C. Fyans. 2011. "Digital musical interactions: Performer–system relationships and their perception by spectators." *Organised Sound* 16(2):166–175.

Holten, D. 2006. "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data." *IEEE Transactions on visualization and computer graphics* 12(5):741–748.

JASP Team. 2019. "JASP[Computer software]." URL https://jasp-stats.org/.

Jordà, S. 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*, pp. 6–9.

Karnik, A., W. Mayol-Cuevas, and S. Subramanian. 2012. "MUSTARD: A Multi User See through AR Display." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 2541–2550. URL https://doi.org/10.1145/2207676.2208641.

Kay, M., G. L. Nelson, and E. 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*, CHI '16, pp. 4521–4532. URL https://doi.org/10.1145/2858036.2858465.

Leman, M., et al. 2008. *Embodied music cognition and mediation technology*. MIT press.

Luebke, D., M. Reddy, J. D. Cohen, A. Varshney, B. Watson, and R. Huebner. 2003. *Level of detail for 3D graphics*. Morgan Kaufmann.

Miranda, E. R., and M. M. Wanderley. 2006. *New digital musical instruments: control and interaction beyond the keyboard*, vol. 21. AR Editions, Inc.

Perrotin, O., and C. d'Alessandro. 2014. "Visualizing Gestures in the Control of a Digital Musical Instrument." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 605–608. URL http://www.nime.org/proceedings/2014/nime2014_406.pdf.

Reeves, S., S. Benford, C. O'Malley, and M. Fraser. 2005. "Designing the Spectator Experience." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pp. 741–750. URL https://doi.org/10.1145/1054972.1055074.

Schloss, W. A. 2003. "Using contemporary technology in live performance: The dilemma of the performer." *Journal of New Music Research* 32(3).

Stollnitz, E. J., T. D. Derose, and D. H. Salesin. 1996. *Wavelets for Computer Graphics: Theory and Applications*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Sung, M., Y. Choi, H. Ko, and J. Hwang. 2014. "Level-of-detail AR: Managing points of interest for attentive augmented reality." In *2014 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 351–352.

Turchet, L., and M. Barthet. 2019. "Haptification of Performer's Control Gestures in Live Electronic Music Performance." In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, AM'19, pp. 244–247. URL https://doi.org/10.1145/3356590.3356629.

Wang, W., H. Wang, G. Dai, and H. Wang. 2006. "Visualization of Large Hierarchical Data by Circle Packing." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pp. 517–520. URL https://doi.org/10.1145/1124772.1124851.

Wu, Y., L. Zhang, N. Bryan-Kinns, and M. Barthet. 2017. "Open symphony: Creative participation for audiences of live music performances." *IEEE MultiMedia* .