Levels of Detail in Visual Augmentation for Novice and Expert Audiences

Olivier Capra,* Florent Berthaut,† and Laurent Grisoni†

*ETHICS EA 7446 - Anthropo-Lab, 14 boulevard Vauban, CS 40109, 59016 Lille cedex, France
†Université Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
olivier.capra@univ-catholille.fr, {florent.berthaut, laurent.grisoni}@univ-lille.fr

Abstract: Digital musical instruments offer countless opportunities for musical expression by allowing artists to produce sound without the physical constraints of analog instruments. By breaking the intuitive link between gestures and sound they may hinder the audience experience, however, making the musician’s contribution and expressiveness difficult to discern. 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 article, we introduce the concept of controllable levels of details (LODs) for visual augmentation. 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 analyze quantitative and qualitative data to reveal contrasts in the impact of LODs on experience and comprehension for experts and novices. Finally, we highlight the diversity of usage of LODs in visual augmentation by spectators and propose a new role on stage, the augmenter.

Digital musical instruments (DMIs) offer countless opportunities for musical expression by allowing 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 hinder the audience experience, however, making the musician’s contribution and expressiveness difficult to perceive. Consequently, it also contributes to degrading spectators’ 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 with 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 DMIs, either as an evaluation method [Barbosa et al. 2012], or as a dimension that should be addressed at the design [Fels, Gadd, and Mulder 2002; Jordà 2003; Correia and Tanaka 2017], at the performance stages [Reeves et al. 2005; Benford et al. 2018]. Artists and researchers alike have designed techniques that augment the instruments with additional information to improve the audience experience and restore the trust of spectators in the musician’s involvement. Although these techniques explore different modalities (visual, haptic, or auditory) and address different aspects of the performance (technical, gestural, or intentional), they mostly offer a fixed level of information to all spectators.

We think that augmenting the audience experience can be more effective when considering spectators from an individual perspective, however. The information needed by each spectator can differ depending on personal sensitivity and expertise. To ensure an optimal experience for spectators, we propose allowing them to dynamically change this level of information individually, using visual augmentation with variable levels of detail (LODs).
Augmenting the Audience Experience

A number of augmentation techniques for spectator experience have been designed. We only provide a few examples here, an analysis in greater detail can be found in the taxonomy that we presented in a paper published at the 2020 International Conference on New Interfaces for Musical Expression (Capra, Berthaut, and Grisoni 2020b).

Perhaps the simplest technique is the organization of preconcert demonstrations, such as that described by Bin, Bryan-Kinns, and McPherson (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 investigated displaying the musical controls used by musicians in an orchestra by means of a video projection to help the audience become aware of the actions of each orchestra member, by representing both gestures and musical parameters. Similarly, Correia, Castro, and Tanaka (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 that can be used to reveal the mechanisms of DMIs. Haptic augmentation can also be created to increase the audience’s engagement, as proposed by Turchet and Barthet (2019). All these augmentation techniques, however, offer only a fixed set of information for the whole audience.

Benford et al. (2018) go beyond fixed information by combining projected visual augmentation during the performance, as well as visual and textual augmentation on a mobile app after the performance, thus allowing spectators to access two different levels of representation. Finally, Capra, Berthaut, and Grisoni (2018) propose adaptive augmentation 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 work with adapting the amount and the type of information provided by visual augmentation using an LOD approach.

Contribution

In this article, we first introduce the concept of LODs for visual augmentation, 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 (Capra, Berthaut, and Grisoni 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.

LODs for the Visual Augmentation of DMIs

The concept of LOD originates from the field of computer graphics (Luebke et al. 2003), in which 3-D models and scene complexity are adapted to reduce rendering load. It takes inspiration from existing signal analysis tools, such as wavelets (Stollnitz, Derose, and Salesin 1996) or more basic simplification systems such as downsampling. LODs are meant to bring some flexibility in terms of computational cost in all possible aspects of 3-D 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 visualization (expectations of users, hardware possibilities, etc.).

In the literature of human–computer interaction, 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 by building and manipulating complex musical structures (Barbosa et al. 2013). Finally, LODs have also been 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 visualization to adapt
quantity of information in order to limit visual overload (Holten 2006; Wang et al. 2006).

Practical existing application 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 cannot be directly represented by any sampling set), in which case LODs may be achieved by drawing inspiration from other communities, such as data visualization and human–computer interaction.

LODs Applied to Augmentation for the Audience

In this article, we apply the LOD approach to the design of augmentation for the audience of digital musical performances. As discussed above, digital musical interactions can prove highly difficult to perceive and to understand, owing to potentially small or hidden sensors and gestures, to potentially complex mappings between sound and gestures, and to complex and partly autonomous [pre-defined or automated] processes of sound generation. Augmentation proposed in the literature aims at compensating for 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 the information from an 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 sensors, and all sound parameters.

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

1. too much information provided at once;
2. information requiring expertise in DMIs to be understood; or
3. differing preferences within the audience, which might range from trying to understand musicians’ 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 augmentation. The LOD approach can help adapt augmentation techniques—in our case visual augmentation—to the variety of expertise levels and preferences of spectators. In the notion of a better adaptation to personal needs, LODs for augmentation could be chosen dynamically by the spectators during the performance, either individually or in groups.

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

LODs in Visual Augmentation

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

Following proposals forwarded by Berthaut et al. (2013), our visual augmentation represent the three main components of the instrument:

1. the physical interface composed of sensors (e.g., a MIDI control surface);
2. the mappings, that is, the connections between sensors and musical parameters (e.g., the first fader controls the volume of the first audio track); and
3. the processes (e.g., tracks, loops, or patterns) that generate the sound.
Figure 1. Visual augmentation is organized as graphic representations of specific components of the interactions with a digital musical instrument: Input (a), Mapping (b), and Output (c). For each component, augmentation can be delivered with gradient levels of detail (LODs) from very basic or few, to finely detailed (circled numbers from 1 to 5). LODs dedicated to a component are called local LODs and can be combined into a set called the Global LOD. Finally, an overview of the experiment and their respective sets of local LODs (d).

An important aspect of visual augmentation is that it does 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 of the musicians’ actions (Fels, Gadd, and Mulder 2002) or the familiarity of the audience with the instrument (Gurevich and Fyans 2011), since these aspects are handled by the augmentation.

The potential complexity of DMIs implies, however, that visual augmentation 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 to alter the audience experience during the performance, for example, to change from expressive to “magical” interfaces Reeves et al. (2005).

We implement LODs by defining dedicated LODs for each of three components of the visual augmentation: interface, mappings, and processes. These local LODs can be chosen independently, or they can be combined, as we describe later in the Global LODs section.

**Local LODs**

As illustrated in Figure 1, we work with four levels of detail for the Interface component, three levels for the Mappings component, and five levels for the Processes component. Each local LOD includes a Level 0 in which the component is not augmented. If all three components are at Level 0, no information is added to the performance. One should note that the information provided by each level can be displayed in different ways, the representation used in our implementation is only one of many possibilities that artists can explore.

**Interface Component**

Over and above the Level 0, with no information about the interface component added to the visualization, we use the following levels:

1. Indication only of the global activity, e.g., when the musician performs a gesture sensed by the system.
2. Representation of the activity of each sensor of the physical interface, allowing the audience to follow fast and complex gestures such as bimanual or multifinger interactions.
3. Description of both the activity and the type of each sensor (discrete versus continuous, shape of sensor, etc.).
4. In addition to the above, a representation of sensor values and ranges is included.

Mappings Component

Again, beyond a Level 0, the visualization levels are:

1. Indication only of the processes to which sensors are connected.
2. Refinement of the previous level by representing the multiple parameters modified through the mapping. For example, the link showing a mapping connection can change color (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).
3. Further refinement by adding a representation of the operation or series of operations that transform sensor values into parameters values, for example, scaling, inverting, combining, and so on [cf. Fels, Gadd, and Mulder 2002].

Processes Component

1. Visualization of the system output as a whole, merging the activity of all sound processes.
2. Addition of detailed activity for each system process. For example, a shape’s size might indicate the volume of the corresponding sound process.
3. Addition of a dynamic representation of parameters (i.e., inputs) that can be controlled by a given process.
4. Parameter names, types, and values ranges are added. That is, the audience can see what performers would see while performing with a GUI.
5. Detailed representation of the complete internal graph of audio synthesis and effects that generate the sound of each process. This corresponds to what the musician would access when designing an instrument, and is potentially similar to the mental model used in performance.

Global LODs

Although the local LOD could be chosen by spectators independently for each component, we believe a simpler solution is to define a number of global levels as presets of local LODs. By combining local LODs, global LODs provide spectators with a convenient way to control LOD by modifying several components at a time. The components are indicated by their initials, so for instance, “Sensors [I4-M0-P0]” is a global LOD called “Sensors” and uses Level 4 for the Interface component and Level 0 for the others.

In the following study, we use seven global LODs with increasing quantities of information (see Figure 2).

1. None [I0-M0-P0] provides no information at all. The performance remains unchanged.
2. Sensors [I4-M0-P0] amplifies the gestures performed by displaying representations of the types and values for all sensors of the interface.
3. Proc [I0-M0-P2] displays the sound processes of the instrument as separate shapes in which graphical parameters associated with extracted audio features are shown (e.g., loudness with size, pitch with color hue, or brightness with color luminance).
4. 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 the internal structure or behavior of either.
5. 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, without going so far as to reveal what is exactly controlled by each sensor.
Each is built as a combination of local LODs for the Interface, Mapping, and Process components (details in the Local LODs and Global LODs sections).
this global LOD uses the maximum level for of each of the local LODs, we chose to limit the Mappings component to Level 2 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 Processes Level 5 adds little information compared with Level 4. This structure was chosen to reduce the gap in quantity of information from the previous global LOD. That is, we do not add a complex audio graph in addition to the details on parameters when going from Full Comb to Full Graph. The latter LOD can be seen as similar to approaches in which the full complexity of an instrument is shown, such as in live coding performances.

Implementation

To provide dynamic LODs, such as those we have discussed, to spectators, it is necessary to access internal parameters of DMIs and to adapt to various display strategies.

Accessing LODs

Depending on the LODs chosen, 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., those with fewer details], data for both the interface and processes components can easily be gathered by directly accessing messages sent by the interface (e.g., MIDI or OpenSoundControl) and by extracting audio features from the audio output of the instrument. With higher LODs, in which the mappings and internal structure of the processes need to be displayed, one must gain access to internal events and data of the software used in the system. The cases of patch-based instruments and open-source software are the most convenient, as they offer deep access to all 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, through a dedicated API. They do not guarantee, however, full access to every setting of the instrument, e.g., the set of mappings. In general, the use of visual augmentation 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 information regarding the instrument’s structure, state, and activity is captured and translated to the visual representations outlined in the previous section, it needs to be displayed to the audience in the form of visual augmentation overlapping the performance and instrument. We envision multiple possibilities for implementing visual augmentation with LODs in a performance setting. A first possibility relies on individual views of the augmentation, allowing each spectator to choose an LOD freely. This can be implemented with a mixed-reality headset or a mobile device, as shown in Figure 3a. In our case, spectators access a web page using their personal mobile devices. Based on OpenCV and WebGL, the Web page uses printed markers placed around the instrument to superimpose the augmentation on the mobile camera image in real time. Updates to the augmentation are received via WebSockets and a slider allows spectators to quickly explore the LODs.

To avoid forcing the audience to wear or hold devices that may impair their experience, another possibility is to use a single spatial augmented-reality (AR) display, either a projection mapping or an optical combiner (e.g., Pepper’s ghost display), such as that depicted in Figure 3b, in which case viewers all perceive the augmentation spatially aligned with the physical instrument. Another possibility is to film and reproject a close-up view of the interface integrating the augmentation, as shown in Figure 3c. This solution moves the focus away from the physical performer, however. In these scenarios, all spectators share one LOD. The performing musicians or accompanying visual artists may control the LOD, “modulating” the audience experience during the performance. But the shared LOD can also be chosen by spectators.
Figure 3. Possible implementations of visual augmentation with LODs, using a variety of augmented reality (AR) approaches: mobile AR with individual LOD control (a), spatial AR with an optical combiner shared between all spectators and mobile control to vote for the LOD (b), and shared close-up with Video AR projected behind the musician and artist defined LOD (c). All stimuli, illustration videos of the conditions, anonymized raw results, statistical analyses, and implementation demos can be found at http://o0c.eu/0NA.

A 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 the audience’s mobile devices [see Figure 3b]. 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 augmentation for groups of spectators, using video (i.e., multiple or multiscopic screens, such as one used by Karnik, Mayol-Cuevas, and Subramanian 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 set up separately for each group.

Usage and Effects of LODs

We now 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 levels of expertise.

To retrieve accurate and individual data on spectator experience, we chose to conduct a controlled experiment in laboratory conditions. We discuss the advantages and limitations of such “in-the-lab” studies in greater detail elsewhere [Capra, Berthaut, and Grisoni 2020a], and we plan to address social and environmental aspects of public performances in future work.

Hypotheses

From our literature analysis, we hypothesize that the different LODs, with their various amounts and types of information, will affect audience experience, improving the audience’s understanding and experience up to a certain level, but differently for novices and experts. We also hypothesize that, if given the choice, participants will select the LOD depending on their expertise with DMIs.

Procedure

There were 18 participants (16 men and 2 women) taking part in the experiment, with a mean age of 29 years (standard deviation [std. dev.] 7.3, ranging from 20 to 43 years). 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-in screen, equipped with headphones and a Pupil Core eye-tracking device [the details of the eye tracking are addressed in a forthcoming study]. We measured participants’ 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 the participants into two groups: nine “experts” and
nine “novices.” The experts had general music practice of 17.3 years (std. dev. 6.4, ranging from 10 to 30 years) and electronic music practice of 10.7 years (std. dev. 7.3, ranging from 2 to 25 years) against 1.6 years (std. dev. 2.6, ranging from 0 to 7 years) 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. The average number of concerts with electronic music performance attended annually by experts was 12.8 (std. dev. 8.3, ranging from 2 to 30), whereas for novices the annual average was 0.6 (std. dev. 1.5, ranging from 0 to 5).

Dynamic Stimuli

The stimuli were videos of short performances of a male musician playing with a DMI. 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 (as shown in Figure 2). We designed three sets of mappings between interface sensors (knobs, faders, and buttons) and parameters. Each set was intended to target a different level of contribution from the musician—that is, how much of the change in sound was due to the musician rather than automated changes. The first set was completely manual, so no changes occurred without a gesture. This corresponds to the maximum contribution level. The second set featured automation of half of the parameters, the rest being manipulated by the musician. In the third set of mappings, most parameters were automated, but the musician was able to take control of some of them temporarily, giving the highest contribution to the computer.

To play the videos with dynamic overlapping visual augmentation, we designed the experiment in the Godot game engine. Videos were synchronized with the playback of control data recorded in Pure Data, so that the sound and the visual augmentation 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 augmentation in real time. The experiment lasted around 45 min and was composed of two blocks.

Block 1: Fixed LODs

In the first block, participants watched seven LODs with three contribution levels each, for a total of 21 videos of short performances (20 sec each). Each video was followed by a questionnaire, consisting of nine questions in random order, to evaluate participants’ 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.)

With the other questions, we evaluated the participants’ subjective comprehension. Doing so, we do not target the objective ability to detect a parameter of the interactions, as 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. They are based on five communication design challenges—Address, Attention, Action, Alignment, Accident—introduced by Bellotti et al. (2002) and transposed to the spectator perspective by Gurevich and Fyans.
We complement them with an additional design challenge we call Association, which targets the capacity to expose the respective and shared contributions of the user (musician) and the system (DMI) to spectators (Capra, Berthaut, and Grisoni 2020a). Participants answered on seven-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.”

The questions, with associated design challenges, were:

1. Address: “In this video, I know when the musician is interacting with the instrument and when not.”
2. Attention: “In this video, I can see when the instrument is responding to the musician’s gesture and when it is not.”
3. Action: “In this video, I can see if the musician is controlling the instrument or if not.”
4. Alignment: “In this video, I can see when the instrument is properly functioning and when it is not.”
5. Accident: “In this video, I can see if either the musician or the instrument made a mistake.”
6. Association: “In this video, I can see the contribution of the musician and that of the computer.”

These design challenges are well adapted to the evaluation of New Interfaces for Musical Expression as they allow for an assessment by components of spectators’ subjective experience.

**Results**

Data were recorded, anonymized, and stored in real time during the experiment by custom software developed with the Godot game engine. Subjective reports were obtained via Likert scales and analyzed with parametric tools when assumptions of normality were met.

**Data Analysis**

The analyses were conducted under the common frequentist paradigm and were combined to Bayesian statistics [Kay, Nelson, and Hekler 2016]. A Bayes factor is reported as $BF_{01}$ when data better support the null hypothesis and as $BF_{10}$ when data better support the alternative hypothesis [note that the subscript ‘01’ becomes ‘10’]. For example, the statement $BF_{10} = 2.4$ means 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 the prior probability that the null hypothesis holds across all comparisons to 0.5. Analyses were performed with SPSS (version 25), the RStudio IDE (version 1.2), and JASP [https://jasp-stats.org].

**Block 2: Dynamic LODs**

In the second block, participants could change the LOD of the augmentation with a scroll wheel while the video was playing. In a first task, they watched three short, 60-sec performances and were asked to select the LOD that gave the best experience, i.e., which performance they preferred. In a second task, they watched the same performances and were asked instead to choose the LOD that allowed them to best understand what the musician was doing.

Contrary to our hypothesis on objective tasks, analysis did not reveal any group effect and any effect of the 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 percent of the stimuli compared with 45.5 percent for novices (see Figure 5). As depicted in Figure 6, 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 the musician or the computer.
Regardless of the LOD, experts perceived a higher contribution of the musician than novices.

When detailing the Bellotti/Fyans challenges, regardless of the group, the Accident challenge consistently had the lowest rating, meaning that participants were less 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, Sensors and Full_comb.

Reading the “experts” portion of the graph in Figure 7 from left to right, compared to None, the control condition, Sensors, the level of detail exposing only sensor 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. In the “novices” portion of the same figure, in a much more volatile distribution, the results nevertheless present Full_comb as the most effective.

The efficiency of Full_comb for novices is also supported by an analysis of the difference with the experts’ scores. For six (out of nine) dimensions, the smallest difference is measured when visual augmentation are presented with Full_comb. This result is a good illustration of the expected role of visual augmentation, compensating for the lack of expertise in novices to achieve a better experience.

**Discussion**

Both tasks, experience and comprehension, show comparable evolution characterized by a minimum for the control condition None and a maximum for the higher LODs (see 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 favored Full_comb and Sensors for experience and Full_graph for comprehension. Experts chose the highest LODs for experience and Full_comb and Full_graph for comprehension.

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 objectively discern the musician’s contribution (objective comprehension), participants favored levels Full_graph and Sensors for understanding the performance, especially when the music became more complex with many fast changes in the sound. This indicates that LODs influence spectators’ subjective comprehension, 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
Figure 6. Even if experts have more trust in what they think they can perceive from the interactions (subjective comprehension: a), they do not perform over novices in the evaluation of the objective contribution of the musician (objective comprehension: b).

Figure 7. Levels of detail of visual augmentation did not equally impact the subjective perception of the interactions. Moreover, compared with novices, experts reported higher evaluations of the Bellotti/Fyans challenges (subjective comprehension) and higher ratings of their experience and the virtuosity of the musician.

**SUBJECTIVE COMPREHENSION**

![Graph showing subjective comprehension for experts and novices](image)

**OBJECTIVE COMPREHENSION**

![Graph showing objective comprehension for experts and novices](image)

**SUBJECTIVE SCORES BY LEVELS OF DETAIL**

![Graph showing subjective scores for experts and novices by levels of detail](image)
interactions as proven by their higher evaluation of the Bellotti-Fyans challenges (subjective comprehension). This contrast is confirmed in Block 2, in which only novices favored the Sensors LOD over no augmentation for better comprehension and experience [Figure 8]. It was 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 utilized Full Comb when they could choose their favorite LOD. But when they had to choose a LOD to better understand the interactions, experts equally used Full Comb and Full Graph, whereas novices overwhelmingly favored 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, Berthaut, and Grisoni 2020a], we underlined the selective impact of visual augmentation on subjective, rather than objective, comprehension. Additionally, we showed that when participants watch DMI performances with visual augmentation, they overestimate the musician's contribution over the computer's, as do the experts when compared with novices in our study. Thus, multiple pieces of evidence support the idea of a rather subjective nature of expertise. The way we evaluate objective comprehension surely holds limitations. Increasing data, however, supports the hypothesis that, from a spectator perspective, experts are experts because they feel they are, not because of more reliable judgement than 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 greater contribution from the musicians than from automated processes.

Thus, by biasing the spectator perception towards a greater contribution of the musician [Capra, Berthaut, and Grisoni 2020a], and by strengthening spectators’ confidence in their representation of the interactions (subjective comprehension), visual augmentation levels novices upwards toward the rank of experts, especially when novices can select their favorite 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 judgment of virtuosity [Gurevich and Fyans 2011]. A solution to this problem could be inspired by music video games in which virtuosity is concretized by screen indications of combinations of successful moves. Such informative contents are efficient and dramatic but imply the restriction of any improvisation or unexpected techniques. Another solution would be to design LODs that
inform virtuosity, such as visualizations of input complexity or extraordinary values for controls and musical parameters.

**LOD Choice Strategies**

Strong differences in the choice of favored LODs at the individual level were revealed by the data and refined by the interviews. When analyzing the answers of participants regarding how they would use the LODs in public performance, we can distinguish three clear strategies:

1. **All or (almost) nothing**: Four participants stated they would alternate between the maximum LOD (or just start with it) to form a mental image of how the instrument works (i.e., its capabilities) and then go back to no augmentation or to the Sensors level to focus on the musician’s gestures.

2. **Adapting to complexity and performance**: Four participants stated they would use LODs as a way to adapt to the complexity of the instrument or music, or to change it depending on the musician playing.

3. **Progression**: Two participants mentioned that their appreciation of LODs evolved over time, the more complex LODs becoming more enjoyable and accessible, so that they would end up not going back to the simpler LODs.

Note that even within these strategies there are interpersonal variations, again highlighting the utility of a controllable LOD for visual augmentation.

**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 happening on stage during digital music performances. The extra mediation of visual augmentation 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. Although this potential mismatch between perception and reality is a common phenomenon, well known to illusionists and neuroscientists, we here have the possibility of inferring the role of the augmentation in the constitution of inner representations that are more reliable.

In the diversity of potential representations for digital musical interactions, those that expert observers can build should have a greater similarity to the ones used by musicians. As we see it, and based on our data, it seems that they do not. In fact, the role of visual augmentation, and what we call “Spectator Experience Augmentation Techniques” in general (Capra, Berthaut, and Grisoni 2020a), may not only be 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 of offering rather balanced and customized information to spectators, preserving them from potential cognitive overload, as in fully descriptive visual augmentation.

To sum up, techniques to augment spectator experience 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 great deal of information and “feeling” 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 “augmentation conductor,” composing with the direct inputs from the musicians’ instruments and connecting them to types of visual augmentation, while selecting LODs to emphasize parts of the interactions. On the other hand, the augmenter could leave some mystery in certain parts, or even deliberately disturb...
audience perception with disruptive augmentation. As with 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 with VJs, whose role is to illustrate the music with rather exclusively graphical considerations, the augmenter would perform as a human mediator between digital systems and human agents, revealing the virtuosity of the musicians and the expressiveness of their instruments.

Conclusion

In this article, we introduce the concept of levels of detail in visual augmentation 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 instruments 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 favor of the musician's involvement, experts and novices are only distinguished by their subjective comprehension of the interactions, in other words, what they think they understand of the interactions rather than what they understand objectively. These results lead us to hypothesize a rather subjective nature of the expertise, from a spectator perspective.

Regarding LODs in visual augmentation, 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 analyzed their respective strategies during guided interviews. Our experimental approach suggests that by strengthening spectators' confidence in their representation of the interactions (subjective comprehension), visual augmentation is a particularly effective way to “level up” novices toward the rank of experts, especially when novices can select their favorite LODs.

Finally, 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 augmentation and their LODs during performances.

Although 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 feel that augmentation 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 augmentation should be investigated.

References


