The experience of spectators of digital interactions: Benefits of visual augmentations and the role of attributed agency in electronic music performances

Olivier Capra

The experience of spectators of digital interactions

Benefits of visual augmentations and the role of attributed agency in electronic music performances

L'expérience des spectateurs d'interactions numériques

Contribution des augmentations visuelles et rôle de l'agentivité attribuée dans les performances de musique électronique

Thèse de doctorat de l'Université de Lille
préparée au sein de l'équipe MINT du laboratoire CRIStAL et
soutenue le 20 octobre 2020 par

Olivier Capra

Jury

Yvonne Delevoye Professeure, Université de Lille (SCALab)	Présidente
Myriam Desainte-Catherine Professeure, LaBRI, Bordeaux	Rapporteuse
Marcelo Wanderley Professeur, McGill University, Montréal	Rapporteur
Stuart Reeves Assistant professor, University of Nottingham (Mixed Reality Lab)	Examinateur

---

Laurent Grisoni Professeur, Université de Lille (CRIStAL)	Directeur de thèse
Florent Berthaut Maître de conférence, Université de Lille (CRIStAL)	Encadrant de thèse

Abstract

The advent of interactive technologies in our everyday life has led researchers and designers to study how we interact with these automated systems. In arts, and particularly in music, the support of digital technologies offers a still growing and almost infinite amount of creativity opportunities for artists. However this evolution can have consequences on the spectators of musical interactions of a new kind. For instance, the disappearing of the physical link between gestures and sound can generate doubts about the genuine contribution of the artists compared to the one of autonomous processes like sequencers or audio tracks. Making music with computers and electronic sensors raises many further questions regarding the audience. To address these issues, both artists and researchers from the Human Computer Interaction (HCI), musicology or cognitive sciences communities have developed protocols and techniques in order to better understand the spectator experience and to improve it. Although these contributions may definitely benefit the spectators of digital performances, to our knowledge, their efficiency has not been formally compared and spectator experience remains an elusive phenomena not totally captured by field studies.

To clarify the notion of spectator experience, we identify its potential components from the literature. From there, we propose and assess a first model of the familiarity an observer has with a digital instrument. We complete this theoretical contribution, by reviewing the existing Spectator Experience Augmentation Techniques (SEATs) and structuring them in a taxonomy.

We then contribute a series of experiments to compare the effect of SEATs on spectator experience. To do so, we decompose the equivocal phenomena of spectator experience in objective and subjective components easier to handle in experimental protocols. In particular, we underline the central role of Attributed Agency, that is the evaluation by an observer of the level of control of someone executing an action. With this methodological approach, we compare the efficiency of 2 SEATs: pre-concert explanations and visual augmentations, which are the real-time display of graphical representations of the musician's interactions and the mechanisms of their digital musical instrument. We show the positive impact of visual augmentations on the subjective experience of spectators and on the improvement of the trust they have in the genuine contribution of the musician. In a complementary study, we detail contrasts between experts and novices especially in terms of subjective comprehension and we demonstrate the impact of the levels of detail of visual augmentations on spectator experience.

This work also proposes guidelines and tools for future developments. Thus, we introduce PEANUT, a conceptual pipeline for the augmentation of music performances. The pipeline is based on a modular architecture that includes the acquisition of electrophysiological and subjective signals from the audience as well as the extraction of information from the musician's interactions with their instrument. Data are integrated in objects called "correspondences" that aim at mimicking the internal models of spectators. In extension of this concept, we reveal preliminary results that support the relevancy of grip force as an objective marker of Attributed Agency.

Finally, in order to facilitate the reuse of our data and to certify the transparency of our analyses, we follow the Open Science's recommendations and publicly publish the integrity of our anonymised raw data as well as our tools and statistical analyses that conducted to the conclusion presented in this work. In the same vein, we pay special attention to the statistical process by including Bayesian methods allowing us to develop sounder interpretations.

Abstract

La multiplication des équipements numériques dans notre quotidien a amené chercheuses et designers à étudier la façon dont nous interagissons avec ces systèmes automatisés. Dans les arts, et en particulier la musique, cet apport technologique ouvre encore aujourd'hui des possibilités de créativité presque sans limite, notamment grâce à la disparition des contraintes physiques de production du son rencontrées avec les instruments acoustiques. Toutefois, cette évolution n'est pas sans conséquence sur les spectateurs de ces interactions musicales d'un nouveau genre. Ainsi, la disparition du lien physique entre la gestuelle des musiciens et le son peut perturber l'expérience des spectateurs et les amener à douter de la contribution de l'artiste comparée à celle de processus automatisés comme les séquenceurs et les pistes audio. Pour répondre à ces problématiques, artistes et chercheurs en IHM, en musicologie et en sciences cognitives ont développé des techniques visant à mieux comprendre et à améliorer l'expérience spectateur. Malgré le concours indéniable de ces techniques, à notre connaissance leur efficacité n'a encore jamais été formellement comparée et l'experience spectateur reste un phénomène élusif qui n'apparaît pas totalement capturé par les études de terrain.

Pour clarifier la notion d'experience spectateur, nous identifions ses composants potentiels dans la littérature. De là, nous proposons et évaluons un premier modèle de la familiarité qu'un observateur peut avoir avec un instrument numérique. Nous complétons cette contribution théorique avec une revue des techniques d'augmentation de l'expérience spectateur (SEATs) et les structurons dans une taxonomie.

Dans une série d'experimentations, nous comparons ensuite l'impact des SEATs sur l'expérience des spectateurs. Afin de permettre l'étude de ce phénomène relativement ambigu, nous le décomposons en composants subjectifs et objectifs plus facilement adressables dans nos protocoles. En particulier, nous soulignons le rôle central de l'agentivité attribuée qui est l'évaluation par une observatrice du niveau de contrôle d'une personne en train d'executer une action. Avec cette approche méthodologique, nous évaluons l'efficacité de deux SEATs: les explications d'avant concert et les "augmentations visuelles", technique qui consiste à diffuser en temps réel les représentations graphiques des interactions du musicien et les mécanismes de son instrument. Nous démontrons l'impact positif des augmentations visuelles sur l'expérience subjective des spectateurs ainsi que sur leur confiance dans la contribution des musiciens. Dans une étude complémentaire, nous apportons des détails sur les contrastes entre experts et novices, en particulier en termes de compréhension subjective et nous démontrons l'impact du niveau de détails des augmentations visuelles sur l'expérience des spectateurs.

Ce travail propose également des lignes directrices et des outils pour les développements futurs. Ainsi, nous présentons PEANUT, un pipeline conceptuel pour l'augmentation des performances musicales. Structuré en modules, le pipeline inclut l'acquisition de signaux électrophysiologiques et subjectifs des spectateurs ainsi que l'extraction d'informations des interactions musicales. Les données sont ensuite intégrées dans des objet nommés *correspondances* dont la fonction est de reproduire les modèles internes des spectateurs. Nous enrichissons ce concept d'une étude préliminaire dans laquelle nous démontrons la pertinence de la force de préhension comme marqueur objectif de l'agentivité attribuée.

Enfin, nous adhérons aux recommandations Open Science et déposons publiquement l'ensemble de nos données anonymisées ainsi que les outils et les analyses statistiques qui ont mené à nos conclusions. Dans la même idée de transparence et de reproductibilité, nous portons une attention particulière aux traitements statistiques en y incluant des méthodes bayésiennes qui nous permettent de proposer des interprétations plus fiables et plus étayées.

This thesis is dedicated to Julie and our son Marcello and to my grandmother who gave me the love of books and knowledge.

Acknowledgements

This work would not have been possible without the crucial help of many people:

I would like to thank Pr. Myriam Desainte-Catherine and Pr. Marcelo Wanderley for having reviewed this work and for their precious recommendations. I also do thank Dr. Stuart Reeves who accepted to examine the thesis defense in remote conditions.

I feel really grateful to my supervisor Laurent Grisoni and my master Head Professor Yvonne Delevoye-Turrell who took the challenge to welcome an "atypical profile" student. Thank you very much Yvonne for the chance you gave me and the trust you have placed in me and thank you Laurent for your inspirational vision of science and for making this multidisciplinary adventure a reality.

For nearly five years, I received daily support from Florent Berthaut, my co-supervisor. Thank you very much Florent for your help, your constant support and all the discussions we had. I am looking forward to our next collaborations (which have already started) !

My sincere appreciation goes to Janick Naveteur who taught me, among many things, to love methodology, statistical analyses and punctuation.

During these last 6 years, I had the full support of my family even though I could not visit them so often. I really thank you all for that, my parents, my sister, my brother and all the Capra family (and related!).

I would like to thank everybody at the doctoral school Science for engineers (SPI) and in particular Thi Nguyen and Malika Debuysschère.

Along the way, I had the chance to cross passionate and compassionate persons, Fabien Torre, Bérangère Stassin, Laurent Sparrow, Antoine Deplancke, Clémence Roger, Sławek Staworko and Daniel Lewkowicz. Thank you guys for your support and the inspiration you gave me by doing your job the way you do.

As temporary assistant professor at Lille University, I often received support from colleagues. Thank you Stéphane Chaudiron, Laure Bolka, Eric Kergosien, Marc Tommasi, Mikaela Keller, Louis Bigo, Charles Paperman, Agnieszka Tona, Marie Despres-Lonnet, Philippe Vervaecke, Laurent Feisthauer and Remi Bardenet.

Without the understanding and the great flexibility of my teammates at Le Fresnoy - National Studio, the adventure would have stopped very early. Thank you Alain Fleischer, Stéphanie Robin, Michèle Vibert, Eric 'hi-teck' Prigent and Romain Lefranc (coucou).

Thank you Xavier Sense for passing me the torch at Lille 3, you are among those who put me on the way. I also thank CAF for having me attended the exhibition *Mathematics: A Beautiful Elsewhere* where I bought Jean Pierre Changeux's book, "L'Homme Neuronal" the book that lighted the fuse ...

A big thanks to all the participants to our studies for their patience and their commitment.

I would like to thank everybody at MINT & IRCICA: Cagan (pff mais pourquoi t'écris ça ?), Yosra (thank you for your precious advices), Patricia (I should have started writing the state of the art when you told me to ...), Ehrsan, Nicolas, Vincent, Erwan, Stevens, Valentin, Saana, Hanae, Sam, Julien, Jean-Luc, Fred, Farzan, Guillaume, Eric, Paul Elian, Anis, Veïs, Natalie, Ahmed, Anne-Sophie and Peggy.

Thanks to the IrDive team for their support during the TOSCADE and WICE sessions: Yann Coello, Alain Fargue, Marion Vincent, Laurence Delbarre, Diane Togbe and Mohamed Ladrouz. Thank you Mathieu Giraud for your piano.

A musical thank you to Olivier for his great help on the grip force project.

Thanks to my Master mates, Juliette, Anthony, Florian, and Augusto José Mendes Tomacagar. You were the dream team !

Many discussions fed the idea of a PhD adventure. Thank you Łukasz, Grégoire, Sequoya, Niels, and all the mates from the J.B. Lebas' house sharing (7fever)!

I would like to address a particular thank you to my new team at the Anthropo-Lab who let me prepare the thesis defense in perfect conditions, Nicolas Vaillant, Matthieu Belarouci, Antoine Deplancke, Vincent Lenglin, Julie Delporte and all the teammates.

This thesis was written in many places. Each time necessary, I found hosts who nicely gave me a place to work in great conditions. Thank you very much Marie and Dominique (St Maur forever), Jacky and Sylvette from Valmeinier (Salade de fruits forever) and Christophe and his municipal library team in Segonzac. Thank you Ofi for the english support from the desert.

Thank you Chloé and Adrien for your support and your magic potion at La Gwape.

I discovered many theoretical insights of Neurocognitive Sciences thanks to the great podcasts collection at Le Collège de France. These precious contents, and in particular the lessons of Stanislas Dehaene definitely helped me to produce this work.

Finally, I would like to address a very very special thanks to the one who supports me every days and nights, who spared me hundreds of diaper changes, meals, and night-time awakenings, eventually making the PhDAD experience a very enjoyable moment, thank you so much Julie, I feel so lucky to have you by my side, I love you.

This thesis received the support of the Hauts-de-France region and the University of Lille.

Contents

Every act of perception, is to some degree an act of creation.

G.M. Edelman

Introduction

Music is one of the most intense connections between our internal universe and the external world. When you think about it, it is fascinating how simple air vibrations can have a so direct yet so complex influence on us. A few notes of music are enough to elicit deep emotions, bring back memories to consciousness, invite to meditation or emphasize motivation and even improve stamina.

When the music is played live, the experience we feel is even more vivid as we face the music being created before our eyes and ears and as we share it with others in audience. One of the main goals of this work is to understand what makes the experience of live music so special.

Context: Intimate relation with music

In a few preliminary considerations, we would like to briefly present historical and cognitive insights that can be useful to introduce our work on the experience of spectators of electronic music performances.

Music is sound and movement

Until recently in the history, live performances were the only configuration we had to enjoy music. For millenaries, before recording and broadcasting technologies, musicians and listeners had to share a relative immediate vicinity. More generally, our relation to sound is not only auditory. From the singing of birds, or even the wind in the trees, to the first human vocalisations (-550 000) or the very first music instrument (-40 000), a flute made in a bone, the perception of auditory signals is almost systematically associated with a physical origin and a location that we can see, touch or smell. And when the causal process of a sound can not be established, magical explanations may arise like the presence of a ghost due to a human-voice-like sound of the wind blowing in a creaky corridor.

With the advent of technology, we started to get used to the fact that a sound may not have a visible or physical origin. A conversation over the phone is an example of distanciation of the link between auditory signals and usually associated modalities like the fact of seeing, smelling or touching someone when speaking with them. Yet, the feeling of being *with* the person we speak to over the phone still exists, even if the caller is not physically present. In this case, the experience we share is more about the link we have with our inner representation of the person rather than the physical person herself. With this simple example, we want to illustrate that the contribution of technology is not only the reduction of physical constraints of the world, it is also a mediation between the external environment and the inner representations we built of it.

Inner experience and deep mechanisms

In many aspects, music is as much an inner experience as a perceptive experience. This very special relationship we entertain with music could find its origin in the same deepest mechanisms that support our interaction with the world.

From a neuroscience perspective, watching someone playing music triggers a large variety of processes. Among them are the ones supporting our ability to integrate information coming from different modalities (hearing and sight for instance). These mechanisms, already effective at birth, are the basis upon which complex abilities are built much later like speaking or walking.

Another determinant asset is our propensity to unconsciously simulate the actions made by others as we watch them. It is supported by a wide spread network of neurons identified as "mirror neurons" that have the distinctive functionality to get activated both when we realise an action and when we observe the action performed by someone else, like a mirror would, hence the name.

By founding our interpretation of the world on the integration of multiple modalities and the unconscious simulation of others' actions, evolution made us very adaptive to our environment, capable of high level social interactions, and as a lucky side effect, particularly equipped to play and enjoy music.

Later in this work (1.3.3), we will see how these insights from cognitive neurosciences will help us study the experience of spectators. In particular, we will analyse the importance of these mechanisms in our ability to evaluate the level of control musicians have over their instruments.

Digital musical instruments

A revolution occurred in music with the appearance of *Digital Musical Instruments (DMIs)*. For the first time in the history of music, artists were not constrained any more to physically produce the sound. Along with the progressive inclusion of computers on stage, DMIs offer artists the opportunity to express their creativity with almost limitless expression modalities. We will discuss in more details on the possibilities provided by digital instruments in section 1.2 but here we would like to introduce important questions of this work.

As we will explore it, cultural and cognitive foundations support the deep connections between music, gestures and experience. So what would be the impact of instruments that disconnect the sound from the gestures ? If a musician can still have a multimodal interaction with the instrument, by seeing it, touching it, hearing it, or even building it, what would be the consequences of the restrictive perception a spectator has of these interactions ? To what extent would the experience of spectators be altered by a confusion in the relationship of gesture and sound production ?

In this work, we address the question of the experience of spectators of digital interactions with a multidisciplinary approach. By considering the cognitive aspects of our relationship to music and the possibilities of real time technologies, we propose to clarify the notion of spectator experience and to evaluate the techniques that improve this experience, particularly through the mediation of visual representations of musical interactions.

Challenges

Methodological challenge: live from the lab

The experiments presented in this thesis address the experience of spectators of music performances and were mostly conducted in lab. The approach of a subjective phenomena like concert experience outside of its ecological environment may seem counterproductive. Actually, it allows for the production of knowledge that would be impossible to gather or identify with a field approach. In this work, we often discuss the contributions and the limitations of "in the lab" versus "in the wild" approaches (especially in Chapter 4). We conclude to the indeniable complementarity of the two approaches, especially in the study of human behavior and we advocate for the spread of mixed-methods in HCI.

Statistical challenge: the Bayesian approach in HCI & Open Science

At several moments in the production of this thesis, we had to question our approach of statistical analyses to propose more robust and reliable interpretations of our data. After extensive reading and trainings, the Bayesian approach appeared to be just what we were looking for. The works of Kay et al. [120] and Wagenmakers et al. [203] (and the forum at https://forum.cogsci.nl) were critically helpful to assimilate the Bayesian fundations and provide concrete advice to apply them on the analyses of our precious data. To provide the evidence required by diverse readers, we accompanied the Bayesian analyses with frequentist statistics and their famous (rather than well known) p-values. Still, our interpretations were most of the time based on the evidence quantification offered by the bayesian framework. Considering the moderate sample size of our experimentations (around 20), we found the approach particularly convenient and sound.

Quite naturally, we extended the cumulative uncertainty paradigm that imbues the Bayesian approach by joining the Open Science movement. Thus, all the data of the research presented in this document are available at: `https://o0c.eu/0TH/`

"So what ?" challenge: Multidisciplinarity in the real world

During the thesis (2016-2020), we received a lot of enthusiastic comments about multidisciplinarity, about the relevance of the openness to other fields and the sustainability of the approach compared to intellectual dead ends that too restrictive paradigms may lead to. Yet, we also received consistent warnings on potential publishing issues. Members of a community could face difficulties to evaluate a multidisciplinary work that de facto embeds contents that are not strictly relevant for their expertise. The challenge was real. So we are pleased to have experienced the openness of the HCI community where our protocols and statistic analyses more common in Cognitive Sciences than in Computer Science were well received and appreciated. We hope that this work can be seen as another illustration that HCI offers a breeding ground to multidisciplinarity and proves the relevance and sustainability of the approach.

Thesis structure and contributions overview

Chapter 1 presents the theoretical framework of the thesis. After a brief review of the neurosciences of music, we propose a historical evolution of Digital Musical Instruments (DMIs), detail the cognitive concepts we use throughout the document and expose the previous contributions in the understanding and the improvement of the experience of spectators of performances with DMIs.

In **chapter 2**, we propose a model of the familiarity spectators have with an instrument. We confront this model through an online study and underline the central role of Attributed Agency (the spectator's evaluation of the musician's level of control) in spectator experience.

In **chapter 3** we present a taxonomy of Spectator Experience Augmentation Techniques (SEATs). This formal description of the various contributions in the literature is meant to help the analysis of existing techniques and the creation of novel ones.

> The contribution of this chapter was published in the proceedings of NIME 2020 and obtained the Best Poster award. [39]

In **Chapter 4**, we present and apply a methodology to compare the impact of two augmentation techniques on spectator experience: pre-concert explanations and visual augmentations. To this matter, we developed a controlled protocol based on subjective and objective tasks. The results reveal the efficiency of visual augmentations on the global experience and its subjective components. Specifically, when the performances are displayed with visual augmentations, participants evaluate a greater musician's contribution compared to the one of autonomous processes. This study also provides methodological insights such as the relevance of Bayesian statistical analyses in HCI and the complementarity of 'in the lab' approaches with the common field studies paradigms.

In a second part, we refine the insights on visual augmentations by focusing on two factors: *levels of detail (LODs)* of visual augmentations, that are the amount and the type of visual information displayed during the performance and *the expertise* spectators have in digital musical instruments and electronic music in general. With this study, we expose refinements on LODs and their impact on spectator experience. We also reveal surprising differences between expert and novice spectators. The former do out perform the latter but only in subjective reports, not in objective tasks. We discuss these contrasts with the idea that expertise here would rather be the confidence observers have in their inner representations of the musical interactions rather than an objective ability to extract components from their perception.

> The contribution of this chapter related to the protocol of evaluation of spectator experience augmentation techniques was published in the proceedings of DIS 2020 [37].

> The contribution of this chapter related to the study of levels of detail and expertise was published in the proceedings of NIME 2020 [38] and has been selected for an extended version in the MIT Computer Music Journal.

In **Chapter 5**, we extend the studies of spectator experience augmentation techniques (SEATs) and present a conceptual pipeline that augments music performances by integrating data from the musician, from the audience and from dynamic databases of augmentations. To link these heterogeneous data, we introduce the concept of *Correspondences*, that are dynamic objects designed to imitate the spectators internal models. They allow for the identification of perceptive lacks of a particular aspect of the interaction and the selection of optimal augmentations.

In a second part, we address the need of continuous measurements during a live experience. Indeed, in ecological contexts, measuring the experience through questionnaires or comprehension tasks can be rather intrusive and artificialise the experience we try to capture. In this preliminary study, we reproduce with musical interactions the grip force pattern observed in previous studies. We also discuss the relevance of grip force as marker of Attributed Agency.

> The contribution of this chapter related to the pipeline was published in the proceedings of CMMR 2017. [36]

Music is a language that doesn't speak in particular words. It speaks in emotions, and if it's in the bones, it's in the bones.

Keith Richards

1

Background and contribution positioning

Before focusing on music performances which are at the centre of this work, we introduce a few preliminary considerations that will help to understand our approach and highlight the origin of the deep relation humans have to music. Most of the concepts and theories presented in this introductory matters are intense research fields by their own and will not be fully deployed. Instead, we would like the reader to envision this multidisciplinary introduction as a guided selection of key concepts and frameworks on which our research is grounded.

1.1 Special relationship with music

Music is one of the most intense connection between our internal universe and the external world. A few notes of music are enough to elicit deep emotions [116, 115, 202], bring back memories to consciousness [106, 105, 8] , invite to meditation or conversely emphasize motivation [119] and even improve stamina [118, 117].

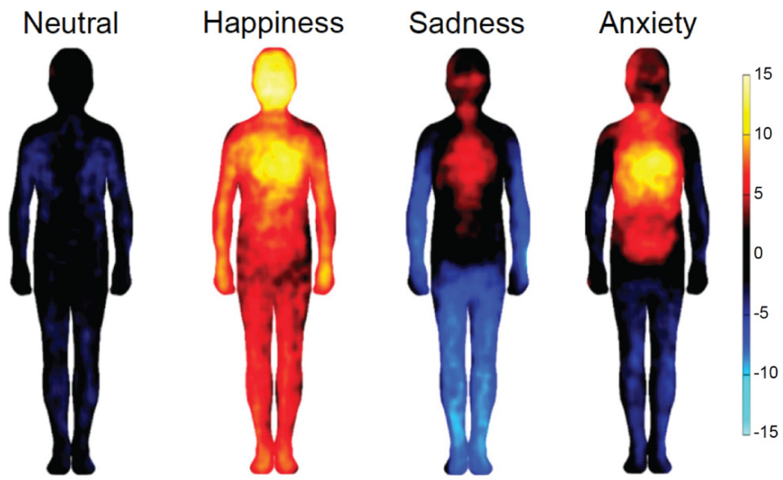
The questions related to our specific relation with music is not a recent matter, from ancient Greeks [30] to recent studies involving high technological equipments [137], researchers and artists continue to explore the subtle mechanisms that make music able to transport our bodies and our souls.

So why do we love music so much ? What makes Nietzsche claim that "Without music, life would be a mistake." or Einstein confessing "I see my life in terms of music." ?

Among the tremendous amount of answers to these questions, a first indication could be that music simply gives us pleasure. With various intensities music is proven to activate neural areas associated with reward and pleasure [215, 24, 65]. As established by Menon et al. [144], our brain is particularly well structured to make "listening to music one of the most rewarding and pleasurable human experiences". From a physical perspective, the pleasure can be bodily experienced with the chill ("'shivers-down-the-spine") that music can trigger with the enjoyable characteristic to be highly reproducible [186]. Furthermore, as our mood is evolving in our everyday life, music is also satisfying because it can accompany us [197] and suit our personal mood-related goal and permit comfort ourself or relieve stress [176, 187].

Emotions A major reason to our attachment to music may also lie in the emotional states it can induce or express [218, 116, 79, 114]. Even if the studies of the emotions elicited by music are beyond the scope of this work, it is worth noting that they represent hundreds of publications [64] in large variety of fields like psychology, philosophy, medicine or musicology. This still effervescent domain of research is a good indicator of the richness and the diversity of the links between music and emotions.

The physical aspect of the emotion-music relation needs also to be considered to understand the approach we developed work. As every emotions, emotions induced by the music is constituted of a physical component (well illustrated by the "heat maps" of Nummenmaa - Fig. 1.1). Furthermore, in their widely recognized theoretical framework of the emotional responses to music, Juslin and Västfjäll [116] expose mechanisms through which music induce emotions. Among them, three mechanisms – the brain stem reflexes, the visual imagery and musical expectancy –, underline the physicality of music that will become more concrete when we will present the importance of gestures and their perception in live music.



Source: Nummenmaa et al. (2014)

Figure 1.1: Emotions are not only a cognitive state. Bodily topography of emotions. Regions whose activation increased (warm colors) or decreased (cool colors) when feeling each emotion. The colorbar indicates the t-statistic range. (From [155])

We saw that one of the reasons we love music is that it triggers emotional as well as physical arousal. When you think about it, it is fascinating how simple air vibrations can have a so direct yet so complex influence on us. This wide range of presence is reflected by the music cerebral processing that is as physical as neural.

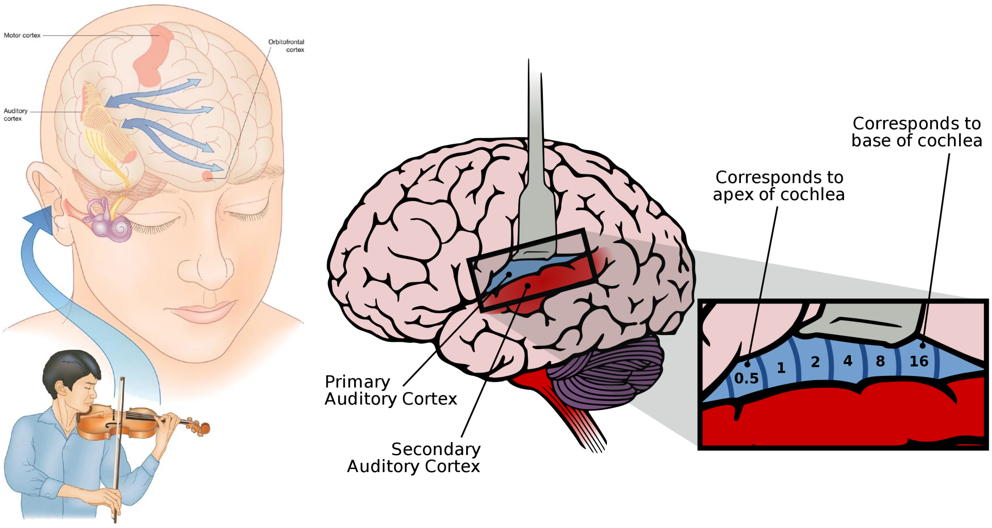
Music is sound and sound is vibrations, most of the time air vibrations, even if it can also diffuse in water or solids like concrete or human bones. Thus listening to music is primarily processing these air vibrations to neural impulses. As depicted on figure 1.2 - left, the inner ear transduces air vibrations into neural signals. From there, the vibratory nature of the sound is transformed in biological signal (transduction). Those signals reach the auditory cortex through the cochlea nerve, the brain stem and the thalamus via chemical and electrical messages. While encoded for neural transport, the information will regain a physical form to process some of the physical parameters of the sounds. Indeed, before leaving the auditory cortex and get addressed by higher level processes (memory, semantic interpretation, ...) the information is decoded in areas physically organised to fit its frequency distribution. This characteristic of a structure spatially organised in correspondence with the tonality is called *tonotopy*. As illustrated on figure 1.2 - right, the primary auditory cortex follows a tonotopic structure meaning that its subareas triggered by given frequencies are spatially arranged following a gradient of frequencies. Even in the brain, the processing of music has physical deployments.

1.1.1 Good vibrations in the brain

To conclude this brief cerebral overview, let us consider a last aspect of music cerebral processing. Notice on figure 1.2 - left how the auditory cortex projects to multiple brain areas. Among them we find the motor cortex that is involved in sensory-motor feedback circuits [215] and in the control of movements. This path _ or wiring _ is a biological evidence of a theory that wants to reunite the body and the mind processes in the constitution of the experience of music: the embodied music.

1.1.2 Music, body and mind

Music instinctively invites to movement. As illustrated by the dozens of web videos of new born babies reacting to music by moving, sometimes even from their mother's womb, or doing their first dance steps when they can barely walk and much later in clubs where people meet to dance in synchrony. Conversely, moving in rhythm with music, tapping with the foot or mimicking



Source: (left) Zatorre, R. (2005). Music, the food of neuroscience?. Nature, 434(7031), 312-315. (right) Chittka L, Brockmann

Figure 1.2: Music is as physical as abstract in the brain. Left: From Zatorre 2005. Illustration of the processing of audio signals coming from an instrument. Right: Lateral view of the brain. The primary auditory cortex (in blue) contains a topographic map of the cochlear frequency spectrum (shown in kilohertz).

an instrument can give the impression of a clearer perception of the music. In other words, everything happens as if the perception of music and music performing were part of a same mechanism. In this subsection, we briefly present a selection of cognitive science, musicology, psychology and neuroscience contributions that constitute the theoretical framework on which we will base the methodology developed in this document.

1.1.2.1 Embodiment

The idea that we interact with the world via mechanisms that fusion perception and action is at the core of philosophical and cognitive frameworks known as embodiment and enactivism [198, 81, 183](Review [85]). They are attempts to explain how we experience the world and find their roots in philosophical works from Immanuel Kant [184] and Maurice Merleau-Ponty [145]. The main idea that the reader should keep in mind is that they posit that consciousness (and by extension conscious perception) can not be differentiate from the body. Even high level processes are rooted in sensorimotor mechanisms "that link the agent to the world in which she is embedded" [179]. The ecological psychology perspective also integrates such ideas by pointing the influence of learned knowledges, current mood and past experiences in our perception of the environment [87]. So embodiment and enactivism are opposed to the cognitivism vision that considers consciousness more like a computer and perception of the world as uniquely materialized by mental representations disconnected from the physicality of the body.

The embodiment theory is also supported by cumulative evidence from the neuroscience field. We saw in the previous subsection that the processing of the sound in the brain can include areas associated with movement. The related neurons, called mirror neurons, are part of a much wider network that is thought to sustain our relation to the world [170, 82] and how we interact with it. The mirror neurons have the characteristic of being activated during the execution of an action and the observation of an action [83, 92]. The activation of the same neurons in action and perception leads the theory to hold that we do not need to explicitly think to understand the intentions, the actions, or the emotions of another person as our brain unconsciously and automatically make us feel them as we feel our own intentions, actions or emotions [156]. From a musical perspective, the embodiment theory is a solid framework that led researchers to propose a dedicated application to an embodied music theory.

1.1.2.2 Embodied music

As mentioned by Schiavio et al [179], the first examples of the application of embodiment to the music cognition are proposed at the beginning of the 00's [96, 217]. Contrary to the cognitivist music psychology for which the brain is like a computer with specialized software for music processing [138], in the embodied music framework, the role of both the performer and the listener's body is crucial in the musical understanding [166, 167, 132]. For Godøy, our relation to music is a "fusion" of auditory and motor sensations [89]. In fact, neuroscience data [148, 216] support the idea that the musical meaning is decoded via action based processes and unconscious motor simulations. These flows of information that constitute the relationship between our body and the external world are called internal models [212].

The concept of internal models is at the core of the embodied music cognition. Indeed, they are the hypothesized mechanisms by which listening to music is sufficient to infer not only the gestures and the interaction that produced the sound but also the intentions and the feelings conveyed by the music [89, 131]. The models operate the gestures-sound association through primary goal-based actions knowledge [180] and get enriched by accumulating experiences.

We precise that our research is based on an embodied approach closer to the one proposed by Leman [132] rather than the enactivist one defended by Matyia and Schiavio [141]. The two paradigms share the main statement that "experience of the world is a result of mutual interaction between the sensorimotor capacities of the organism and its environment" [71]. But Leman includes in his embodied music theory the role of purely representational processes and maintains a theoretical space for a mediation between the mind and the body. In this work, we will support the idea of a mixed contribution of "pure" enactive/embodied and "pure" cerebral interpretations of the multiple aspects (physical, emotional, intentional) conveyed by music. The debates are still lively about these questions. They strongly rely on multidisciplinary approaches as contributions from neurosciences, philosophy and even Human Computer Interaction and Artificial intelligence constantly interoperate.

1.1.2.3 Multimodality and crossmodality

A last aspect of the experience of music stands in its multimodal nature. Indeed, we saw that music experience is an embodied phenomenon that relies on action-perception mechanisms. These mechanisms are not only based on the single auditory modality but integrate visual cues, motion sensations (proprioception), effort and dynamic [90]. Besides their individual integration, the multiple sensory inputs can also interfere, meaning that the information perceived by a modality can be modulated by the information of an other modality [142]. This crossmodality in the musical experience is a crucial characteristic to encompass when studying spectator experience [200]. In section 1.3.1.2 we will address in more details the consequences of the multimodal integration of cues that constitutes the experience of music, especially considering the musical gestures.

1.1.3 Conclusion: harmony for mind and body

In this section, we saw that our relation to music is a wide multidisciplinary question that can not be reduced to a simple translation of vibrations into neuronal signals. Whether as a performer or as a spectator, mind and body form a system from which the experience of music emerges. Still, musicians keep an extra very special link with the music as they can create it, thanks to music instruments. In fact, the coupled system they create with their instrument can be seen as a cognitive system by itself [45] where the instrument is embodied in the musicians [98] as an extension of their body [154]. In the next section, we address the questions of digital musical instruments and focus on the technological evolution that gave musicians the ability to design and customize their instruments.

1.2 Digital Musical Instruments

From the very first known music instrument (-40 000), a flute made in a bone [47], to the inclusion of artificial intelligence [70], musical instruments always followed or contributed to (see Fig. 1.13 - page 35) the technological evolution of humanity. As pointed by Cadoz [32], the tight link between music and technology do raise questions at a moment in history where technology is undergoing dramatic change. From this, the emergence of digital musical instruments (or DMIs) is a major step in the history of music.

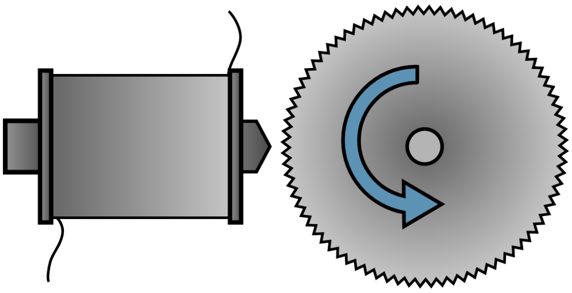
In this section, we propose a historical overview of DMIs, a description of their main mechanisms and the interaction they enable. An exhaustive presentation of DMIs and the related works of the prolific community they aggregate could not stand in one PhD ! Here, we will highlight how technological events, like the generation of sound from electricity modulations and the generalisation of personal computers, led to an explosion of the creative possibilities offered to artists and started to raise new questions for the audience.

1.2.1 History

1.2.1.1 Origins of DMIs

Electronic musical instruments appeared at the end of the XIXth century or the beginning of the XXth century. The Telharmonium is considered as one of the first. Invented by Thaddeus Cahill in 1897, the instrument looks like a massive organ and generates sound from electromagnetic interaction between "tonewheels" and coils. As depicted in Figure 1.3 - Left, a tonewheel is a metal disc with bumps or cogs that concentrate the electromagnetic field when they pass in front of the coil. Doing so the electrons inside the coil are "pushed" (cog) and gradually "retained" (inter cog). The signal forms a sinusoidal wave which frequency depends on the rotation speed of the wheel or its size. That wave is transported along wires, amplified and diffused as a tone through speakers (which were telephone speakers at this time). From this basic electromagnetic phenomenon and the multiplication of the tonewheels, the instrument was polyphonic and the musician could play with the harmonics.

Sadly for Cahill, the Telharmonium did not meet a large popularity, probably because a full room was required to contain the giant tonewheels. Yet, the technique of producing sound by generating electrical sinusoidal waves and adding them on top of each other, called additive synthesis, was a huge contribution with countless successors.



Source: Public Domain

Source: Public Domain

Figure 1.3: Telharmonium (1897) and the Theremin (1920) were the first instruments to allow the generation sound with contactless processes. Left - Detail of a tonewheel used in the Telharmonium. Right - Alexandra Stepanoff playing the theremin on NBC Radio, 1930

1.2.1.2 Synthesizers

Following the Telharmonium, the Theremin (1920), invented by the Russian inventor Lev Sergeyevich (known as Leon Theremin), introduced the first musical instrument with contactless interaction. Here again, tones with any pitch can be generated. The modulations are controlled by the movement of the musician's hands around the two metal antennas of the instrument 1.3. These two instruments are the ancestors of the synthesizers which will generalize the sound synthesis with the addition, the subtraction and frequency manipulation of electrical signals. Their popularity rapidly grows with iconic devices like the ones invented by Robert Moog (Moog Synthesizer - mid 60's, minimoog - 1970) or the Yamaha DX7 (John Chowning, 1983) which was the **first actual digital instrument**. Digital synthesizers pushed further the synthesis by replacing the analog processing, meaning directly supported by voltage control using knobs, buttons and faders, by a digital signal processing where the synthesis is numerically-controlled. With components like Digital Signal Processors (DSP) which can transform any real-world signal like voice, audio or motion into manipulable numeric data, Digital Musical Instruments look more and more like computers.

1.2.1.3 Samplers and sequencers

Somewhere between synthesizers and computers, samplers are a central piece of many electronic music setups. Their main functionality is to play on demand pre-recorded audio sources. Historically, audio sampling was introduced in the 1940's by the French composer Pierre Schaeffer. In his experimental creation *Musique Concrète*, Schaeffer manipulated pre-recorded magnetic tapes to create sound collages and play them in loop. In 1963, the Mellontron is one of the first instruments to propose audio sampling, still via magnetic audio tapes that are read accordingly to the key pressed by the musician. Since the first digital sampler *Musys* released by EMS in 1969, the miniaturisation and the parallel evolution of computer music, audio sampling was integrated in synthesizers like in the Fairlight CMI instruments, delivered as a stand-alone device like in the Akai S1000 and virtualized in computer software. The ability to trigger pre-record sounds but also to modify their pitch, volume, to reverse them or play them in loop made the sampler a staple of modern music exploited by a wide diversity of artists from all styles and generations.

Sequencers Often associated with a sampler, a sequencer can record, edit and play back music. The information stored and manipulated by a sequencer is not an audio signal but a control data like MIDI or OSC (See Section 1.2.1.6). Whether as software or a dedicated device, the sequencer typically sends commands to a sampler that plays associated sounds. This functionality is particularly used to play autonomous sequences of music. Thus, a musician can launch a sequence, like a drum loop, and overdub it with a bass melody for instance.

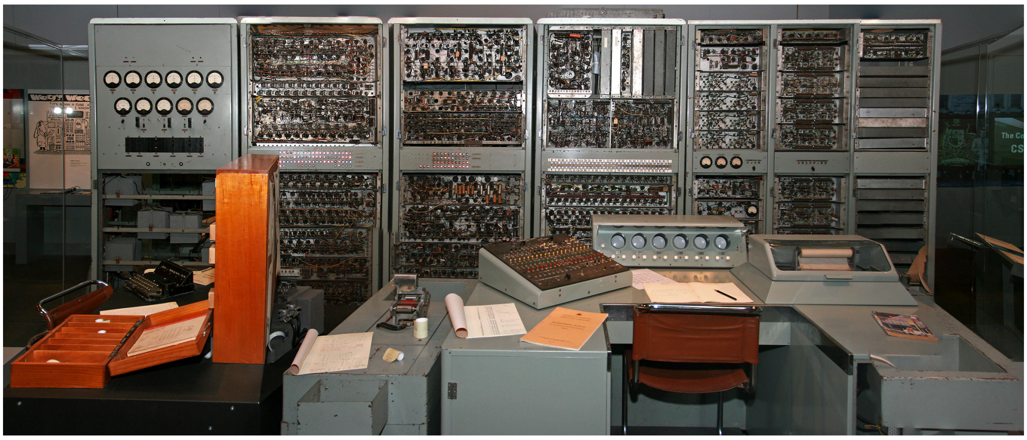
1.2.1.4 Computer music

Computer music really spread with the democratisation of personal computers in the 1990's. Before this date, the initiatives of making music with computers are mostly supported by mathematics researchers and the machines could require a full room to get set up. (Fig 1.4). Named after the CSIRAC, the CSIR Mark 1 (Trevor Pearcey and Maston Beard - late 40's) is the first computer that played music in public. Forty years later, with the miniaturisation of the machines and the democratisation of their price, computers became a very popular tool for artists. Especially in music where a large part of the music creation, recording, production and diffusion will progressively be manageable with a single computer and a few accessories. With the evolution of computers and digital storage, the level of customisation goes one step higher. Musicians can play and configure multiple tracks at the same time, in real time, thanks to the rich user interfaces of optimized dedicated software.

From the 1980's, software evolved in parallel of the hardware (computers) and also became more affordable, ergonomic and responsive. In the 2000's, a computer equipped with software solutions like Avid Pro Tools, Apple Logic, Ableton Live or Propellerhead Reason, can reproduce any analog, often expensive, device allowing musicians to produce professional music contents from their home studio. Finally, in the late 2010's, the miniaturisation gives smartphones and tablets the capacity to compute heavy sound processing and run full Digital Audio Workstations (DAWs). With these late innovations taking a large advantage of the high rate internet connections, musicians can record and produce music wherever they want.

1.2.1.5 Instruments programmers

Besides music production or edition software, a range of programming languages progressively offered musicians the ability to control their music with great refinements. In 1957, Max Matthews officially gave the first computer music performance with his *Music I* that produced



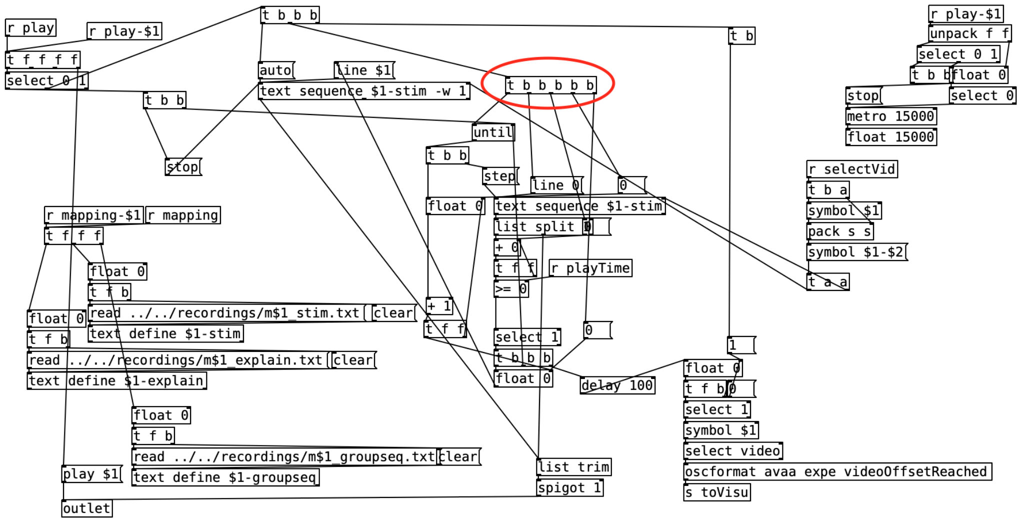
Source: John O'Neill - GFDL

Figure 1.4: CSIRAC (1949) is one of the first stored program computer. Despite rumours, this model was not proven to be able to play music. One of its evolution, the CSIR Mark 1 is officially the first computer that played music (early 1950's)

a 17 seconds preprogrammed melody, running on an IBM 704. Twelve years later, a real time control was possible with the music synthesis system GROOVE. With the development of personal computing at the end of the 1980's, MAX proposes a graphical interface to control external hardware devices before becoming MAX MSP and real time sound synthesis. In 1997, MAX's creator Miller Puckette launched an open-source initiative: Pure Data. Both MAX MSP and Pure Data (PD) propose modules to create complex graphs of processes with a high level of customisation. As depicted in Figure 1.5, a PD patch is a collection of connected modules. These modules are boxes with inputs and outputs (Fig. 1.5 - red circle) connected by graphical wires. Some modules can process sound while others convey discreet or continuous flow of data. With physical devices, the costly operation of connecting the frequency output of a sinusoidal generator to the volume input of a digital sampler becomes a trivial mouse click work with such programming languages. Thus, with a relatively small time investment, a computer running a pure data patch becomes a full custom musical instrument. The mouse or the keyboard of the computer are not the only way to play the computer. The control can be devoted to an external device, like a digital piano keyboard or a drum pad, thanks to the ability of DMIs to exchange information.

1.2.1.6 Communications protocols

Musical Instrument Digital Interface (MIDI) As the sound and parameter modulations became electrical, and then digital musical instruments allowed for the exchange of information between each other. At the beginning of the 1980's, companies looked for a standardisation



Source: John O'Neill - GFDL

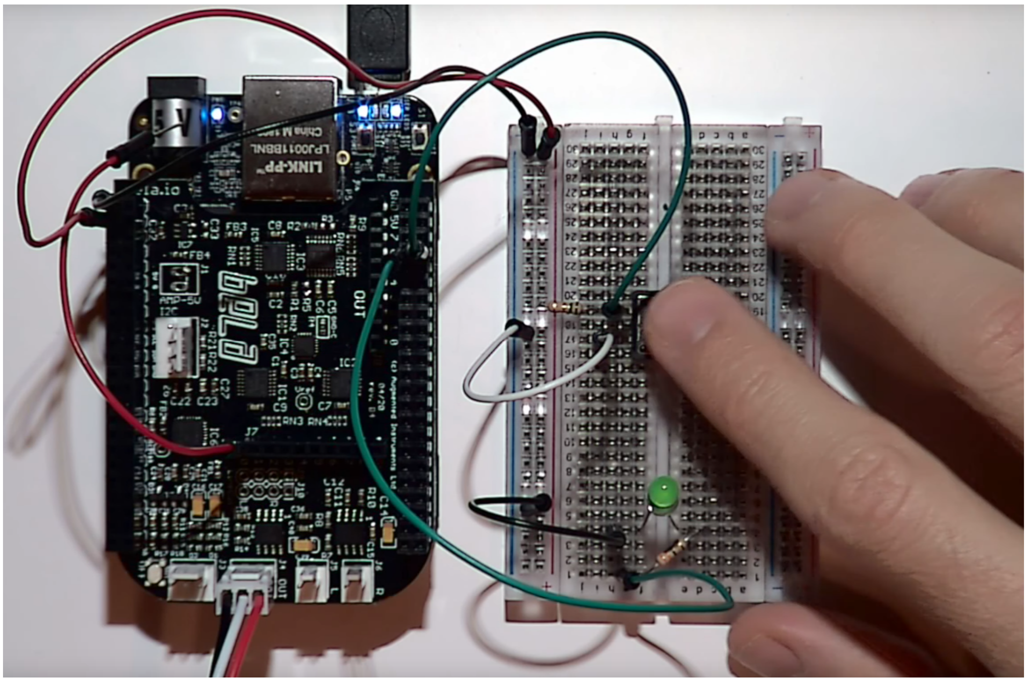
Figure 1.5: A Pure Data patch. Programming languages allow musician for refined, complex and custom sound processing.

of the communication protocols between instruments. After common discussions between manufacturers, the initiative finally came up with the Musical Instrument Digital Interface or MIDI in 1982. With the MIDI protocol, devices from different manufacturers can communicate, synchronize or control each other. For instance, a common MIDI setup is a digital piano keyboard linked to an audio sampler. When a key is pressed on the keyboard, the reference of the key, along with various information like the velocity the key was pushed, is transmitted via the MIDI protocol to the audio sampler that triggers the corresponding sound. We will describe in more details the relation between controllers and sound processes in section 1.2.2.2.

Open Sound Control (OSC) In 1997, Adrian Freed and Matt Wright [72] present Open Sound Control, a protocol inspired by the network communication protocols allowing for better performances than MIDI in terms of time resolution and flexibility of usage. In particular, it offers higher data resolutions (32 or 64 bits) to manipulate more complex control streams. Additionally OSC allows musicians to route the messages between instruments over common communication networks, including the Internet, and also to send messages to multiple recipients/modules/instruments.

1.2.1.7 Instruments makers

During the late 1990's, in parallel of the generalisation of personal computers in the creative industry, the creation, the modification or the hacking of electronic devices becomes popular.



Source: Bela - Youtube video

Figure 1.6: A Bela electronic board in a basic setup (from the Bela Platform Youtube channel) - In the 2010's customisable electronic boards became more and more popular making musicians able to build their instrument from the ground up.

Companies like Arduino or Bela now propose microcontrollers mounted on electronic boards, diverse ready-to-mount sensors or compatible components. All these electronic equipments are designed to be modular and easily integrable in already existing music setups. Conveniently, their microcontroller can be configured with program sent from a personal computer. Coupled with a large amount of tutorials and documentation, affordable prices and enlarged customisation possibilities, these equipments allow musicians to literally compose their instrument and to link with refined dynamic relations a large variety of sensors to any audio signal and effects.

1.2.1.8 NIME

The success of a workshop at the ACM Conference on Human Factors in Computing Systems (CHI) in 2001 inaugurated an new international conference dedicated to research on new technologies in musical expression and artistic performances: New Interfaces for Musical Expression (NIME). Each year, in a place around the world, artists and researchers expose their work on the instruments of the future and discuss diverse issues in terms of design of controllers and interfaces, real-time gestural control or perceptual and cognitive issues related to instruments control ([110] for a complete review of fifteen years of NIME 2002-2017).

Finally, by intertwining artistic performances and scientific contributions, by supporting discussions between fields like design, HCI, computer, or musicology, research in musical expressiveness is a symbiosis, "where artists could not have been so far with computers without the help of researchers and engineers, and vice versa, from a scientific point of view, we could not have reached such complex questionings about our field if artists were not involved in the process[1].[61].

At the time this thesis was written, we just attended our first NIME edition and finally exchanged (even if remotely because of COVID19) with this welcoming and very enthusiastic community !

1.2.1.9 Conclusion

In this overview of the evolution of musical instruments, we saw that artists are offered continuously growing creative possibilities. In fact, musicians can play pre-recorded materials or generate new ones in real time, customize and build their own instruments or augment existing ones to eventually shape refined custom interactions and deploy their artistic envision. In the next subsection, we will describe the main technical aspects of the DMIs that allow for such a diversity of usage.

1.2.2 Structure and functioning

Digital musical instruments (DMIs) allow musicians to generate, trigger and control any sound in an infinite amount of ways. Tanaka an instrument is not utilitarian, and is not designed for a single well-defined application To achieve this diversity of usage, a multidisciplinary approach involving fields like engineering, arts and design [143], is required in the creation of DMIs. The global structure of DMIs can be summarized in three main components: control interface, mappings and audio processes.

1.2.2.1 Control interface

The control interface is the part of the DMI that is used by the performer to control the instrument (See Fig. 1.7). It is commonly composed of physical sensors like buttons and faders or graphical elements of touch interfaces. Actually, any sensor can take part in the

[1]"On peut réellement parler de symbiose parce que on ne peut pas imaginer que les artistes aient pu aller aussi loin qu'ils ne sont arrivés aujourd'hui avec l'ordinateur sans l'aide des chercheurs et des ingénieurs et vice versa du point de vue scientifique, on est arrivé à un degré de complexité des questions qui se posent dans le domaine qui n'aurait jamais pu être atteint si les artistes n'avaient pas participé."

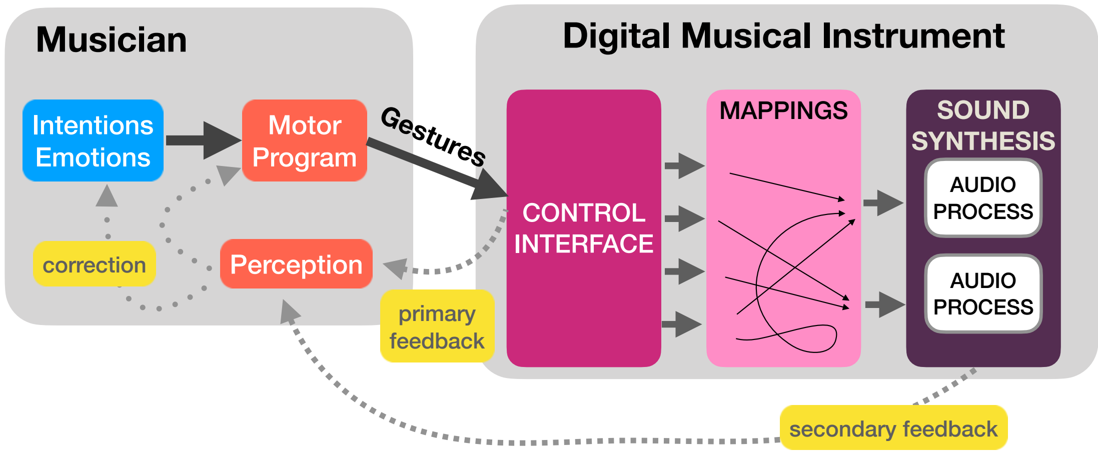


Figure 1.7: Conceptual framework for New Interface for Musical Expression and model of a digital musical instrument (Adapted from [209, 13])

control of the instrument as long as its output remains interpretable by the system: wearable sensors for control by motion, computer vision, electrophysiological signals ([147]) ... The integration of several sources, called sensor fusion,[134] is also possible and even encouraged as it is shown to improve the experience of the performer by giving the DMI a better accuracy (compensation of sensor deprivation and individual spatial and temporal limitations) [143].

While in an acoustic instrument the interface is "inherently bound up with the sound source" [100], in DMIs the interface is mostly independent from the sound source. This separation allows for a wide range of control gestures and opens artistic and scientific questions about the approach to adopt in the design of controllers [49, 205, 147, 33]. For a DMI to be responsive, the interface should rely on sophisticated engineering solutions and refined sensors but it not so often a respected practice [143]. Beyond engineering, phenomenological approaches propose to maintain an inherent integrality by adopting a logical approach of control interface with physical sensing of an object [103]. At the end of the day, "Musical interface construction proceeds as more art than science, and possibly this is the only ways that it can be done" as pointed by Cook [49] showing that the question of control interfaces in DMIs remains vivid and at the heart of the The International Conference on New Interfaces for Musical Expression (NIME) community [110].

Gestures As control interfaces are a mean for gestures to become sound, the question of musician's gestures is central in music research (e.g. in [205, 201]). Before covering the embodied approach of gestures in the next section, here we would like to briefly review the contributions that aim at clarifying and describing musical gestures.

The notion of a gesture can be too broad and difficult to define [201], especially because gestures are multifunctional and have multiple significations, from physical to metaphorical. Furthermore, some gestural inputs that can be used to interact with sound processes are not necessarily movements, like data obtained from image or sound analysis [147] or physiological data (EEG ...). In music research, these various interpretations are often different facets of rich musical experiences [111]. By clearly leaving the ambition to cover the whole multidisciplinary question of gestures, Jensenius et al. nevertheless extracted [111] four functional categories of gestures from a large review of contributions [86, 31, 60, 205] : sound-producing gestures, communicative gestures, sound-facilitating gestures and sound-accompanying gestures. From there, sound-producing gestures can be further subdivided in excitation gestures (like a violin bow stroke) and modulation gestures (like a pitch bend)[33], furthermore, according to Bowers and Hellström [25], non sound-producing gestures with DMIs preserve an *expressive latitude* that can join the Jensenius's definition of communicative gestures.

Finally, DMIs are not necessarily direct sources of sound by themselves [205], so contrary to acoustic instruments, the term "sound-producing" gestures should be temperate with DMIs as the gestures are not sufficient to generate sound. Furthermore, we will see below that a musical gesture with DMI is not constrained to a particular relation with sound production.

1.2.2.2 Mappings: bringing an interface to life

The mapping is the correspondence between the performer's gestures, acquired by the control interface, and the musical parameters. Unlike acoustic instruments where the mapping is imposed by the instrument structure, in DMIs, the mapping must be defined and this is an extremely difficult task [100]. Mapping is "bringing an interface to life." for Hunt et al. [100], and it can "make or break" interesting user interfaces according to Malloch & Wanderley [140]. In this work, we will focus on two important characteristics that mappings provide to a DMI,

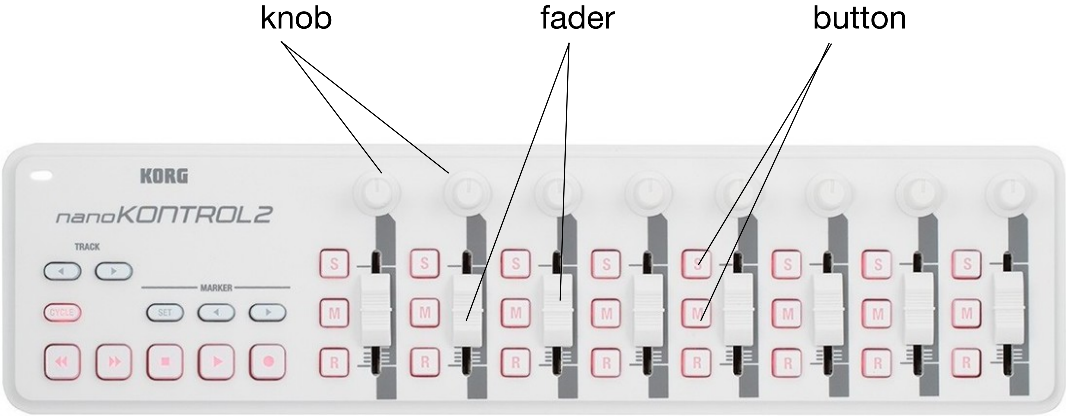


Figure 1.8: A MIDI controller with knobs, faders and button. Through the mapping, each sensor is configurable to alter one or multiple parameters of one or multiple audio processes.

the definition of its identity and the complexity of the interaction it allows.

Defining an instrument The role of the mapping in a DMI is crucial because it defines the user interaction [140] and thus, it has been the object of voluminous research (eg:[99, 204, 5]). Choi considers the mapping as a primary factor to distinguish a DMI from an other [44]. From a performer as well as a spectator perspective, the character of a instrument is changed by altering the mapping, even when keeping constant the interface and the sound [100]. In this work, we will use mapping alteration to create multiple instruments from a same control interface.

Various complexity Depending on the number of inputs and outputs, a mapping can be more or less complex, more or less expressive. Kirk and Hunt [99] propose to classify mappings as follows:

- **one-to-one** : in this straight forward mapping, a single control is connected to the alteration of a single sound parameter. In DMI, this mapping is the one of a fader controlling the frequency parameter of the sound synthesis.(Fig 1.9 - top left)
- **many-to-one** : in this mapping, also called *convergent mapping* [175], multiple inputs are required to control one parameter, like the control of the volume when playing violin. In this case, the bow-speed, the bow pressure and the finger positions all contribute in the control one that single parameter. (Fig 1.9 - top right)
- **one-to-many** : in this mapping, also called *divergent mapping* [175], one control can alter multiple parameters of the sound synthesis, like the bow of violin can influence many aspects

of the sound synthesis such as volume or timber. (Fig 1.9 - bottom left)

- **many-to-many** : in addition to the mappings listed by Kirk and Hunt, the many-to-many mapping is a complex configuration where multiple controllers are required to influence many parameters of the sound synthesis in non one-to-one relations. (Fig 1.9 - bottom right)

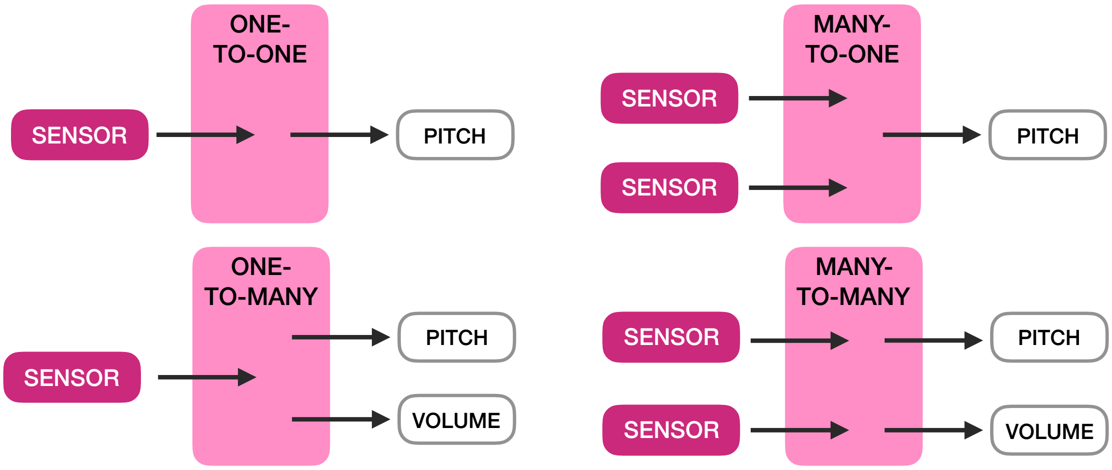


Figure 1.9: Four types of mappings (adapted from [99]). Non one-to-one mappings are more engaging and expressive but too complex mappings can make sound expressions exceedingly difficult to reproduce, especially in public conditions [149].

Thus, depending on their configuration, mappings can be obvious like the ones mimicking acoustic instruments eg: a percussive gesture on a pad that triggers a percussive sound. They can also be more complex and even invisible eg: a single arm gesture could trigger multiple sounds at the same time with a pitch modulation proportional to the amplitude of the arm when the gesture is performed from top to bottom and with a vibrato modulation when the arm gesture in performed from bottom to top. This example illustrates the complexity mappings can develop and how they can appear too complex and become exceedingly difficult to reproduce in public conditions [149].

1.2.2.3 Audio processes

An audio process is the audio part of a DMI. Contrary to most acoustic instruments, DMIs can manage multiple audio processes that are combined to form the instrument audio output. From the two preceding notions, interface and mappings, we can draw a typical use of DMIs: Via the interface, musicians control audio processes accordingly to their related mapping. We will describe common configurations in which audio processes can be fully dependent from the performer interaction but can also have a static or dynamic autonomous behaviour.

Fully dependent An audio process can be fully dependent of the performer. In that case, an input (e.g. : a gesture) is necessary to hear something from the instrument like most acoustic instruments would behave.

Static autonomous An audio process can follow a predefined score without any access of the control interface apart from the initial start signal and the stop signal. This is a typical functionality of the demo songs proposed by digital pianos. Musicians can still play over the playback but they can not modify it.

Dynamic autonomous This type of audio process is particularly useful for interactive performances with a variable ratio of autonomous processes and direct interaction of the musician. The process can follow a predefined score but can also be modified by the control interface. For instance, the performer can control the speed of the playback, alter the pitch or even reverse the audio process via the control interface.

An other common practice is to record a sequence of interactions during the performance and play it back automatically in loop, a technique called live-looping. In that case, the behaviour of the audio process is mixed. Starting fully dependent during the recording of the sequence, it goes autonomous when the musician triggers the playback of the recorded sequence. The performer is thus freed of playing that particular sequence and can then freely control the sequence parameters (dynamic autonomous) or focus on any other audio process (static autonomous).

Source: Cook, P. (2017). 2001: Principles for Designing Computer Music Controllers. In A NIME Reader (pp. 1-13). Springer, Cham.

Figure 1.10: From [49], the Java Mug (top right), is an illustration of the opportunity for musicians to build instruments that fit their internal models. Here, the mug is the author's favourite object. The gap with spectators' internal model can lower their experience during performances using such instruments.

1.2.3 Conclusion: the instrument artists have in mind

We saw that digital musical instruments concentrate many technological improvements of this past hundred years. They offer artists countless configurations thanks to their modular design that removes the physical constraints of sound production and allows for custom and complex mappings. These characteristics are invitations to experimental development often leading DMIs performers to be their own instrument makers and testers. From this all, the shift with acoustical instruments is radical.

Adapting DMIs to the performers' needs, creativity and projects will continue to raise challenges for the NIME community in terms of expressiveness, reliability, reproducibility or durability. Still, in 2020, musicians never had such a profusion of ways to shape their instruments to fit what they have in mind, to fit their internal embodied models [62] that animate their relation to the world (see Section 1.1.2.2. These opportunities to shape the instrument to tightly fit one's internal model can be illustrated by the JavaMug (Fig. 1.10), an instrument made out of the author's favourite object.

What does this relation to their instrument become in public performance ? How can artists deploy the artistic envision, the skills or the sensitivity they intended in the customisation of their instrument ? Is it possible for the audience to fully appreciate the performance like they would with acoustic instruments ?

These questions are central to this work, in the next part, we will address them and for dedicated research and methodologies in favour of the spectator experience.

1.3 The challenges of spectator experience

We saw in previous sections that the experience of music is supported by cerebral and bodily mechanisms also called internal models. We also described how the evolution of digital musical instruments freed the musicians from many physical constraints in production and manipulation of sound and allow them to build and customize instruments. With such conditions, DMIs become extension of musicians' internal models.

Paradoxically, the creative opportunities and the proximity artists have with their instrument can distance spectators with perceptive difficulties during live performances. Indeed, we will see that, contrary to acoustic instruments, interactions with DMIs do not provide enough cues to observers for them to build reliable internal models and share the richness of the musicians' envision. In this section, we will present how, in response to these new challenges, research emerged and contributed to define, measure and improve spectator experience.

Figure 1.11: From a spectator point of view, musical gestures with Digital Musical Instruments can be difficult to perceive and to decode compared to acoustic instruments.

1.3.1 Performances with DMIs: gaps in perception for the spectators

1.3.1.1 Live concerts

The live performance is a very particular moment for musicians. For many of them, this is the primary goal of playing music. Many conditions are in place to make the event special. The stage, whether a virtually delimited space in the street or a large stage with dedicated lights, is a decorum that intensifies the moment. When playing, the sound is louder than in rehearsals and the music produced is more condensed in intentions, freed from unintentional and casual playing that happens at home.

For centuries, the live performance was actually the unique way to share music. Today, listening to music has never been so easy. Technology in 2020 offers the best listening conditions for a relatively reasonable investment, video techniques allow for immersive contents with extreme realistic renderings and social media let people talk to artists without almost any barrier. Still, people love to go to concerts, whatever the conditions, in small bars or in stadiums, sometimes whatever the distance or the cost.

So why are concerts more popular than ever ? Maybe because a live performance is a show. In the literal sense, performing is showing, exposing, embodying an artistic envision. So, despite some rare exceptions, a concert is a dedicated moment for sharing. Sharing emotions, pleasure, aesthetic, all in a delimited time, in a delimited space. The intensity and the ephemerality of the moment are probably strong factors to explain why people attend concerts. Why does music feels more intense, deeper and live in these conditions ?

1.3.1.2 Performances with acoustic instruments: an embodied and crossmodal experience

One of the reasons of the deeper experience of music in live concerts lies in the embodied and crossmodal nature of music experience (See section 1.1). For spectators, seeing a musician playing an acoustic instrument increases the experience of music because the complementary perception of music and musicians participates in the construction of a richer internal model of the interactions occurring on stage (See section 1.1.2.2). The virtuosity of musicians is particularly visible even for naive spectators with no expertise. Indeed, with the explicit and known affordances [88] of acoustic instruments, the audience can most of the time "delineate action relationship between the instrument and the musician" [201].

© Kmeron

Figure 1.12: Gestures with Digital Musical Instruments do not always allow spectators to perceive the intention or the contribution of the performer.

Multiple work confirmed that musical gestures with acoustic instruments are not only instrument control contingencies (e.g. : [204]). They are expressive, communicate aspects of the musicians internal states [59], and may be different during concert than when executed alone [60]. For Visi et al., *they act as bridge between bodily movement and meaning formation* [201]. Thus beyond movement, musical gestures encode intentions, goals, and expressions [130] and their undemanding decoding is handled by internal models.

Besides, the consistency between a gesture and the related sound (like the gesture amplitude and volume for instance) fits well with the the predictions listeners can build upon their internal models of the interaction. Additionally, musicians' body movements are known to convey emotions by themselves [57] and so enrich the experience with non exclusively audio contents. Finally, audiovisual perception also improves the ability to extract musical properties like the level of tension or creativity [200, 125].

1.3.1.3 Performance with DMIs: gaps in perception for the spectators

From a spectator point of view, many aspects of performances with acoustic instruments fade out in performances with digital musical instruments (DMIs).

Unusual instruments Thanks to the modularity and the relatively low cost of components, musicians can easily build or configure their own DMIs to fit their creative projects. Consequently, concerts are often performed with devices that exist nowhere else and so the audience can not build any expertise from earlier expositions or identified repertoires. Besides, contrary to acoustic instruments for which physical affordances can inform about their potential sound or functioning, no clear interpretation of the musical possibilities can be inferred from the appearance of DMIs.

Control dislocation Because DMIs separate sound from control, a characteristic called *control dislocation* [147], a series of difficulties arise for the spectators. The control interface of the instrument is not always visible (hidden sensors, tiny controllers, see section 1.2.2.1 for details) and their activation with gestures is not always obvious. A gesture can activate a sensor at a moment and later the same gesture can lead to nothing if the controller is temporally deactivated, when the gesture was not sensed for instance. Furthermore, the confusion is reinforced as the same controller can be used to control different sound parameters throughout the same performance and multiple artists can use the same interface in different instruments.

Complex mappings Additionally, depending on the mapping, the physicality of a gesture is not necessarily reflected in the audio output, if not reversed. Thus, a greater amplitude in the gestures will not always trigger a greater modulation in the sound or a single percussive gesture can be used to slowly and continuously modulate multiple sound parameters at the same time (divergent mapping, see 1.2.2.2). Consequently, contrary to acoustic instruments, no clear outcome expectation or explicit link is accessible for spectators when subtle gestures can trigger large audio consequences and vice versa.

New sounds Additionally to these control and gestural aspects, sound synthesis is an other shift with acoustic instruments. As pointed by Tanaka [194], sound synthesis is powerful precisely because *it creates new and never heard sounds* and so can disorient the listener. In a context where gestures can be connected to complex mappings, or transformed via live algorithms, autonomous processes like audio loops or sequenced partitions (see section 1.2.2.3) can also contribute to the disorientation of the audience especially in the context of multiprocesses instruments, e.g. multiple tracks or loops.

A historical example of an instrument which control interface (a keyboard) has not a direct physical link with sound process (here, strings).

Figure 1.13: Viola organista - Leonardo da Vinci - XVIth century
Television report from AFP: Aleksandra Gawlik - organ teacher from Krakow - when attending a concert of a Viola organista, the mixed instrument between violin and keyboard invented by Leornado Da Vinci : " I'm truly amazed about what I heard, **this is a total surprise to see a keyboard and hear a string instrument !** It's really an incredible impression, this is very surprising ! This is the first time I've heard such an instrument, it's fantastic !"

Virtuous innovations Not all technological innovations in DMIs are to be associated with disturbances for the audience experience. For instance, the improvement of haptic feedback for control interfaces is a great musician experience improvement that has no negative consequences for spectators. Some innovations can also benefit both performers and observers like devices that can dynamically and explicitly adapt their shape to fit richer interactions [171].

We will see next that the negative counterparts of performances with DMIs led researchers to adopt new approaches and put the spectator experience at a central place.

1.3.1.4 Interest, trust and understanding

The generalisation of interactive technologies in public contexts progressively invited the HCI community to preserve a space in the effervescence of its contributions in design of interactions for the question of the spectators of these interactions. Many contexts like museums, theatres and concert hall are concerned by public interactions and so the interest was not limited to the musical context (e.g.: [164, 165]). In his final talk at the Stockholm Music Performance

Symposium 2002, Andrew Schloss exhorted the community to consider the understanding of the audience of DMIs music performances. In particular, he claimed that the disappearing of cause-and-effect relationship between the performer and the instrument makes the performance seems like magic. "Magic is great; too much magic is fatal" [181]. One of the key idea of this call is to stress the term *performance* when proposing a *musical performance*. To do so, artists should give enough visual cues to spectators for them to regain interest in the performance, understand the interaction and trust that the musicians are actually performing.

1.3.1.5 Adopting the spectators' perspective

To our knowledge, one of the first initiatives to classify digital interaction with an approach exclusively based on the spectator experience is the one proposed by Reeves et al. [164]. The motivation of the approach can be resumed in one sentence: *"How should a spectator experience a performer's interaction with a computer ? "* .

Thus, Reeves et al. choose to name what is happening on stage with respect to the spectator perception. To do so, they [164] regroup the gestures with no distinctions under the term *manipulations* and the outcomes under the term *effects*. Then, they propose to describe the events on stage "*in term of the extent to which spectators experience a performer's manipulation versus their effects*". So, as depicted on Fig. 1.14, manipulations and effects can be *hidden*, *partially hidden*, *transformed*, *revealed* or *amplified*. With these markers set, four design strategies are exposed:

- *Secretive* : manipulations and effects tend to be hidden
- *Expressive* : manipulations and effects are revealed
- *Magical* : manipulations are hidden and effects revealed
- *Suspenseful* : manipulations are visible "*but effects only get revealed when the spectator gets to take their turn as a performer*"

Even if originally grounded in the larger context of observers of digital interaction, like in demos, museums or theatre performances, this taxonomy fits well with DMIs performances. The variety of interaction encompassed by the taxonomy can include musical interaction and their impact on spectator experience. For instance, *secretive* interactions are described as the ones that get revealed when the spectator eventually becomes a performer. The interpretation could be correlated with the expertise of spectators as they could perceive a manipulation or an effect as revealed (the interaction is not secretive any more) where novices would still evaluate

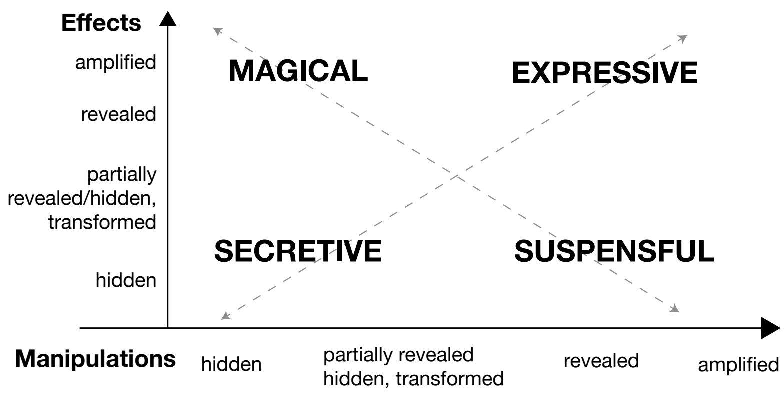


Figure 1.14: The correspondence between manipulations and effects lead to four approaches to designing the spectator's view: secretive, suspenseful, magical, expressive. (From [164])

interaction components as hidden, or partially hidden.

1.3.1.6 Synthesis

In this subsection, we saw that great creative opportunities for artists can also decrease the experience of spectators. Consequently, we saw that HCI and NIME communities decided to adopt new research approaches in favour of audience experience. In the next subsection, we will review the research initiative that contribute to the definition of spectator experience.

1.3.2 Defining the experience of spectators

The experience of spectators of music performances is an elusive phenomenon whose exploration and definition elicited numerous contributions from HCI, NIME, musicology or philosophical communities. In this subsection, we propose a selective review of potential components of spectator experience and their related research in a music context. Most of the time, the contributions do not ambition a definitive description of the music experience but rather focus on identified emotional or cognitive states which intensity or valence is modulated by performance components.

1.3.2.1 Definitions and components of spectator experience

Tension A definition of musical experience emerged in the works of psychologists and philosophers about the notion of *tension* [146, 125, 158]. As pointed by Vines et al. [199] : "Musical experience may be characterized as an ebb and flow of tension that gives rise to emotional responses". Here again, the link between musical experience and emotion [2]is underlined and includes physiological markers like cardiac or respiratory functions [125]. Many contributions (e.g. : [146]) associate the quite intuitive notion of tension with structural features of music like pitch range or harmonic relations but also with implicit expectations based on a musical genre experience as mentioned in [199]. Eventually, from the different definitions in literature, tension is not exactly a synonym but rather a proxy for the musical experience. Furthermore, the notion is more related to musical content and structure rather than a full experience of attending a music performance.

Flow Another component of the experience, or rather, a specific state of the experience is the flow. This immersive cognitive and emotional state is accessible to performance spectators [55] even if it is far more documented with cases of active behaviour like playing music, doing sport or working. As defined by Csikszentmihalyi [55], the flow is " The state in which people are so intensely involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it" [54]. This broad definition is greatly detailed in the work of Csikszentmihalyi and is completed with many others (e.g. : [67, 168], "being in the zone" [121]). Nevertheless, a consensus seems to merge on the conditions to access the flow and includes:

- the perception of situations in which challenges and skills are equal [67]
- thoughts that are led by a purpose and meaning [54]
- little distinction between oneself and the environment [54]

Prediction / Expectation When listening to music or attending a concert, being able to predict what is going to happen is a source of pleasure and satisfaction [101]. The expectation is also a characteristic that can induce the flow state described above [54] and plays a central role in creating musical tension and relaxation [125]. Importantly, the expectancy mechanisms are part of our capacity to feel and decode music (see Section 1.1.2.2) and, as pointed by Zatorre et al. [216] are crucial in setting up temporal and melodic expectancies at the heart of musical understanding [101]. Included in these mechanisms, our internal models help predict the outcome of musical interactions. A rewarding process is triggered when predictions are confirmed but can be frustrating when they are not [177].

[2] The study of the relation of music and emotions is a wide subject which detailed review is out of the scope of this document. We very briefly explored the related literature in section 1.1 and we saw how much music and emotions are tightly intertwined. In section 1.3.4, we will see that measurement tools used in the study of emotions, especially electrophysiological equipments, also contribute in the study of music experience (e.g.: [66]). Besides, emotional responses are of course not specific to music and can be elicited by an infinity of external like internal circumstances. We invite the interested reader to read the two main reviews on music and emotions [115, 146].

Communication design issues Another approach in the definition of the spectator experience is proposed by transposing a user-system communication approach in a performer-spectators perspective. In [9], Victoria Bellotti pointed five design issues – *attention, address, action, alignment, accident* – in the communication between a user and a system. Fyans et al. [75] proposed to transpose them to the perspective of observers of digital interactions, leading to five challenges for spectator experience:

- **Address**: How does the spectator know that the performer is directing communication to the system ?
- **Attention**: How does the spectator know that the system is responding to the performer?
- **Action**: How does the spectator think the user controls the system?
- **Alignment**: How does the spectator know that the system is doing the right thing?
- **Accident**: How does the spectator know when the performer (or the system) has made a mistake?

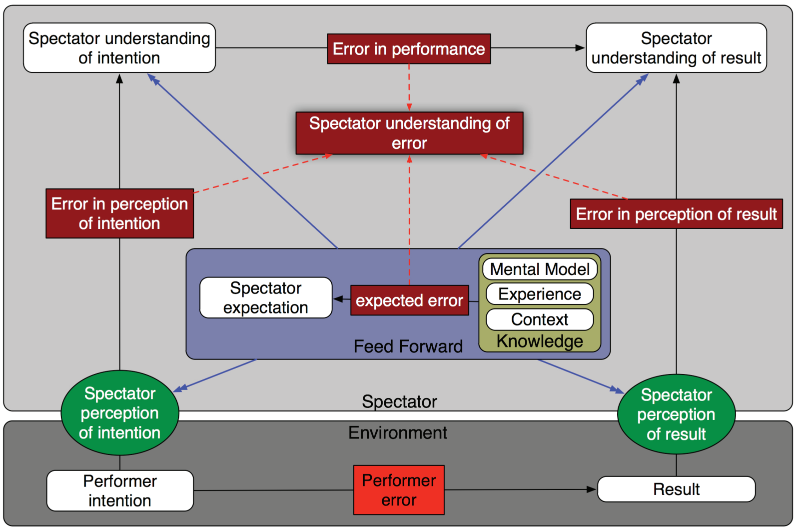
In the studies developed in this work, we will in our turn adapt the Bellotti-Fyans challenges to evaluate how spectator experience can be improved.

Error perception The possibility of an error is an element of risk that makes the performance more engaging [17] and a failure on stage can also be aesthetic [97]. Thus, the ability to perceive errors, to distinguish a desired outcome and its original gesture from a mistake, can contribute to the experience of spectators and their judgement of skills and thus virtuosity [78].

From the transposition of communicative design issues between a user and a system to a spectator's perspective (see above), Fyans et al. [76] choose to focus first on the impact of error understanding on the experience – i.e. the *Accident* item. They propose a model describing how different sources of error judgement impact the understanding of error by spectators of DMI performances. Their model is built upon the parallel analysis of the internal models of both performers and spectators.

We will see in this work that this modelling approach allows for interesting methodological advantages like the capacity to isolate and assess a potential component of the model.

Liveness In his book "Liveness: Performance in a mediatized culture"[2], Philip Auslander defines the liveness as "the kind of performance in which the performers and the audience are both physically and temporally co-present to one another". He completes the description of



Fyans, A. C., Gurevich, M., & Stapleton, P. (2009, June). Where Did It All Go Wrong? A Model of Error From the Spectator's Perspective. In NIME (pp. 171-172).

Figure 1.15: The possibility of failure makes performance more engaging. To better understand its role, Fyans et al. [76] proposed a model of spectator understanding of error including the internal model of spectators.

liveness with the *authenticity* [2] that is the character of a performance exposing perceptible effort and skills in opposition with pre-recorded contents. However the definition is diversely interpreted in the literature. As pointed by Bin [17], in HCI, a more common meaning of liveness is related to the perception of causality that we specifically address next.

1.3.2.2 Audience experience: instruments characteristics

From these descriptors of music experience, characteristics of instruments that support the spectator experience can be extracted. Most of the time obvious in acoustic instruments, these characteristics are challenges for the NIME community. Here is a brief review of the main definitions reported in the literature.

Expressivity Although there is no unique definition in the literature, *expressivity* can be described as the capacity of an instrument when performed to convey emotions and meaning. We saw (see 1.3.1.2) that performances with acoustic instruments, because of musician's gestures, are particularly rich in terms of meaning and emotions. When performances include DMIs, when gestures can be less meaningful, *expressivity* becomes a crucial design issue for performers and spectators (See the interesting work of Astrid Bin for review [17]).

Familiarity Familiarity is the feeling of knowing about the behaviour and the possibilities of an instrument. It can be associated with other concepts in the literature as *mental model* [75] or as an extension of *transparency* [69] (see below). It is a rather large component of the experience as it provides the spectator the ability to detect the intentions and the virtuosity of the performance of the musicians as well as their errors.

Transparency Defined by Fels et al. [69], *transparency* is the resulting property of an instrument which behaviour is well understood by musicians and audience. Specifically, it " provides an indication of the psychophysiological distance, in the minds of the player and the audience, between the input and output of a device mapping " [69]. *Transparency* can be associated to the notion of *intimacy* developed by Richard Moore [150], even if first intended with a musician perspective, as they both underline the importance of a clear relation between the musical possibilities of an instrument and the physiological capabilities of the performer. Thus, in a spectator perspective, *transparency* is the ability of an instrument to make the audience understand which control produces which sound.

Instrumentality The notion of instrumentality encompasses the criteria that makes an device an instrument. For Cadoz, instrumentality is present " when, for any object (real or simulated) and for any interaction (in real world, in virtual world or in mixed ones), the variables that represent or characterize them are mechanical (for example forces and displacements), and the processes respect energy consistency." [31]. With the evolution of instruments and the disappearing of physical constraints, the notion evolved to include criteria also based on the audience perception like liveness, expressiveness, causality or expectation [53]. As the interest for audience perception increases, instrumentality becomes a criterion for the evaluation of novel digital instruments [3, 27].

1.3.2.3 Synthesis

We saw that spectator experience can be accessed through many notions. Sometimes their meaning overlap or describe the same phenomenon but either from the instrument perspective or the spectator's. For instance familiarity and transparency encompass the same attributes (intention, causality) the former as a perceptive characteristic (in familiarity), the latter as a design capability (in transparency). Besides, it is not clear if the listed notions are descriptors or components of spectator experience as some notions are constitutive of others like the role of familiarity in the emergence of flow.

Still, among the multiple definitions and components, the notion of causality seems central. As it is embedded, with variable saliency, in all the notions presented above, we address it specifically in the following section.

1.3.3 Attributed agency and perceived causality

We saw in the last subsection that the literature of spectator experience is diverse and includes multiple potential components. Among them, the perception of causality between the musician's gestures and the sound is recurrent.

In this subsection, we will see that perception of causality is a prerequisite of higher-level considerations. We will show how the notion of apparent causality can be explored through the cognitive framework of *agency* and its transposition to *attributed agency*. From there, we will draw methodological opportunities to explore this component of the experience in a controlled environment.

1.3.3.1 A prerequisite for higher-level perception

A common point to the previously reported definitions and models of spectator experience is the role of the perception of the causal link between the behaviour of musicians and the sound. In fact, many contributions (e.g.: [15, 68, 26]) expose it as a prerequisite to identified components of the experience like evaluation of skills, or perception of errors.

For Jon Croft [53] perceived causality between performer's action and the sound is a key component of liveness in DMI performances. This proposition is supported by research underlining that gesture-sound causality is required to understand the functioning of an instrument, experience liveness, and access higher-level evaluative concepts like skills or virtuosity [15, 68].

Liveness through perceived causality In order to improve the liveness of DMI performances, Berthaut et al.[13] ground the design of visual interactive content (i.e. visual augmentation) on a better understanding of the underlying mechanisms of perception of causality by spectators. The authors propose an approach based on the *apparent mental causation*, a model of *agency* developed by Wegner et al [208]. These notions are primary related to actions we conduct ourselves but Berthaut et al. propose to transpose these self-centred notions to the perception of actions pursued by others. Before focusing back on Berthaut's approach, we need to present a few details about the concept of agency.

1.3.3.2 Agency

Agency, or sense of agency, or self-agency, is our ability to consider ourselves as authors of our own actions. It refers to the link between execution of voluntary movements and the modifications they apply to our environment [95]. Disturbed in patients of mental afflictions like schizophrenia, it is a common ability for healthy people and stands *as a basic and constant underpinning of our interaction with the world* [193].

The central position of agency in our relation to the environment triggers a lot of debates and research contributions from diverse fields like philosophy (e.g.: [81, 95]), cognitive neurosciences (e.g.: [94, 58]), or psychiatry in the understanding of schizophrenia (e.g.: [109]). Depending the context and the framework, we will see that it can be address as a judgment or as a feeling.

Two main models From this dense literature, two main models of agency emerge: the comparator model [73, 22] claims that agency emerges thanks the same *predictive* mechanisms that support voluntary action. The second model is the one developed from Wegner's work. We will call it *postdictive model* [193] as it explains agency emergence from post-hoc inferences (see *Apparent mental causation* above).

The models are often challenged on their efficiency to explain behavioural or pathological observations. Their detailed presentation is out of the scope of this document, instead, we propose to introduce to the reader their main characteristics on which we will later ground our research approach.

Predictive - Comparator model The main claim of the comparator model [212, 73, 21] is that feeling of agency emerges from the correlation between a desired goal and sensory-motor

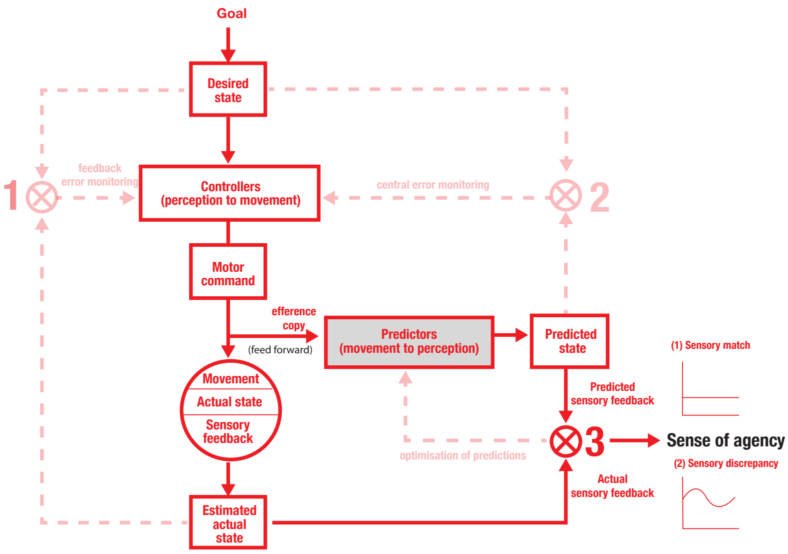


Figure 1.16: The neurocognitive mechanisms underlying the sense of agency [193]. There is a consensus on the role of comparator 3 in the emergence of sense of agency. Note that the computational model presented here, with feedback control loops and feed-forward comparators (1,2,3) is the one used to explain the underlying mechanisms of voluntary action.(adapted from [193, 211, 22, 73])

feedbacks. The comparisons between predicted states and actual outcomes are conducted thanks to *efference copies* that are copies of the motor commands sent to motor control. These copies allow for parallel motor simulations to predict sensory consequences and correct the action toward the desired goal thanks to continuous comparisons (feed forward). In the comparator model, sense of agency arises when the sensory feedback of the ongoing action are close to these motor simulations (Fig 1.16).

To sum up, in the comparator model, agency is a *feeling* of being in control based on predictive mechanisms.

Postdictive model Introduced by Wegner & Weathley [208], *apparent mental causation* is a theoretical proposition to explain our feeling of free will, of voluntary control, when we act in the world. The main idea proposed by the authors is that we infer causality between an action and its consequence from the a posteriori integration of three criteria – priority, consistency, exclusivity – introduced like so: "When a thought appears in consciousness just before an action (priority), is consistent with the action (consistency) and is not accompanied by conspicuous alternative causes of the action (exclusivity), we experience conscious will and ascribe authorship to ourselves for the action" [208]

This definition exposes apparent mental causation as a post-hoc inference rather than an immediate feeling emerging from sensory-motor cues. It also highlights the fact that this integration is a prerequisite to our ability to ascribe the authorship of our actions to ourselves, i.e. sense of agency. In other words, the natural feeling of being in control during an action could be a mind *trick* [207] as agency occurs *after* the action.

To sum up, in the postdictive model [208], agency is rather a judgement based on a posteriori criteria: priority, consistency, exclusivity.

1.3.3.3 Attributed agency: a first insight

In previous sections, we saw that the perception of causality is crucial in many definitions of spectator experience. We also explored the work of Wegner et al [208] who link apparent mental causality to the wider phenomenon of agency, that is our ability to consider ourselves as actors of our actions. Besides, we saw that Wegner et al. consider agency as a postdictive mechanism where agency is correlated to the action-posterior consideration of three distinct criteria.

From these theoretical background, Berthaut et al. [15] propose to link the liveness of DMI performances to the causality between performer's actions and sound as perceived by the audience. Furthermore, the authors propose to study this perceived causality by transposing Wegner's self-agency model (I am author of my actions) to the ability to ascribe authorship to someone else's actions, a phenomenon authors call **attributed agency**.

With the transposition to the observer perspective, the criteria for the emergence of agency (see section 1.3.3.2) become criteria for the emergence of Attributed Agency [15]:

- *Priority*: originally refers to the necessity of a prior conscious thought before the action. For spectators, the temporal order can be associated to the temporal nature of the performer's actions in relation to the musical outcome.
- *Consistency*: originally refers to the necessity of the action to be consistent with the prior thought. For spectators, this can be transposed to the consistency between actions and the musical outcome. Unintuitive mappings can link gestures of any kind to any type of sound parameters with changes in scale, linearity or continuity.
- *Exclusivity*: originally refers to the number of possible causes of an action. For spectators, exclusivity issues can appear with multiprocess DMIs where the origin of sound is not clear between autonomous processes and musician's gestures. Complex mappings can also impact exclusivity when multiple changes occur in the sound while multiple gestures are being made.

We will see in section 1.3.5 that the transposition of Wegner's theory to the perception of causality by spectators can offer interesting design guidelines for the creation of techniques to improve the experience of spectators.

1.3.3.4 Attributed agency: the need for richer models

The seminal work of Berthaut et al.[15] on Attributed Agency is a distinctive step in the comprehension of spectator experience. As Attributed Agency is the perception of the authorship of someone else's actions, this a particularly relevant approach to explore the relation of an observer and a performer. Besides, the criteria from which Attributed Agency emerges constitute methodological items that can conveniently be addressed individually.

However, we saw that the postdictive model of agency from which Berthaut et al. conceive Attributed Agency is not the only relevant model of agency and it may too strictly exclude predictive aspects supported by the *comparator model*. Another argument in favour of the relevancy of the comparator model as precursor of Attributed Agency is that it shares mechanisms with the internal models of interaction we often refer to in this document. Additionally, taken individually, these two models are not sufficient to fully explain agency and more recent research expose the relevancy of a weighted composition of them [193].

Besides, for the sake of clarity we exposed only two models of agency but the subject elicit many interesting approaches that could complement the consequent definition of Attributed Agency. For instance, personal background beliefs could influence our judgement of someone else's agency ([80, 189] reported in [193]). If one personally knows the performer and knows that they always practices their performances with a great contribution and few autonomous processes, one can perceive a clear Attributed Agency during a performance even if live conditions and obscure mappings disturb the perception of causality.

In this work, Attributed Agency is a central theoretical ground. We will extend the previous works with Attributed Agency and propose to enhance the underlying components upon which previous research stand to capture richer aspects of our capacity to associate a musical outcome to a performer's behaviour.

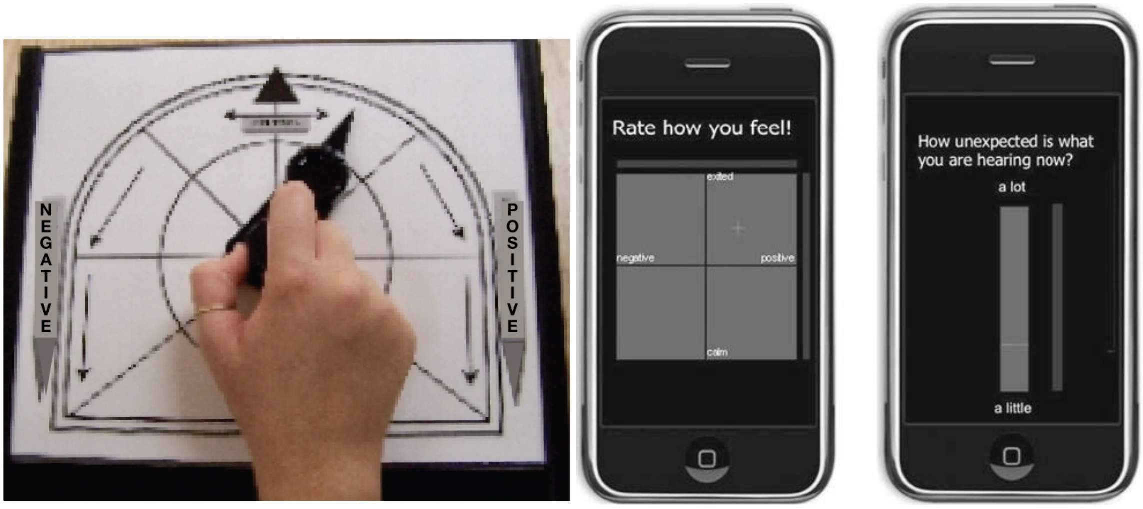
1.3.4 Measuring the experience

The measurement of spectator experience is a sensitive operation as the method employed can impact the experience we try to evaluate. Asking a participant about their experience during a performance can be disturbing and interrupt the moment. Conversely, questioning a participant after the performance may elicit biased answers as the participant should live again from their memory the experience they had long minutes ago and report a revisited feedback. Objective non verbal measurements like electrophysiological signals are useful to lower those biases as they can discreetly provide data while the participant is enjoying the performance. Unfortunately, the interpretation of objective signals is not as rich and specific as introspective reports. Indeed, the analysis should be conducted with precautions as no marker, to our knowledge, can precisely and exclusively variate in function of the dependent variable we target, especially with high level aspects like musical experience.

In this section we will review measurement methods and targeted components of spectator experience. Most of the time, researchers base their work on introspective reports of participants after a performance [19, 213, 11, 35] and gather fruitful knowledge. Alternative initiatives are also presented and discussed along with the ongoing debate of "in the wild" versus "in the lab" experiments.

Questionnaires and interviews Fyans et al. analysed interviews of participants after having watched videos of performances to address the perception of skills[74], mental models, intention and error perception [78]. Benford et al. [11] use a questionnaire and ask spectators of two performances to rate their enjoyment and their understanding of the performance, after the first performance and again after the second performance. The authors completed the study with voluntary video interviews "to capture more detailed thoughts and reactions". Brown et al. also use questionnaires to evaluate the perception of dual, i.e. human and machine, agency in musical performances [27]. Attributed agency is also explored by Berthaut et al. [15] who asked participants through a computer questionnaire their rating of the influence of the performer's gestures as well as their confidence in their answers.

Spectator experience can also be addressed to evaluate DMIs (e.g.: [128, 3, 213]). In [3], Barbosa et al. use a questionnaire as an evaluation method to address **the comprehension of a DMI** by the audience. The authors base their questions on the **Bellotti-Fyans challenge**s (see 1.3.2.1) and explore causality with questions like "How understandable are the actions made by the user for interacting with the system?", effect comprehension: "Did the system provide enough audiovisual information for the audience to understand what is happening between the user and it?", mapping comprehension: "How clear is the relationship between



(a) A continuous response digital interface. A one-dimensional non verbal measurement device used in the study of musical tension. (adapted from [185])

(b) Continuous rating interface - Left: two-dimensional scale (valence - arousal) for emotional rating – Right: one-dimensional scale for unexpectedness of musical events (From [66])

the user's actions and the resulting sound?", Intention comprehension: "How successful was the user to express himself using the system?"; "Was the user's intention well understood?", Error comprehension: "Were the system's errors perceived (e.g. technical problems and software bugs)?"; "Were the user's errors noticeable?"[3].

Continuous responses We saw previously (see sub-section 1.3.2.1) that tension is a potential component of the musical experience, preferably described as a proxy of the experience. The tension judgement is also associated to methodological approaches that advantageously rely on a single dimension continuous non verbal measurement thanks to a device named continuous response digital interface (CRDI) [139]. As depicted on figure 1.17a, while listening to music, participants are asked to move the CRDI indicator to the right to indicate when they perceive a tension or to the left to indicate when they perceive a release. Alternatively, in [6], authors use a linear slider from a USB MIDI controller.

Without the memory biases of post concert questionnaires and relatively little disturbing, continuous reports during performances are a relevant compromise to measure spectator experience, they can also take the form of votes. Bin et al. [19] explore the connection between error perception and enjoyment thanks to post performance questionnaires but also by collecting data during the performance via a dedicated mobile app and a simple two-buttons interface.

Objective signals Alternatively to subjective reports, physiological measurement is a promising way to gather objective data about the audience experience. In [43], Chapados & Levitin showed that tension judgement ([200]) and electrodermal activity were particularly correlated when spectators could watch and listen to a music performance compared to unimodal conditions.

Complementary measures A single electrophysiological signal is often not sufficient to produce rich interpretations. Besides, it can be precious to accompany objective measurements with introspective reports to clarify the interpretation of raw signals. Egermann et al. [66] provide quantitative data to the study of the link between expectation and emotions thanks to an extensive set of physiological, continuous and introspective measurements:

- Psychophysiological measurements:
  - device: respiration belt — dependent variable: respiration
  - device: photoplethysmograph — dependent variable: blood volume pulse
  - device: finger electrodes — dependent variable: skin conductance
  - device: electromyography sensors — dependent variable: expressive muscle activations
- Continuous subjective measurement:
  - device: touch screen with a two-dimensional scale (valence - arousal) — dependent variable: emotion state
  - device: touch screen with a one-dimensional scale — dependent variable: unexpectedness of musical events
- Questionnaires and interviews:
  - device: clipboard questionnaire — dependent variable: familiarity with the performance
  - device: paper questionnaire — dependent variable: socio-demographic characteristics, musical training, music preferences
  - device: cameras & microphones (x2) — dependent variable: behaviour

Such extensive measurements provide rich data materials with the counterpart of a proportionally extensive processing and analyses (See [66] for details).

Insights from related fields Here are examples that leave the musical context but the approaches could be exported to music performances. They also underline the relevancy of objective measurement of spectator experience. Wang et al. [206] developed their own

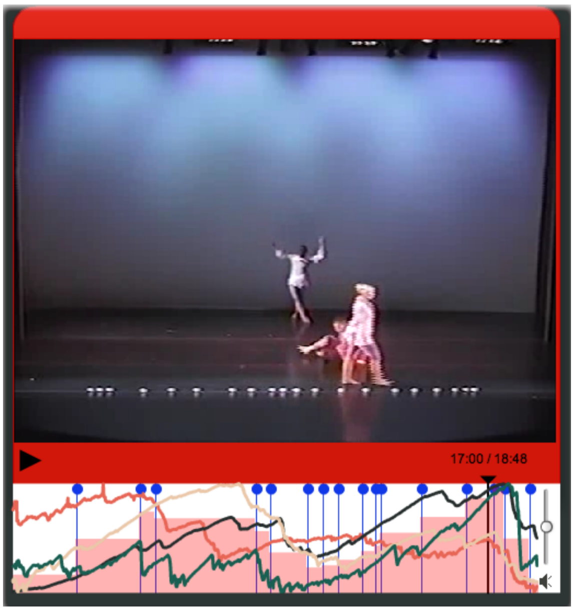


Figure 1.18: Electrophysiological measurement can inform about the spectator engagement. Here, the galvanic skin response of a spectator watching a dance performance.(From [129])

electrodermal activity sensor to measure the level of engagement of spectator in a theatre context. An interesting approach was developed by Latulipe et al. [129] to measure the engagement of spectators of a dance performance. Beyond the measurement of electrodermal activity, Latulipe et al. includes a secondary feedback of experts (choreographs) who watched the video accompanied of the signals previously measured on participants (See Fig. 1.18).

1.3.4.1 In the wild - in the lab

In the study of audience experience, most research rely on an "in the wild" approach [41, 10], utilising observations, post-performance questionnaires and guided interviews of spectators. In contrast, few "in the lab" experiments have been conducted on audience experience. They all rely on videos of performances to address components of the experience. In a controlled environment, experiments from the lab allow for refined measurements with fewer distraction for the participants.

However, data analysis methods for controlled experimentation can be considered as a risk for misleading or overly narrow results [153]. Thus, experiments should mix objective and subjective trials to limit the bias often met in field studies [29]. Another limitation of the 'in the lab' approach naturally lies in the evaluation of a live phenomena in a controlled environment. In this context, important aspects of the experience of live music are missing, from the social environment to the ceremonial of the concert hall. Still, these rich aspects are more grounded in the field of sociological studies, the interested reader is invited to read the related literature [157, 63, 28] on the subject.

1.3.4.2 Synthesis

We saw that the measurement of spectator experience can be supported by introspective reports, continuous measurements and objective data capture. Some components are relatively straight forward to assess like the tension, while the measurement of others, like understanding or expectation, induces methodological challenges. Often conducted "in the wild", studies can also take place "in the lab" in less ecological but more controlled environments for refined measurements. In fact, it appears from this review that no unique methodology can capture the entire reality of the experience of participants without biases and limitations. Therefore, it seems relevant to ambition the study of spectator experience with complementary approaches, mixing objective and subjective measurements.

1.3.5 Improving the experience

These last fifteen years, the growing interest for spectator experience led researchers and artists to work on techniques to improve the experience of audiences of DMI performances. Numerous initiatives include design guidelines, demos, explanations of the instruments or visual augmentations. In this section, we will review these contributions and expose their respective advantages and limitations.

1.3.5.1 Building a repertoire

We saw in section 1.3.2.2 that the familiarity spectators have with a DMI can positively impact their experience. Thus, a first way to improve the experience is to increase the familiarity of people with a particular instrument. It can be initiated by promoting the use and dissemination of the instrument and building a repertoire of compositions around it. Once a majority of spectators have seen the instrument played by many musicians in different contexts, or even practiced it themselves, they are aware of its potential for musical expression and of its behaviour. Thereby, the familiarity issue does not exist any more. This kind of "natural familiarity" is definitely effective but requires a large amount of time and energy to be achieved. Moreover, this method is not compatible with the very idea behind DMIs and the exponential creativity they embed. Whether a musician wants to evolve its instrument or let other musicians modify it, as soon as the instrument changes, the whole process of natural familiarity needs to start all over again.

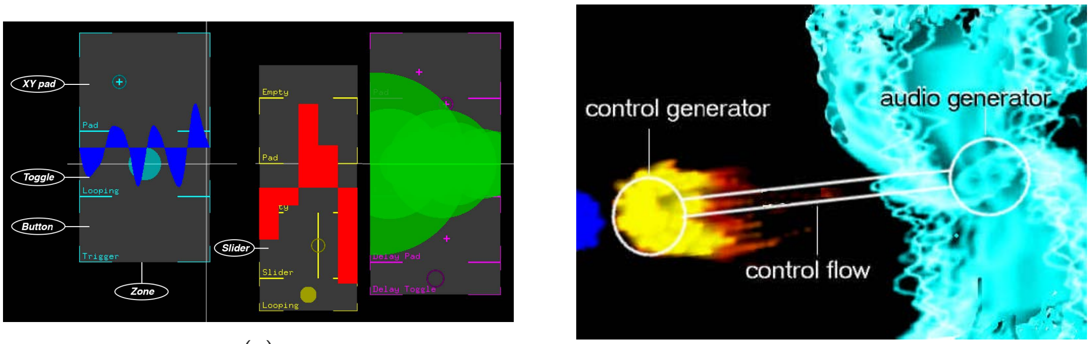
1.3.5.2 Explanations

Spectator experience lies partially in the understanding of the operations of a DMI. Thus, demonstrating the behaviour of the instrument can be a valuable solution. Building the familiarity with a pedagogical method is a simple way to make an audience understand what is going on on stage. Before, after and even in breaks during the performance, musicians can explain how their instrument works. Prior hands-on demos, where the audience can actually play the instrument, are also a good way to increase their experience during the actual performance. However, both these methods trigger some reservations as the technical understanding of an instrument is not necessarily linked to a better appreciation of the performance. For example, Bin et al. [18] demonstrate that explanations before the performance do improve the understanding of the instruments and its mechanisms but do not increase either the appreciation or interest. Besides these results, the audience may forget or unconsciously misreport important details from the performance [136] and the demos may not be possible when dealing with a large audience.

1.3.5.3 Designing approach

Another strategy to increase the experience of spectator lies in the design of the instrument itself.

Designing for transparency The idea is to expose clear and intuitive mappings so the link between the gestures and the sound modifications operated by these gestures are obvious for the spectator. We saw that this fluency of perception of mappings is called "transparency" (See 1.3.2.2). Aiming for transparency from the very first steps of the design of a musical device can lead to "easier to perceive" instruments [152]. Following that lead, an interesting way to increase the transparency is the use of metaphors [69]. In that case, the common background of inexperienced audience, their general knowledge, are used as a mould to grow new knowledge about an instrument mapping. For instance, the timbre of a sound can be modified as the shape of an associated graphical representation that is getting sculpted. However, the instruments designed according to these specifications tend to maintain the familiarity we have with acoustic



(a)

(b)

Figure 1.19: (a) Example of AVUI built with the toolkit proposed by Correia et al. The interface of the instrument used by the musician is composed of audiovisual elements that allow the audience to better understand the performance. (From [50]) – (b) Interface with visual feedback of graphical representations of sound synthesis (organic shape) and controls (straight lines).(From [113])

instruments and physical laws and so may require the performers to restrain their expressivity with respect to all potential combinations of gestures, sensors, mappings and content usually available for DMIs.

Audiovisual interfaces The restrictions of design for transparency can be lifted by audiovisual interfaces. Indeed, they are designed to expose the instrument capabilities through visual feedback and facilitate the comprehension of both musicians and audience (e.g. :[113, 50]). For Jorda et al. [113], visual feedback enables the audience "to watch the music and how it is being constructed" and therefore improves the understanding of spectators in addition to exciting elements they provide to the show through video projections. In [50], Correia et al. propose to intricate sound and images inside a same interface called AVUI (AudioVisual User Interface). In the sense of musicians becoming instruments makers that we saw earlier (See Section 1.2.1.7) and 1.2.3), the contribution of Correia et al. includes a toolkit and guidelines to implement and customise these audiovisual interfaces.

1.3.5.4 Haptic augmentation

Haptic feedback is a promising way to amplify musicians' gestures [1, 196]. In [1], Joanne Armitage addresses the issue of gestures in live coding. Making music exclusively by coding through the small surface of a laptop can indeed lead to expressivity issues for spectators [181]. In response, the author developed a haptic system called 'Key' that transduces the keystrokes the artist trigger while coding into variable vibrations that spectators can feel by holding a device. Doing so, Armitage bridges a highly mediated practice with physical signals to enhance the embodied relation of spectators with the performer. Additionally, haptic feedback can be used in parallel of DMI performances to improve the audience comprehension (significantly at low tempo only) [195] but it can also be used with augmented traditional instruments to push spectator experience in unexplored crossmodalities [196].

1.3.5.5 Visual augmentations

A last strategy explored by both artists and researchers is to augment the performance with information that complements the audience experience without modifying the performer's interface. This augmentation can be done for various goals regarding the audience experience, such as increasing the audience enjoyment, engagement or comprehension, e.g. turning *magical* interfaces into *expressive* ones [163]. It may also take many forms that mostly aim at providing optimised graphical representations of the ongoing interactions.

Revealing the inner mechanisms of DMIs Berthaut et al. propose [14] to extract musical interactions parameters and graphically represent them to reveal the inner mechanisms (controllers, mappings, sound processes) of DMIs (See Fig. 1.20). The visual augmentations are then projected or included via augmented reality setups and overlap the device so spectators can keep their attention on the interaction space rather than on a deported screen. Hence, by recreating the causal link between the performer's gestures and the sound, visual augmentations support an increase Attributed Agency for spectators.

Additionally, the setup also contributes to clarify the performer's intentions and expressiveness by exposing the details of the interactions. From there, the augmentations also enable the distinction between autonomous processes and actual live music production. Such a discrimination between live and pre-recorded music is crucial in the experience of live music. It provides cues to audience that the performance is actually *live* and thereby comfort their interest and their trust in the musician.

Visual augmentations for orchestra Following the seminal work of Berthaut, Perrotin et al. [161] propose a projection of visual augmentations behind the performers of an orchestra. The augmentations inform the audience about the role of each musician in a voice synthesis performance by associating the performers and the parameters they control with a colour. The information displayed can be related-to-gesture parameters (Fig. 1.21b) or related-to-sound

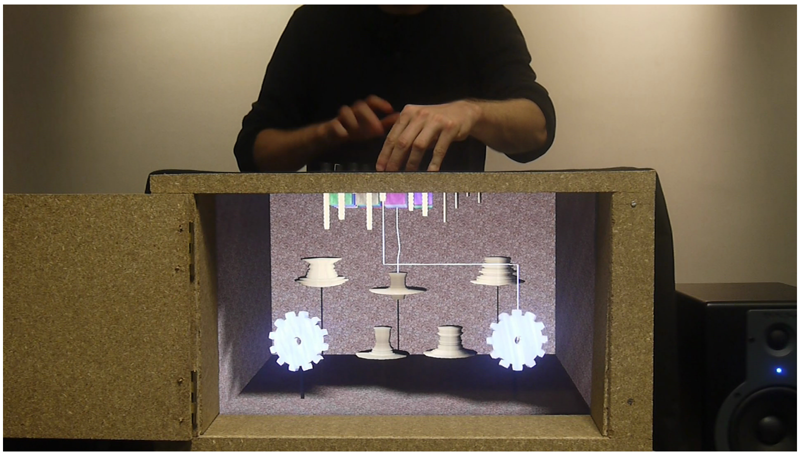
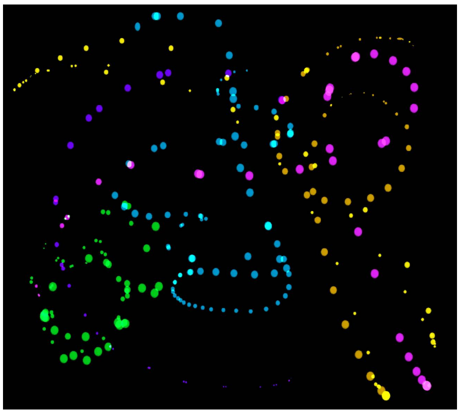
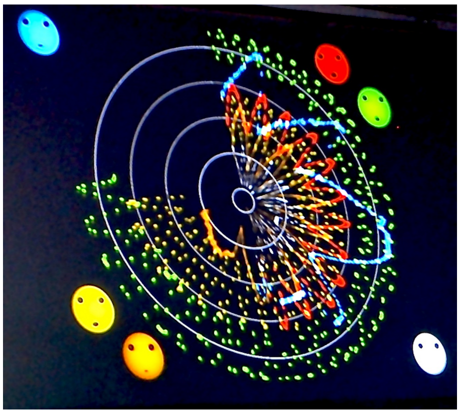


Figure 1.20: With Rouages, Berthaut et al. propose graphical representations of musical interactions parameters - Berthaut et al. (2013)



(a) Visual feedbacks of related-to-gestures parameters of an orchestra. Performers control a voice synthesis with a stylus on a tactile surface. The trajectory of each stylus is represented with the color associated to the performer. (From [161])



(b) Visual feedbacks of related-to-sound parameters of an orchestra. The circular lines represent the pitch versus the time. The avatars represent the articulation control. Each colour corresponds to a musician. (From [161])

parameters (Fig. 1.21a). As depicted on Figure 1.21b, additionally to the evolution of notes through time, authors provide embodied cues through avatars whose expression metaphorically indicate the articulation one should do to emit an approaching sound of the one controlled by the musician.

Multichannel augmentations In [11], Benford et al. use visual augmentations to inform the audience about the progression in a non-linear musical piece. The music is actually composed of micro pre-recorded sequences triggered by the pianist and is intertwined with the poetical narration of a journey up a mountain. Beyond the projections of visual augmentations on stage, a mobile app, synchronized with the performance, provides didactic information to the audience like the previous steps taken by the artist in the journey, her current location on the map and the potential next step. Finally, the engagement of the audience is maintained even after the performance thanks to an online retrospective archive to review the different versions of the performance.

1.3.5.6 Synthesis

In this section, we reviewed the contributions of artists and researchers to improve the experience of spectators of DMI performances. In particular, visual augmentations expose many advantages as they do not constrain the artists neither in their exploitation of the technical possibilities of DMIs, nor in the design they choose to deploy. Furthermore, by providing cues in a modality deeply linked with experience (one neuron out of two is dedicated to visual processing in the brain), visual augmentations have decisive potential to fill the perceptive gaps for spectators of digital interactions.

1.4 Our approach

1.4.1 Context of research

In this chapter, we saw that our relation to music is a deeply embodied experience, especially in live conditions. We overviewed the evolution of digital musical instruments (DMIs) and explored how they offer infinite creative opportunities for artists by erasing the physical constraints of sound production and control. However, without this link between the musician's gestures and music, we observed that spectators lack embodied cues and can face difficulties in enjoying the artists' contribution, their intention or their virtuosity.

In our review of contributions to address these issues and better understand such an elusive phenomenon, we saw that spectator experience embeds multiple components with no clear hierarchy of their individual influence. We found that, consequently, its measurement raises methodological challenges that would benefit from a complementary vision between field and lab studies.

Furthermore, we exposed augmentation techniques that both artists and researchers developed in order to improve the spectator experience. We saw that these techniques can take many forms and rely on variable strategies to support the better accessibility and enjoyment of audiences. Although these techniques may definitely benefit the spectators of digital performances, to our knowledge, their efficiency has not been formally compared and no insights have been provided about the components of the audience experience that they respectively affect. Besides, a formal analysis of these techniques has not yet been conducted that would greatly facilitate their development.

1.4.2 Contributions

To address these issues, we propose to develop a multidisciplinary approach grounded on conjoined frameworks of HCI and cognitive sciences. Thus, in the studies described in this work, we develop theoretical, methodological and practical aspects towards a better understanding of spectator experience.

In complement to the main outline of the thesis, we propose to the reader an alternative navigation in the document that can be presented in a global threefold contribution.

1.4.2.1 Defining - theoretical contribution

To comply with its multiple facets, we believe that the definition of spectator experience would benefit from a modelling approach. In a first study, we propose to evaluate the role of Attributed Agency among other components in the emergence of *familiarity*, an important aspect of spectator experience (chapter 2). We also propose a decomposition of spectator experience based on subjective and objective elements which relevancy significantly appears in the literature (chapter 4).

1.4.2.2 Measuring - methodological contribution

To contribute to the methodological challenge of measuring the spectator experience, we propose a controlled protocol that aims at evaluating the efficiency of augmentation techniques on subjective and objective components of the experience (chapter 4). In particular, we investigate the perception of the contribution of performers (Attributed Agency) compared to the one of autonomous processes. We then refine our protocol to take into account the participants' expertise and the impact of controlled levels of detail in the augmentations (chapter 4). Finally, we present a preliminary study that evaluates the relevancy of grip force as an objective marker of Attributed Agency in the context of musical performances (chapter 5).

1.4.2.3 Improving - practical contribution

This work highlights the role of visual augmentations in the improvement of spectator experience, especially in terms of Attributed Agency. To help future research and pave the way for better augmentations, we propose three contributions. We introduce a taxonomy of spectator experience augmentation techniques (SEATs) and use it to analyse previous contributions (chapter 3). We present visual augmentations with selectable levels of detail for a more personal experience (chapter 4). And finally, we present a conceptual pipeline that generates personalised augmentations based on the integration of physiological signals from spectators and interaction parameters of performers (chapter 5).

La musique exprime ce qui ne peut être dit et sur quoi il est impossible de rester silencieux.

Victor Hugo

2 The role of attributed agency in familiarity

Main contributions

- Decomposition of familiarity in evaluable dimensions
- Demonstration of the importance of Attributed Agency in spectator experience
- Experimental contrasts between acoustical instruments and Digital Musical Instruments in terms of spectator experience

> The contribution of this chapter was reviewed and corrected following a rejected submission to the 12th conference on Creativity & Cognition (2019).

2.1 Introduction

Digital Musical Instruments (DMIs) used in public performances are often customised by the performers, if not created from scratch. As we saw in last chapter, the modularity of DMIs offers great creative opportunities for artists to develop their own controllers, mappings our sound processes. In the end, the DMIs they conceive are integrated in the cognitive models of musicians so they precisely know their capabilities, mechanisms and autonomous processes.

From a spectator point of view, instruments used in DMI performances are most of the time unknown. We noted in section 1.3.1.3 that the lack of familiarity with an instrument can trigger difficulties for the audience to perceive the causal link between the performers' gestures and the sound produced on stage. Therefore, without a reliable mental model of the interactions, the experience of observers unfamiliar with the instrument is directly impacted.

In response to this issue, we observed in section 1.3 that artists and researchers proposed solutions to improve spectator experience through the design of the instruments or dynamic augmentations. We saw in section 1.3.4 that these research are supported by measures of spectator experience and related aspects like enjoyment, error perception and Attributed Agency that appears to have a recurrent position in the literature.

2.1.1 Measuring components rather than consequences

However, measuring such features might not be sufficient if the overall *familiarity* of the audience with the interaction is not understood correctly. Indeed, the perception of *expertise, skill, error or intention* of the performer can be considered as **consequences of the familiarity** with a musical interaction **rather than dimensions** of the familiarity itself.

The more familiar we are with an instrument, with the way the sound is controlled and produced, with the context of the performance, the better we can perceive features like errors, skills, intention of the musician and then assess the quality of the performance. Consequently, while they can help measure the audience experience, with which they are correlated, these features do not provide insights on the reasons for familiarity.

We think it can be insightful to investigate the components which, when changed, have a more direct impact on the overall familiarity, and in turn on the perception of skill, error or intention. Furthermore, familiarity is by definition (knowing an instrument and its possibilities) highly dependent on the spectator's profile, including their expertise in technology and musical genres.

2.1.2 Benefit of controlled conditions

Additionally, existing research has focused on the study of full performances, which participants were asked to attend or watch videos of. During these performances many factors of familiarity may vary with no possible control of interference with each other, such as the order of gestures performed, the mappings used, or the specifics and constraints of the musical genre. In this uncontrolled context, it is not possible to precisely isolate the effects on spectator experience of the different components of a performance that are the mapping, the instrument appearance or the type of gestures. Consequently, decisions on the design of an instrument or performance are more difficult to take.

2.2 Our approach

The work presented in this chapter aims at a better understanding of the familiarity, a key component of the experience of spectators of DMI performances. To this end, we propose a decomposition of *Familiarity* which includes *Attributed Agency*, *Instrumentality* and *Expertise*. We address these components through an internet study with controlled stimuli[1]extracted from a dataset of simple musical interactions (Unit Perceived Interaction - UPIs - See 2.2.2). Finally, we extract models from the data and compare them to identify the most relevant approach.

2.2.1 Familiarity dimensions

As other high-level cognitive abilities, familiarity is not a unitary notion. It can be decomposed into components that differ by their nature, their inner rules as well as the type of information they convey. These diverse aspects relate to perception, expertise or attention and all contribute to the constitution of a mental model of the instrument for musicians and spectators (See section 1.1.2.2).

Taking this into account, our approach is to decompose familiarity into dimensions that we can both evaluate and improve. Based on results in HCI, NIME, and cognitive sciences that have explored various aspects of spectator experience (see section 1.3.2), we propose to

[1]All stimuli are available online. Stimuli used in the study: `https://o0c.eu/0WS` — Full dataset: `https://o0c.eu/1WS`

decompose familiarity into five components: *Attributed Agency*, *Instrumentality*, *Instrument expertise*, *Musical culture* and *Musical genre expertise*.

These in turn contain dimensions that can be evaluated independently:

2.2.1.1 Attributed agency

As detailed in section 1.3.3, *Attributed Agency* can be approached by a transposition to the spectator perspective [15] of the apparent mental causation, e.g. the judgment of causality of one's action, defined by Wegner et al [208]. As the original model, Attributed Agency can be decomposed in three dimensions:*Priority*, *Consistency*, *Exclusivity*

The consistency aspect is also closely related to the notion of naive physics, that is a set of innate or common knowledge about the physical world like gravity or friction that can also be used in digital interfaces as physical metaphors [104].

2.2.1.2 Instrumentality

Instrumentality relates to common knowledge that allows one to predict the range of sound possibilities of a musical instrument from its appearance. It is composed of two dimensions: the *composition* , i.e. shape and material, and *behavior*, i.e. mechanisms and degree of autonomy from the musician's actions.

2.2.1.3 Instrument expertise

Instrument expertise relates to the exposure the spectator has had to the instrument, from a first-time observer to an expert player.

2.2.1.4 Musical culture

Musical culture corresponds to a basic knowledge of musical theory, that can for example be used to represent pitch as vertical position of a graphical element, i.e. mimicking a staff.

2.2.1.5 Musical genre expertise

Musical genre expertise is composed of dimensions that correspond to the familiarity with the specifics of a genre, such as of *structure* and *constraints*. The effect of this category is obvious in the study by Bin et al. [18], where the same instrument played in two genres has a different impact on familiarity.

2.2.1.6 Decomposition of familiarity

Finally, the complete decomposition of familiarity that we propose includes the following dimensions:

$$
\begin{aligned}
Familiarity = \; & AA_{exclusivity} + AA_{priority} + AA_{consistency} \\
& + I_{composition} + I_{behavior} \\
& + E_{expertise} \\
& + M_{culture} \\
& + G_{structure} + G_{rules}
\end{aligned} \tag{2.1}
$$

2.2.1.7 Adaptation of the model to the experimental context

A simplified model To maintain a decent amount of questions in the survey developed to assess this model of familiarity, we opted for a simplified version and kept three main variables: Attributed Agency, Instrumentality and Expertise.

Decomposition of attributed agency Besides, we saw that *Attributed Agency* (see section 1.3.3) can theoretically be decomposed in *Exclusivity*, *Priority* and *Consistency*. Additionally to the evaluation of the model of familiarity, we take the opportunity of this study to assess this decomposition through its simplified version with *Exclusivity* and *Consistency* only.

Thus, the simplified equation is now:

$$
\begin{aligned}
Familiarity = \; & AttributedAgency(AA_{exclusivity} + AA_{consistency}) \\
& + Instrumentality \\
& + Expertise
\end{aligned} \tag{2.2}
$$

2.2.2 Unit Perceived Interactions

In this study of the familiarity spectators can have with a musical interaction, we want to gather specific data to determine the weight of each of the components of the model of familiarity we hypothesize (Equation 2.2). Because artistic expression can be extremely complex and rich, many aspects of a performance are indistinctively in play. A controlled approach can help to exclude uncontrolled variable as much as possible and so provide refined results.

To this end, the stimuli we developed for this study are basic performative actions which perceptual parameters are controlled. They take the form of videos where a single gesture triggers a change on identified parameters of the interaction. We call these videos **Unit Perceived Interactions** (UPIs).

2.2.2.1 Parameters

As indicated on table 2.1, a UPI combines eight perceptual parameters (and their modalities):

- Instrument: the instrument used in the interaction
- Point of view: the angle and the distance from which the interaction is observed (3 modalities)
- Gesture: the type of gesture executed by the performer (5 mods.)
- Gesture profile: the way the controller evolves from a start to an end position (2 mods.)
- Gesture amplitude: the amplitude of a gesture (2 mods.)
- Sound parameter: the sound parameter modulated in the interaction (3 mods)
- Mapping: the mapping configured in the instrument (4 mods.)
- Intro: an audio signal played before the interaction, while the musician is standing still at his instrument (5 mods.)

example A UPI can be for instance the video of a musician turning a knob on a midi controller that triggers a pitch modulation seen from a side point of view. Such UPI can be used as a stimulus alongside another one presenting the very same interaction but triggering a modulation of timbre rather than pitch to evaluate the impact of the sound parameter. One could also compare it with a UPI shot from a different angle to evaluate the impact of the point of view and so on.

2.2.2.2 A first dataset of UPIs

In the preparation of this study, we built a first dataset of UPIs with 4 instruments selected to represent a 'technological range' from the most archetypal to the most disruptive: an acoustic guitar, a digital piano, a MIDI controller with knobs and faders (see Fig. 1.8), a hand tracking device for midair gestures (Leap Motion). Two takes were shot for each UPI for potential variability needs. After a shooting session and automatic post production processing to cross

INSTRUMENT	
	The instrument used in the interaction
POINT OF VIEW	
FRONT	As seen from front, the instrument may be partially visible
FRONT BOTTOM	As seen from a lower front position , the instrument may be not visible
SIDE	As seen from a side eye view, the instrument is always visible
GESTURE	
TURNING	The gesture of turning a knob
SLIDING	The gesture of push/pull a slider
MIDAIR-CIRCLE	The hand draws a circle over the instrument
MIDAIR-TRANS	A vertical movement of the hand
USUAL	The gesture commonly associated with the instrument (only used with generic instruments : guitar, piano)
GESTURE PROFILE	
0-1-0	The gesture starts from a minimum goes to a maximum and come back to the origin
0-1	The gesture starts from a minimum goes and remains to a maximum (pending gesture) *not yet available in the dataset*
GESTURE AMPLITUDE	
SMALL	The gesture is deployed with a small amplitude
LARGE	The gesture is deployed with a large amplitude
SOUND PARAMETER	
PITCH	The pitch of the sound
TIMBRE	The timber of the sound
USUAL	The sound commonly associated with the instrument (only used with generic instruments : guitar, piano)
MAPPING	
LINEAR	The sound modulation follows the gesture
STAIRCASE	A continuous gesture triggers discreet stepped modulations
REVERSED	The sound modulation is reversed from the gesture
DIS2CONT	A discreet gesture (e.g. a single guitar picking) triggers a continuous pitch modulation (as in V13)
INTRO	
The intro is an audio signal played before the interaction, while the musician is standing still at his instrument, not interacting with it	
Modality	**Description**
NO	No introduction
INTRO_0	The exact same sound
INTRO_1	The same sound parameter through the same mapping but the nature (speed, rythm, …) of the modification is a bit different
INTRO_2	The same sound parameter through a different mapping
INTRO_3	A different sound parameter and a different mapping

Table 2.1: List of perceptual parameters and their modalities manipulated in the dataset of *Unit Perceived Interactions*

the different perceptual parameters, the dataset counts up to 2000 UPIs [3].

2.3 The study

In this section, we present the study that allows us to evaluate our hypothesised model of familiarity. We proceeded in three steps.

First, as detailed below, we carefully selected UPIs from our dataset to cover the largest range of effect on each dimension we want to assess. See section 2.3.1

Second, after the experiment, we conducted a first analysis of the collected data to confirm our previsions of effect range on each dimension. See section 2.3.2

Third, with the confirmation that our stimuli covered a large range of effect, we analysed the data to validate our hypothesized model of familiarity. See section 2.3.2

2.3.1 Methodology

2.3.1.1 Stimuli selection

To compose the stimuli of the study, we extracted UPIs from our dataset with the ambition to cover the largest range of effect on each dimension *Expertise*, *Consistency*, *Instrumentality* and *Exclusivity*. Thus, we chose 4 subsets of UPIs according to their expected score in *Expertise*, *Consistency*, *Instrumentality* and *Exclusivity*. Practically, by induction from their perceptual parameters, we chose the two UPIs for which we predicted minimum and maximum values for a given dimension as well as two or three other videos expected to sit between these extreme values.

Example of stimuli selection For instance, as the *Consistency* is the dimension that relates to the causality between an action and its consequences, we hypothesised this dimension score to be maximum with a video (UPI-02) of a string picking on a guitar (see table 2.2) and minimum with a video (UPI-10) showing a musical interaction that produces a discrete volume variation (OFF/ON/OFF) from a continuous midair translation gesture above a Leap Motion. All in all, as a same UPI could lay in multiple subsets, 13 stimuli were extracted (Fig. 2.2).

Balanced order The order of the stimuli was designed in order to shuffle the different perceptual parameters, to evenly distribute stimuli that targeted the same dimension and to avoid

UPI	INTRO	INSTRU	MAPPING	SND PARAM	GESTURE
UPI-01	NO	NANO	LINEAR	TIMBRE	TURNING
UPI-02	NO	GUITAR	LINEAR	NOTE TRIG	PICKING
UPI-03	INTRO_0	NANO	LINEAR	PITCH	TURNING
UPI-04	NO	LEAP	REVERSED	PITCH	MA-TRANS
UPI-05	INTRO_3	NANO	LINEAR	PITCH	TURNING
UPI-06	NO	PIANO	LINEAR	NOTE TRIG	TAPING
UPI-07	NO	GUITAR	LINEAR	TIMBRE	MA-CIRCLE
UPI-08	NO	PIANO	HIGH-SOFT	NOTE TRIG	TAPING
UPI-09	INTRO_2	NANO	LINEAR	PITCH	TURNING
UPI-10	NO	LEAP	CONT2DIS	PITCH	MA-TRANS
UPI-11	INTRO_1	NANO	LINEAR	PITCH	TURNING
UPI-12	NO	LEAP	LINEAR	TIMBRE	MA-CIRCLE
UPI-13	NO	GUITAR	DIS2CONT	PITCH	PICKING

Table 2.2: The 13 video stimuli (Unit Perceived Interactions) and their perceptual parameters

interference due to the succession of conditions too similar. However, even if the survey is relatively short (less than 20 min), we took the risk of a limited order effect since all participants viewed the videos in the same order.

2.3.1.2 Participants

32 participants (20 males, 12 females, 36.7 y.o ± 4.3) were recruited to answer a survey at home via Internet. They were previously invited to plan a 20 minutes window where they would not be disturbed and they had to shut down their phone during the experience. They were informed that it was a survey and not a quiz with good or bad answers so they had to answer freely with their personal opinion. The exclusion criteria were hearing and vision impairments. After control of the reported answers, no participants were excluded. A quick technical check at the beginning of the survey was made to ensure the good quality of the sound before the first part of the survey. The first page presented the study and its goal. Participants were informed about their rights and conditions to access and remove their personal data.

2.3.1.3 Expertise profiles

The first part of the survey consisted of subjective and objective questions related to the participants skills and knowledge in music practice, electronic/experimental music performances and technology[2]. Depending on their score, participants were associated to one or more ex-

[2]See appendix A for details.

pertise profiles: *Musician*, *Electro*, and *Techno*. The *Musician* expertise was associated with participants playing music on a regular basis or with solid musical education. The *Electro* expertise was associated with participants who listen to electronic music and attend digital or even experimental performances. The *Techno* expertise was associated with participants that proved knowledge about digital devices, such as Arduino boards or infrared camera, and recent technological innovations. Participants with no expertise were associated to the profile *No_Expertise*. A single participant could hold several areas of expertise, e.g. Techno_Musician. The participants of the current study were composed of 17 people with at least one expertise profile and 15 participants with no expertise.

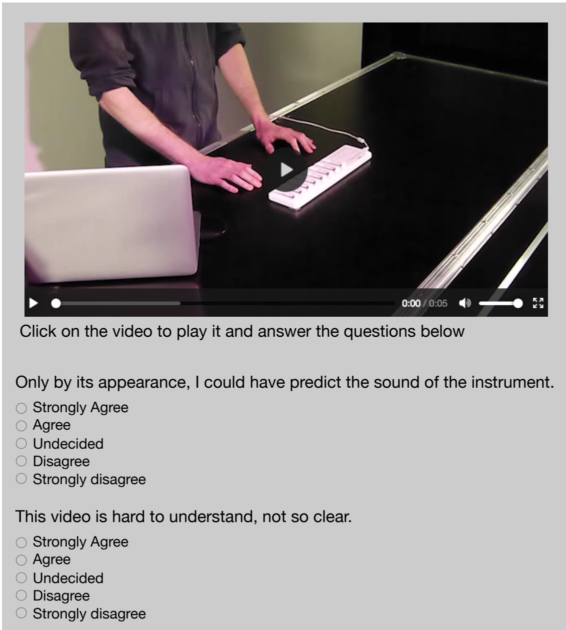


Figure 2.1: Survey interface (zoomed) with the video player and the first questions.

2.3.1.4 The survey

Apart from the general and profile questions described above, the survey is composed of 13 pages. Each page presents a video stimulus and a set of 6 questions related to the displayed interaction. The questions can appear in the form of multiple choices, Likert-style scale or yes/no/other radio buttons. The confidence is randomly asked and is computed as a multiplier of the answer corresponding score.

To access the next page, participants have to answer every questions. They can not navigate backwards. Participants are free to replay and pause the video at will. After the 7th page, participants are invited to take a short a break if necessary before resuming the survey for the last 6 questions. The last page thanks the participants for their time.

Details of the questions On each page, 2 questions individually targets *Familiarity* and *Attributed Agency* and 4 questions individually targets the components *Consistency*, *Exclusivity*, *Instrumentality* and *Expertise*. To reduce the redundancy of the questions and keep the participant's attention, the questions are extracted of a set of 3 to 4 equivalent questions, some with an negative form. In that case, a negative coefficient was applied to the answer.

Affirmations were followed by Likert-style scales of the degree of agreement:
Strongly agree / Agree / Undecided / Disagree / Strongly Disagree

Examples of questions by dimensions (The complete survey is available in appendix B.1) :

- Agency:
  - "The musician is totally in control."
  - "The musician does not seem to control his instrument."
- Familiarity:
  - "Globally, I find this interaction familiar, natural."
  - "This video is hard to understand, not so clear."
- Consistency:
  - "Is the sound consistent with the musician gesture ?"
  - "The sound fits what the musician wanted to do."
- Exclusivity:
  - "In this video, the sound is controlled by: the musician / the computer / both / other"
  - "It is certain that in this video, the musician is the only one in commands."
- Instrumentality:
  - "Considering its appearance only, do you think the sound matches the instrument ?"
  - "Only by its appearance, I could have predict the sound of the instrument."
- Expertise:
  - "Have you already seen musicians playing like this, that is with the same gestures but maybe with another instrument ?"
  - "Have you already played the current instrument (at least once) ?"

2.3.2 Predicted impact of stimuli on targeted dimensions

2.3.2.1 Research question

Before analysing the link between *Familiarity* and its hypothetical dimensions, in this part of the study we evaluate the predicted impact of the stimuli we selected on these dimensions (Table 2.2).

2.3.2.2 Results

No order effect By deciding to expose participants to questions in the same order, we took a relative risk of order effect. Data analysis through Spearman's rho excludes such order effect ($\rho s < -0.34,\ p < 0.004$).

Predictions partially confirmed As shown on figure 2.2, the subsets of stimuli did gradually impact their targeted dimensions even if the rank order was not always perfectly predicted. The post-hoc results confirmed the gradient effect of each set on its target dimension and generally followed our prevision for the minimum and maximum values. However it also revealed some discrepancies in the hypothesised ranking order especially for the *Exclusivity* dimension.

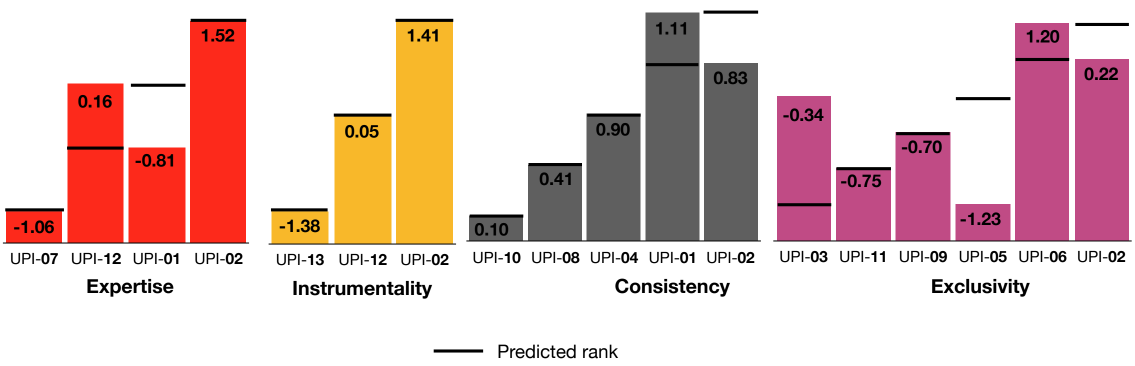


Figure 2.2: For each familiarity dimension, we composed a subset of stimuli that we predict should gradually impact a target dimension (e.g. UPI-07, UPI-12, UPI-01, UPI-02 for Expertise). This figure shows the predicted versus the measured rank of each stimuli. The actual measured score is indicated at the top each bar.

2.3.2.3 Discussion

Lack of literature To our knowledge, this is the first study to evaluate the effect of perceptual parameters of musical interactions on components such as *Exclusivity* or *Instrumentality*. This lack of literature constrained us to select our stimuli by arbitrarily inferring their impact on familiarity and the other dependent variables. We debated with the team to validate the selection of UPIs, still it remains an arbitrary inference from perceptual parameters.

Predictions globally confirmed As illustrated on Figure 2.2, post-hoc results globally confirmed the expected gradient effect for each component. The less effective rank hypothesis concerns the *Exclusivity* dimension (Fig. 2.2). Initially, we decided to manipulate the *intro* parameter to support an impact on *Exclusivity* dimension. As described in Table 2.1, the *intro* parameter indicates the presence of an audio signal before the interaction, while the musician is standing still and not interacting with the instrument. We anticipated that the presence of an introduction would be a negative factor for the exclusivity judgement depending on the nature of the introduction. The more the introduction sounds like the audio signal triggered by the interaction, the more likely the sound triggered by the musician could be associated to an autonomous source, and therefore downgrade the exclusivity judgement. The measures did not follow our ranking predictions but we kept the stimuli subset in the results as a solid gradient effect, even if not in the right order, was confirmed.

No bias in the data To sum up from this first analysis, the stimuli used in the study ensure a gradient effect on each of the components we target. We can now proceed to the model analysis with the insurance that no particular bias can downgrade the consequent interpretations.

2.3.3 A first model of familiarity and its limitations

2.3.3.1 Research questions

Now that we confirmed the predicted impact of our stimuli on *Consistency*, *Exclusivity*, *Instrumentality* and *Expertise*, in this part of the study, we evaluate the model of *Familiarity* based on *Attributed Agency*, *Instrumentality* and *Expertise*.

2.3.3.2 Results

Correlations To get a a first overview of the results, we computed individual correlations that confirmed the strong link between *Familiarity* and the proposed dimension as shown on figure 2.3.

Correlations	Agency	Cons	Exclu	Instru	Expert
Agency	1	.576**	.286**	.349**	-0.021
Consistency	0.576**	1	.195**	.289**	0.021
Exclusivity	0.286**	0.195**	1	.240**	0.038
Instrumentality	0.349**	0.289**	.240**	1	.295**
Expertise	-0.021	0.021	0.038	0.295**	1
Familiarity	0.613**	0.546**	0.234**	0.388**	0.092

Table 2.3: Correlations table: Familiarity and its hypothesised components show multiple correlations.

Expertise Surprisingly, *Expertise* did not correlate directly with *Attributed Agency* nor with *Familiarity* ($p > 0.325$). However, a positive correlation with the *Instrumentality* ($r_{414} = 0.295, p < 0.001$) may give the *Expertise* an indirect link on the level of control perceived (*Attributed Agency*) or the *Familiarity*. The *Expertise* dimension is discussed in more details in its dedicated section.

2.3.3.3 Linear model of Familiarity

Individual correlations indicate a link between our variables of interest. We will now build the model of familiarity from the hypothetical dimensions and evaluate it.

Methodology Before processing the data, outliers where trimmed following the Cook distance. This preprocessing is part of the Automatic Data Preparation from SPSS (V25). All the following comparisons of models are based on the Akaike information criterion (AICc) and the accuracy.

A relevant model and questions on Expertise To analyse the contribution of each dimension, we computed a linear model of *Familiarity* with *Attributed Agency*, *Instrumentality* and *Expertise* as factors. Data show a strong contribution of *Attributed Agency* ($\beta = 0.883, p < 0.001$) and a moderate contribution of *Instrumentality* ($\beta = 0.117, p < 0.001$). *Expertise* does not reached the significance level. This model has an accuracy of 40.8% and a Akaike information criterion (AICc) of 53.725 (Table 2.4).

2.3.3.4 Familiarity $\neq$ Agency and decrease of Instrumentality

The primary analysis of the results partially confirmed the proposed dimensions but also highlighted an almost mirror relationship between *Attributed Agency* and *Familiarity*. However,

Familiarity = Attr. Agency + Instrumentality + Expertise			
Model term	**Coef.**	**Sig.**	**Importance**
Intercept	-0.002	0.974	
Attributed Agency	0.555	0.000	0.883
Instrumentality	0.205	0.000	0.117

AICC : 53.725 **Accuracy : 40.8%**

Table 2.4: Linear model of familiarity based on *Attributed Agency*, *Instrumentality* and *Expertise*. *Expertise* contribution did not reached the significance level and was therefore excluded from the model.

the interpretation as a strict equivalence between the two notions can be excluded from the clear difference in the focus of their related questions in the survey.

From there, the exploration of cases where *Familiarity* and *Attributed Agency* are not correlated can be useful.

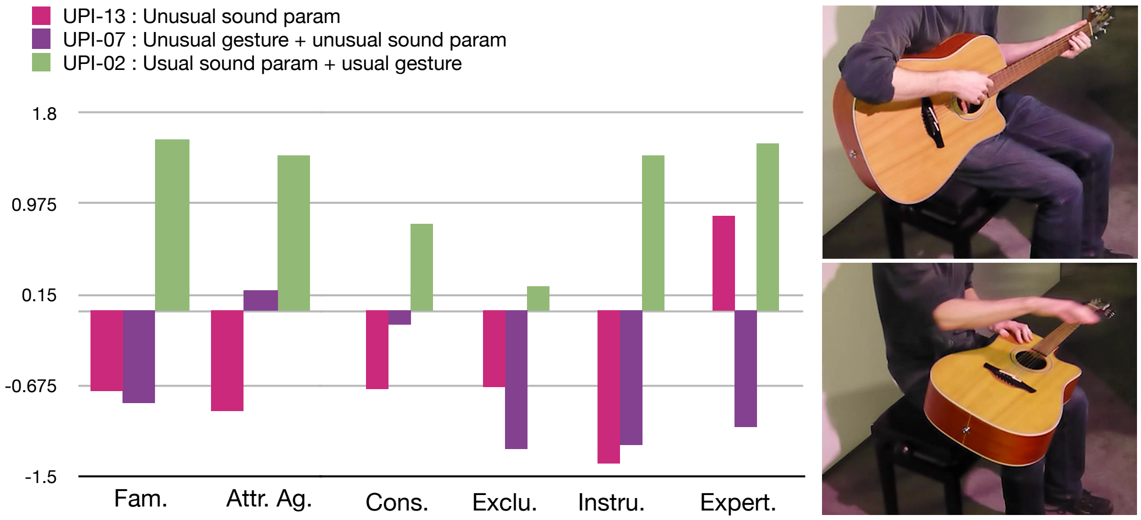


Figure 2.3: *Left* - Often highly correlated, *Familiarity* and *Attributed Agency* differ in some situations associated with a very low *Instrumentality* score and contrast in *Expertise* and *Attributed Agency*. *Top Right* - Usual picking gesture with an acoustic guitar. *Bottom Right* - MidAir gesture with no visible tracking device.

Cases with no correlation between Attributed Agency and Familiarity The strong correlation between *Attributed Agency* and *Familiarity* disappears with some UPIs. In that case, observers judges the interaction with a very low *Instrumentality* score.

In video UPI-07, the guitar triggers a digital sound (timbre modulation) from a contactless gesture, a MidAir Circle, without any tracking device visible on screen. Compared to a UPI showing a classical picking gesture triggering a usual acoustic guitar sound (UPI-02), a radical decrease in instrumentality is logically observed ($t_{62} = 13.644, C = 1.99, p < 0.001$). The same low *Instrumentality* score is also observed in UPI-13 where the musician triggers the pitch modulation of a synthetic sound with the same usual picking gesture than in UPI-02 still without any device or cable visible on screen.

Participants reconsider their Expertise from Instrumentality evaluation In these 2 examples, *Familiarity* is identical and the interaction did not impact the other dimensions the same way they impacted *Instrumentality*. As shown on figure 2.3, *Familiarity* does not change between UPI-13 and UPI-07 but an important difference in *Attributed Agency*, carried by *Consistency*, is observed.

Second remarkable divergence, *Expertise*, that is the expertise a participant reports having with an instrument, is positive for UPI-13 (guitar with unusual sound output) and negative for UPI-7 (tilted guitar with unusual gesture and sound) even though they both display an acoustic guitar. In the context of UPI-07, **observers seem to consider the instrument as a new one and therefore, reconsider their expertise and moderate their ability to judge (Attributed Agency $= 0.188 \pm 0.5$) the level of control of the musician.**

An interpretation of Fig 2.3 is that spectators judge unfamiliar with low *Exclusivity* and *Consistency* an interaction with a instrument they know but that does not have the usual sound outcome (UPI-13). When the guitar is tilted, and both outcomes and gestures are unusual, they also judge the interaction as unfamiliar but the acoustic instrument is considered by participants as an instrument they have no expertise about (UPI-7).

Here we have the demonstration of how an acoustic instrument (guitar) becomes a DMI for spectators. DMIs are not considered as instruments defined by their physicality or shape, but by they entire interaction possibilities. From this distinctive appreciation, they logically evaluate their own expertise on multiple criteria, physical aspect, mappings, and sound output, rather than on the physical aspect alone as they might for acoustic instruments

2.3.4 Agency dimensions

In the survey, we ask question about the global *Attributed Agency* and specific questions that target its dimensions *Consistency* and *Exclusivity* [208, 15] (See 1.3.3 for details). From these data, we confirm the link between *Attributed Agency* with a linear model. *Consistency* is by far the main component and supports 85 % of the explained variance.

Attributed Agency = Consistency + Exclusivity			
Model term	**Coef.**	**Sig.**	**Importance**
Intercept	0.130	0.001	
Consistency	0.676	0.000	0.851
Exclusivity	0.266	0.000	0.149
AICC : -1049	**Accuracy : 30.6 %**		

Figure 2.4: The survey results confirm the role of *Consistency* and *Exclusivity* in the emergence of *Attributed Agency*.

2.3.5 Limitations

Based on a web survey, this study avoid some biases due to the usual lab conditions: experimenter in the room, unusual environment or schedule constraint. However, the absence of an experimenter to confirm that the questions were perfectly understood could have led to a lower variability on some questions. In this regard, we excluded from the results a question that appeared after the experiment to be too ambiguous.

Furthermore, all the questions appeared on the same page, below the video player. This proximity may have encouraged the participants to answer the same way to questions that seemed similar. However, as our results showed correlations and discrepancies, a strong effect of this bias can be excluded.

Finally, as for any model, our model of familiarity requires a large amount of data to improve its capability to produce useful indications on random interactions. The model presented in this chapter was built from only 13 UPIs and could be refined by applying the present methodology to the other interactions registered in our dataset.

2.4 Conclusion

In this chapter, we proposed and evaluated a decomposition of the familiarity. Our data show a tight relation between *Familiarity* and *Attributed Agency*, confirming the importance of *Attributed Agency* in the spectator experience. Our survey did not clearly exposed the role of *Expertise*. In further studies, we will address *Expertise* with refined profiles evaluation. Finally, thanks to an approach by components, we revealed contrasts in the evaluation of *Attributed Agency* by observers for a same *Familiarity* evaluation. In particular, we propose a detailed hypothesis about how observers change their perception from an acoustic instrument into a DMI.

Music has always pushed the envelope of what defines interaction.

Atau Tanaka

3
A taxonomy of Spectator Experience Augmentation Techniques (SEATs)

Main contributions

- Taxonomy of eleven dimensions to organise the Spectator Experience Augmentation Techniques contributions
- Review and normative description of Spectator Experience Augmentation Techniques

> The contribution of this chapter was published in the proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2020) [39] and received the Best Poster Award.

3.1 Introduction

As we saw in section 1.3.5 and as illustrated on Figure 3.1, artists and researchers proposed techniques to improve the experience of spectators of performances with Digital Musical Instruments (DMIs). Because of the perceptive issues induced by musical interactions with DMIs (see section 1.3.2.2 for extensive listing), these techniques facilitate the accessibility of performances with existing and future interactive systems.

We propose to name these techniques Spectator Experience Augmentation Techniques (SEATs).They range from textual information [11] and pre-performance demos [16], to augmented reality displays [14, 160] and haptic feedback devices [1, 195].

To our knowledge a formal analysis of SEATs has not yet been conducted, restraining the development of such techniques. Most classifications related to DMIs are defined from the musician or instrument designer point of view, such as the dimension space proposed by Birnbaum et al. [20]. Previous work on audience experience has led to classifications such as challenges for the design of DMIs from an audience perspective [93] and to a 2D space for the analysis of action and effect perception [163].

In this chapter we present a taxonomy for Spectator Experience Augmentation Techniques (SEATs). This taxonomy is aimed at artists and researchers as it can help inform the design of performances to compensate for the alteration of audience experience in performances with digital interfaces.

3.2 Spectator Experience Augmentation Techniques

It is essential to point out that, by using the term *augmentations*, we focus on techniques which have no impact on the design of the performer's interface, i.e. they do *not change the way performers interact*. Rather, they provide additional information to the audience on various components of the performance, such as performer's equipment, gestures or intentions.

SEATs may be designed with various goals in mind. They may be: pedagogical, with a focus on helping spectators understand performers' interactions; aesthetic or engaging, with a focus on increasing the implication of the audience; or even disruptive, so that performances appear as magical or secretive [164]. Therefore, the values chosen for each dimension of our taxonomy may have an impact on multiple aspects of the audience experience.

We do not include interfaces or techniques for audience participation in our taxonomy as actively involving spectators with a direct control on the artistic content developed "on stage" implies a specific approach with numerous extra considerations and dimensions.

Finally, for genericity reasons, our taxonomy does not take the display technology into account. For instance, visual augmentations with the same dimensions could be achieved through projection mapping, with video AR on a smartphone or with an optical see-through head-mounted display. Our aim is that this taxonomy will remain usable by designers of performances with future technologies.

3.2.1 Taxonomy dimensions

Our taxonomy is composed of eleven dimensions, described below. For each of them, we provide the range of possible values and discuss their potential impact.

3.2.1.1 Spatial Alignment

This dimension describes how augmentations are spatially aligned with the performance components. Possible values include: *aligned*, when the augmentations are perceived as coming from the component itself; *co-located* when they can be perceived while maintaining the focus on the performance; *distant* when there is a shift in focus required to access them.

For instance, visual augmentations of a music performance could be aligned if using a Pepper's ghost augmented-reality display, co-located if projected on a screen behind the musician, or distant if displayed on the spectators' mobile devices. While aligning augmentations seem ideal because it preserves the focus on the performance, the presentation of complex information might cause sensory or cognitive overload for the audience, in which case a more distant presentation is preferable.

3.2.1.2 Temporal Alignment

This dimension pertains to when spectators can access to the augmentations: *Before*, *During* or *After* the performance.

For instance, in the case of verbal explanations, they may be provided as a pre-performance demo, as comments during the performance, or as a post-performance discussion. If presented before the performance, augmentations may provide cues for the audience without any alteration to the performance. However if they are too complex or given too long before, some of the information may not be remembered when needed. If presented during, the information is given when needed but it might distract or overwhelm the audience. If presented after, the information might be lacking to understand events when they happen, but the magical aspect [164] of the performance is preserved.

3.2.1.3 Temporal Density

This dimension relates to the temporal range that the augmentations cover. It can be broadly defined as a *low*, *medium* and *high*.

For example, augmentations of a music performance could range from displaying only the last played note to displaying the whole score. While a high density will help the audience forge a stronger sense of performer's intentions or actions over time, and may help them in perceiving virtuosity or errors, providing too much information might also distract them from the current actions.

3.2.1.4 Temporal Control

This dimension describes the possibility for the audience to control the temporal range covered by the augmentations, i.e. on which part of the performance they will access information. It ranges from *none* to *full*.

An example would be textual explanations provided one by one either synchronously with events of the performance or on request, with spectators being able to freely scroll through them. From a design perspective, allowing for temporal control may help spectators build a better understanding of the performance, but it might also lead them to miss augmentations that should be perceived at a specific time.

3.2.1.5 Semantic Density

This dimension relates to how much information is provided by a SEAT. It can also be seen as the information level-of-detail. Like temporal density, it can be broadly classified *low*, *medium* and *high*. For instance, augmentations on a digital musical instrument may range from only showing which synthesizer is being played, to showing the detailed audio graph including the activity of all synthesis parameters. We will see in more detail in chapter 4 that levels of detail do have an impact on the spectator experience.

This dimension is essential as one needs to ensure that the density is sufficient to provide useful information but not too high so that it does not result in sensory or cognitive overload

	Spatial Alignment	Temporal Alignment	Temporal Density	Temporal Control	Semantic Density	Semantic Control	Pres. Nature	Pres. Modality	Content Nature	Content Reactivity	Agents
Perrotin [160]	co-located	during	medium	none	medium	none	abstract / conceptual	visual	technical / gestural /	reactive	avatar / origin
Benford [11]	co-located / distant	before / during / after	high	full	medium	limited	abstract / linguistic	visual	technical / intentional	semi-fixed	—
Armitage [1]	distant	during	low	none	low	none	figurative	haptic	gestural	reactive	—
Turchet [195]	distant	during	low	none	low	none	figurative	haptic	gestural	reactive	—
Bin [16]	aligned	before	medium	none	low	none	linguistic	auditory / visual	technical / gestural	fixed	—
Berthaut [14]	aligned	during	low	none	medium	none	abstract	visual	technical / gestural / causal	reactive	—

Table 3.1: Analysis of Spectator Experience Augmentation Techniques along the dimensions of our taxonomy

for the audience.

3.2.1.6 Semantic Control

Symmetrically to temporal control, this dimension indicates if the level of semantic detail can be chosen by the audience, allowing for the personalized journeys suggested by Benford et al. [11]. It ranges from *none* to *full*.

In the case of textual explanations, it can for example go from a single level of detail given to all spectators to multiple versions targeted at different levels of expertise, e.g. children / adults or novices / experts.

This level may also be automatically selected according to emotional or cognitive states measured through wearable physiological sensors, as we proposed it with our conceptual pipeline (See section 5). This dimension has an impact on the accessibility of the augmentations. Increasing the control helps avoiding too simple or too complex information depending on spectator's expertise, but it also increases the complexity of implementation for the performance designers. One also needs to ensure that the control interface is not too difficult or that it does not distract spectators from the performance.

3.2.1.7 Presentation nature

This dimension relates to the form given to the information provided to the audience. Possible values include *figurative*, *abstract*, *conceptual*, and *linguistic*.

For instance, to indicate that a certain key was played by a musician the augmentation can display respectively a close-up of their hands and the keyboard, a colour changing shape, the note on a scale or the note name. This dimension involves a trade-off between the information explicitness, maximised with linguistic or figurative augmentations, and its compactness which can be optimised with abstract or conceptual augmentations showing only essential information. It also has implications on the aesthetic integration of the augmentations in the performance.

3.2.1.8 Presentation modality

This dimension describes how the augmentations are displayed. The modality can be *visual*, *auditory* and *haptic*.

For example, the subtle gestures of a performer on a sensor can be amplified using changes in a visual shape that represents the sensor or through vibrotactile feedback reproducing the gestures. The choice of modality depends on the type of performance and its scalability. While visual and auditory displays can be generalised, haptic ones imply more restrictions as individual devices need to be designed.

3.2.1.9 Content nature

This dimension pertains to the nature of the content displayed by the augmentations, i.e. on which aspects of the performance the SEAT provides information. We identified four possible aspects: *technical* to reveal the mechanisms of the interface [14]; *gestural*, to amplify subtle/hidden movements [160]; *intentional*, so that the audience understand what performers are trying to accomplish [77]; *causal*, so that spectators have a clear perception of who from the performer or autonomous processes is responsible for variations in the sound.

The choice of content nature depends strongly on the aim of the augmentations. Showing the intention can highlight the performer's virtuosity or affect the audience emotional response, while technical augmentations may increase the level of comprehension of the audience.

3.2.1.10 Agents

This dimension describes the representation a SEAT can give of the agents in a performance and of their interactions. Agents are entities that can interact with instruments and interact together, like 3 musicians on stage or a composite crew including virtual agents. We envision 4 values for this dimension. *Origin*, so that spectators perceive who is the source of events in the performance, for example amongst members of an orchestra [160] or between the performer and automated processes. *Avatar* indicates if the SEAT uses representations of the agents to provide additional information. This modality could be used in performances where the musicians are not physically located in the same place or when an agent is virtual. *Communication* when musicians exchange information not directly linked to sound production, such as synchronisation or support signals. *Interactions* indicates the capacity of a SEAT to augment the interactions between agents. This can be useful for instruments with shared controls and rich musician-to-musician interfaces such as bf-pd [56].

3.2.1.11 Content reactivity

This dimension describes the relation between the augmentations and the performance. It can be *fixed* when the augmentations are pre-defined and do not change with the performance, *semi-fixed* when some elements of the augmentations change according to the performance, and *reactive* when augmentations are generated from information extracted in realtime during the performance.

For instance, visual augmentations can be pre-recorded videos provided at pre-defined moments or synthetic graphics triggered and adapted dynamically based on the performers actions. While a reactive content guarantees that the augmentations will adapt to changes in the performances, it involves a technical complexity not always achievable. For example performers intentions can not be extracted dynamically during the performance. It may also lead to information being provided in a less accessible way than fixed content if it is not designed carefully, e.g. with multiple visual indications overlapping.

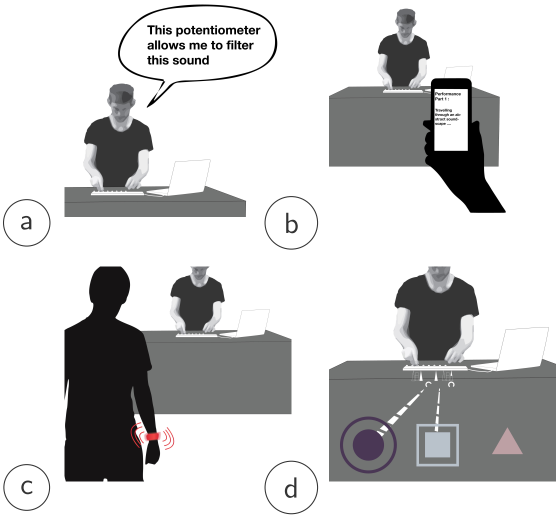


Figure 3.1: Spectator Experience Augmentation Techniques: a) pre-performance explanation, b) textual information on mobile, c) haptic wearables, d) visual augmentations

3.3 Analysis of existing SEATs

Table 3.1 provides a decomposition of the following SEATs from the literature.

Perrotin et al. [160] use visual augmentations projected behind a digital orchestra to display the musical parameters (e.g. notes) played by each musician.

Benford et al. [11] use visual augmentations, both conceptual and textual, projected on stage and available on spectators' mobile devices to provide an overview of the structure and intentions in a musical performance.

Armitage [1] uses haptic augmentations to provide the audience with vibrotactile feedback that amplifies the performer's key presses in a live coding performance.

Berthaut et al. [14] use visual augmentations to reveal the mechanisms of a DMI (including sensor values, sound processes activity and sensor to processes mappings) to the audience with an augmented reality display.

Bin et al. [16] uses pre-concert explanations of the instrument to increase the spectators' level of enjoyment and comprehension.

Turchet et al. [195] use a wearable haptic device to provide vibrotactile feedback to the audience to amplify the musician's gestures.

3.4 Using the taxonomy for design

When creating SEATs for a particular performance, our taxonomy can be used to explore different designs with various effects on the audience experience. Starting from a very common visual display of information behind the performer, the following directions could for example be chosen, depending on the desired effect:

- In order to improve the perception of the performer's control, i.e. the audience's Attributed Agency [15], value of the *Spatial Alignment* dimension can be increased, for example using augmented reality displays which would place the information directly on top of the performer's gesture, therefore improving the causal link between actions and effects. However, attention must then be paid to the quantity of displayed information, so as not to overload visually the performance and reduce the visibility of the performer's gestures.
- In order to increase the audience engagement in the performance, the *Presentation Modality* dimension could combine visual and haptic feedback, for example through the use of belts or wristbands providing vibrotactile feedback. This direction however poses issues for

implementation, as it requires to equip each spectator with a individual device or constrains them to fixed positions (e.g. seats) with built-in feedback.

- Finally, one could choose to adapt the augmentations to spectators with different levels of expertise. In that case, the value of the *Semantic Control* dimension could be increased. This would however require an individual or grouped device for the selection of the level of information provided by the augmentations, for example using a voting system. In addition, this dimension requires choosing which levels of information will be made available and how they will each be displayed, potentially constraining the aesthetic choices for the performances.

3.5 Conclusion

In this chapter, we proposed a taxonomy for Spectator Experience Augmentation Techniques (SEATs), which complement digital performance interfaces in order to improve the audience experience. We showed how our taxonomy can be used for the analysis of existing techniques and the design of novel ones. As future work, we believe a formal evaluation should be conducted, through both ecological and controlled experiments, which would precisely determine the impact of each of these dimensions on the audience experience. Artists and researchers would then benefit from a better understanding of how to design the spectator experience while preserving the expressiveness and appropriation possibilities of digital performance interfaces.

The nature of reality
Is pure subjective fantasy
Space and time and here and now
Are only in your mind

Andy Bell

4

A methodology for the study of SEATs

Main contributions

- Controlled protocol for the study of Spectator Experience Augmentation Techniques
- Decomposition of spectators experience in objective and subjective components
- Bayesian analyses of spectator experience
- Improvement of the trust in the musician's contribution
- Proposition of a new design challenge for spectator experience: the Association

> The contribution of this chapter related to the protocol of evaluation of spectator experience augmentation techniques was published in the proceedings of the ACM conference on Designing Interactive Systems (DIS 2020) [37].

> The contribution of this chapter related to the study of levels of detail and expertise was published in the proceedings of the International Conference on New Interfaces for Musical Expression (NIME 2020) [38] and has been selected for an extended version in the MIT Computer Music Journal.

4.1 Introduction

In the previous chapter, we presented a taxonomy to structure the Spectator Experience Augmentation Techniques (SEATs) proposed by the HCI and NIME communities. The large amount of contributions illustrates the important effort of artists and researchers in the improvement of spectator experience in electronic music contexts. As we saw in 1.3, this effort is a response to the multiple challenges induced by performances with Digital Musical Instruments (DMIs) when considering the perspective of spectators. The proposed techniques range from pre-performance demos [16] to augmented-reality displays [14] and provide various types of information about the instrument, its mechanisms and the ongoing interactions of the musician.

While they constitute a promising solution to the issue of audience experience with digital performances, the efficiency of SEATs have not yet been formally compared. Therefore, their respective effect on the many aspects of the spectators experience remains unclear.

Among them, doubts about the performer's involvement is one of the identified risks for the audience's interest and experience [181, 190]. Indeed, when the collaboration between humans and machines happens in public, the audience can face difficulties in distinguishing the actual human contribution from the contribution of autonomous processes. As described in section 1.3.3, this issue is a direct consequence of a disturbed Attributed Agency that SEATs should compensate.

Moreover, existing SEATs do not take sufficient account of important individual factors like the expertise of spectators with DMIs or their personal sensibility to the amount or type of information conveyed by the augmentations.

In this chapter, we first introduce a controlled approach to study the efficiency of SEATs on multiple aspects of spectator experience, including Attributed Agency. We apply this protocol to compare two common SEATs, pre-concert explanations and visual augmentations. In a second section, we complete our methodological contribution with a specific evaluation of the impact of the levels of detail (LODs) exposed by visual augmentations and the role of expertise in the spectator experience.

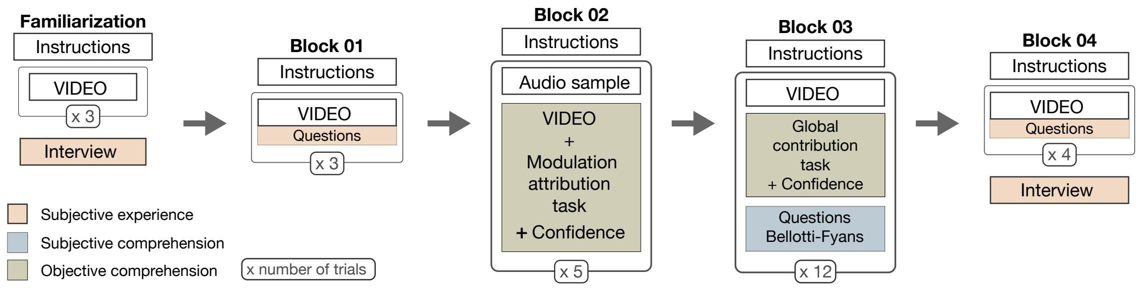


Figure 4.1: After a familiarisation phase, participants followed 4 successive blocks targeting the components of the experience: subjective experience in Blocks 01 and 04, objective comprehension in Blocks 02 and 03 and subjective comprehension in Block 03.

4.2 Controlled protocol for the evaluation of Spectator Experience Augmentation Techniques

In this section, we evaluate the impact of two Spectator Experience Augmentation Techniques (SEATs), pre-performance demo and visual augmentations. We investigate their effect on multiple aspects of the audience experience, including subjective and objective comprehension, especially looking at the perception of the performer's contribution (Attributed Agency). From the results, we derive insights which can inform both the design of future techniques and the evaluation of the audience experience with digital performances.

4.2.1 Study: comparison of two SEATs

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

4.2.1.1 Overview

For the participants, the experiment mostly consists in watching videos and answering subsequent questions. The experiment lasts one hour including consents signature, equipment setup and interviews. As shown on Figure 4.1, the experiment is composed of a familiarisation phase and four experimental blocks. Throughout the blocks, objective and subjective aspects of the spectator experience are addressed.

4.2.1.2 Hypotheses

Following the work of Berthaut [14, 15] on visual augmentations, and the conclusions of Bin [16] that even if explanations can familiarise the audience with a new digital instrument, they may not suffice to increase their experience, we hypothesised a greater impact of the visual augmentations on the spectator subjective experience and their objective comprehension. We also hypothesised that this impact is due to the provided information, not only to the increased visual richness of graphical material inclusion.

4.2.1.2.1 Taxonomy comparison Our hypothesis is also supported by the preliminary comparison of the SEATs through their dimensions in the taxonomy we present in chapter 3. As indicated on Table 4.1, the main differences lie in two dimension:
- *temporal alignment*: the explanations about the instrument behaviour occur *before* the performance while the visual augmentations happen *during* the performance.
- *content reactivity*: the content of pre-concert explanation is considered as *fixed* in the sense that it does not change with the performance. This modality for *content reactivity* is a direct consequence of the *before* value for *temporal alignment*. Visual augmentation in the contrary propose a *reactive* content reactivity throughout the performance.

	Spatial Alignment	Temporal Alignment	Temporal Density	Temporal Control	Semantic Density	Semantic Control	Pres. Nature	Pres. Modality	Content Nature	Content Reactivity	Agents
Explain	aligned	before	medium	none	low	none	linguistic	auditory / visual	technical / gestural	fixed	—
V-AUG	aligned	during	low	none	medium	none	abstract	visual	technical / gestural / causal	reactive	—-

Table 4.1: Preliminary analysis of the two SEATs addressed in this study, along their dimensions of our taxonomy. *Explain* : Pre-concert explanations - *V-AUG* : Visual augmentations. (The complete taxonomy is presented in chapter 3.)

From this preliminary analysis, the hypothesis of a greater efficiency of the visual augmentations technique is supported. Indeed, reactive content during the performance clear the memorisation issues of pre-concert explanations and so are more likely to improve the experience of spectators.

4.2.1.3 Decomposing the experience

To handle the elusive concept of experience, we split it into three measurable features, the *objective comprehension*, the *subjective experience* and *the subjective comprehension*. Each aspect is evaluated through one or more tasks and distributed across blocks of our protocol. This methodology maintains the participants attention with short and diversified tasks, and by addressing the experience under multiple aspects prevents from too narrow results [29].

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

- A guided rating task where the participants are asked to independently rate constituents of the performanc: the music, the virtuosity of the musician, the visual aspects and their overall appreciation of the performance. This task is tested twice and constitutes Block 01 and Block 04 (Figure 4.1). We placed the task at the beginning and at the end to let the time to the participants to integrate the SEATs' behaviour and control for a potentially too flat learning curve.
- Semi-structured interviews allow for more personal and detailed feedback and are crucial to fully understand what the participant experienced. For this purpose, with the consent of the participant, the experiment is fully audio taped. The interviews occurred after the familiarisation phase and at the end of the experiment (Figure 4.1).

4.2.1.3.2 Objective comprehension

The objective comprehension refers to the ability of the spectator to integrate factual elements of the interaction, like who from the musician or the machine is the author of a sound modulation, addressed in Block 02, or who is more contributing in a performance, addressed in Block 03 (Figure 4.1).

4.2.1.3.3 Subjective comprehension

Midway between subjective experience and objective comprehension, subjective comprehension relates to the inner feeling of being able to perceive an aspect of the interaction. This aspect of the experience is tested in Block 03 (Figure 4.1).

4.2.1.4 Experimental conditions

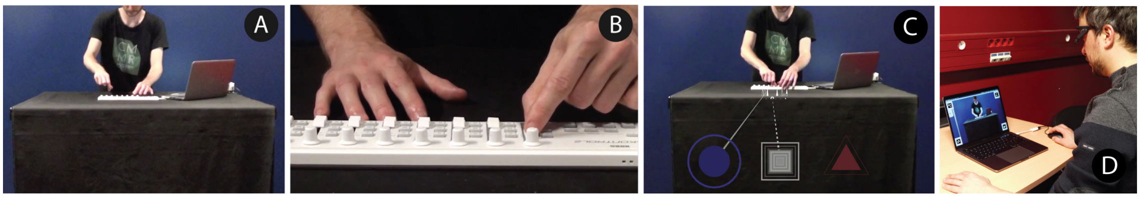


Figure 4.2: Conditions and *Spectator Experience Augmentation Techniques* evaluated in our stud: A) *Control*, B) *Explain* (pre-performance explanations), C) *V-AUG* (visual augmentations), D) The experimental setup

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

4.2.1.4.1 Control condition

In the control condition, the performance is showed as is, without any cues or explanation. (See Figure 4.2-A)

4.2.1.4.2 Preliminary explanations (Explain)

In the *Explain* condition, a short video demo (65 seconds) is played before the performance. This video is a close-up on the musician hands and the instrument that will be used in the subsequent performance. The musician demonstrates each mapping, from sensors (button, slider, knob) to sound parameters, with verbal explanations (See Figure 4.2-B). The participant can hear each sound processes played individually as well as their modulations allowed by the instrument. The performance is played right after the demo.

4.2.1.4.3 Visual augmentations (V-AUG)

In the *V-AUG* condition, visual augmentations are overlaid on the performance (See Figure 4.2-C). Visual augmentations are graphical representations of the behaviour of the instrument updated in real time. Similar to Berthaut et al. [14] they represent:

- The sensors and their activity in order to make them more visible (Figure4.3, I)
- The mappings between sensors and sound processes, which appear when the musician manipulates them and disappear after. (Figure4.3, II)
- The three sound processes with three shapes whose appearance changes according to changes in their sound parameters (from either the musician or the computer) : colour hue represents pitch, size represents volume, rotation represents scrolling in the sound, contours represent a delay effect. (Figure4.3, III)

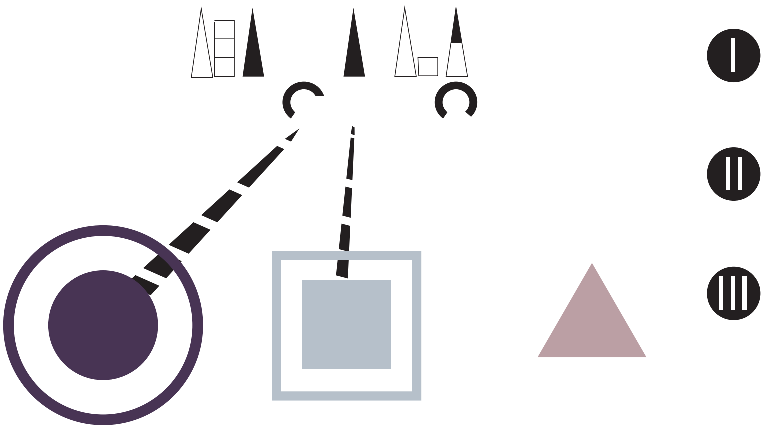


Figure 4.3: Visual augmentations are graphical representations updated in real time of the behaviour of the instrument. I) Sensors II) Mapping III) Processes. See Section *Visual augmentations (V-AUG)* for details.

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

[1]All stimuli, illustration videos of the conditions, anonymised raw results and statistical analyses can be found her:
`http://o0c.eu/0DA` (under the "Have a SEAT on stage" folder.)

4.2.1.5 Controlled stimuli

The stimuli used in this study are videos of short electronic music performances (about 30s) shot from an identical point of view[1]. Each video presents a musician behind a table on which a digital musical instrument (DMI) is placed next to a computer (see Figure 4.2-A). The video starts by a fade in with music already playing and the musician manipulating the DMI. The video ends with a fade out.

The control device used by the musician is a MIDI control surface (Korg NanoKontrol) composed of sensors: linear potentiometers (faders), rotating potentiometers (knobs) and buttons (see Figure 4.2-B). Linked to the device by USB, the computer is running a music software (Pure Data) allowing for the control of three sound processes (and their associated parameters) : a bass track (pitch, note_length, filter, volume), a beat/percussion track (trigger_note, echo, filter, volume) and a textural track (position_in_sample, filter, volume, echo). This setup (controller + computer) constitutes the common architecture of many DMIs.

4.2.1.5.1 The mapping The mapping is the relation between the sensors of the control device and the sound parameters. It impacts multiple aspects of the performance [100]. In particular, the mapping influences the contribution of the musician by setting the ratio between the controls given to the musician and those given to autonomous processes. Thus the role of the mapping is crucial in the definition of an instrument [100]. In order to vary the contribution level of the musician, from very high to very low, we designed three instruments from the MIDI controller:

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

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

4.2.1.5.2 Performance masking - Learning bias

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

4.2.1.6 Participants

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

4.2.1.7 Procedure

After having read and signed the consent form, the participant was equipped with two watchbands measuring cardiac and electrodermal activity, and with a light eye-tracking device. The physiological aspects of the experiment are not mentioned in the present document and will be covered in a future study. During the whole experiment, the participant was seating at a desk in front of a laptop equipped with a mouse (see Figure 4.2 - D). She or he was informed that the experiment consisted in watching videos and answering the subsequent questions via a survey form. After each block, the participant was offered to take a short break.

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

4.2.1.8 Data analysis

Data was recorded, anonymised and stored in real time during the experiment by a bespoke experiment software developed in Python. For technical reasons, one participant did not finish the second block and was excluded of the analysis for this block. Subjective reports were obtained via Likert scales and were analysed with parametric tools when the normality assumptions were met [191, 40]. All the analyses were conducted under the common frequentist paradigm and were combined to Bayesian statistics.

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

Bayesian statistics offer sounder analysis when dealing with relatively small samples like ours (n=21) [120, 102]. Besides, by clearly exposing the evidence in the data that support either the null or the alternative hypothesis, they also provide a more intuitive interpretation [126] and are increasingly considered as the researcher-centred design of statistics [120]. In this work, whenever possible, the ANOVA, the pairwise test [172, 151, 174, 173] and the linear regression [46, 48, 133] are expanded with their Bayesian version and a Bayes factor is reported as BF_{01} when data better support the null hypothesis and as BF_{10} when data support the alternative hypothesis (note that '01' becomes '10'). For example, the statement $BF_{10} = 2.4$ means that the data are 2.4:1 in favour of the alternative hypothesis, so 2.4 times more likely to occur under a model including the corresponding effect. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons[107, 210]. Analyses were performed with IBM SPSS v25, R studio 1.2 and JASP (Bayesian statistics) [107, 203].

4.2.2 Tasks and results

4.2.2.1 Familiarisation

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

4.2.2.2 Subjective experience

The subjective experience is assessed in Block 01 and in Block 04 (See Figure 4.1) through a within subjects design (N = 21).

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

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

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

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

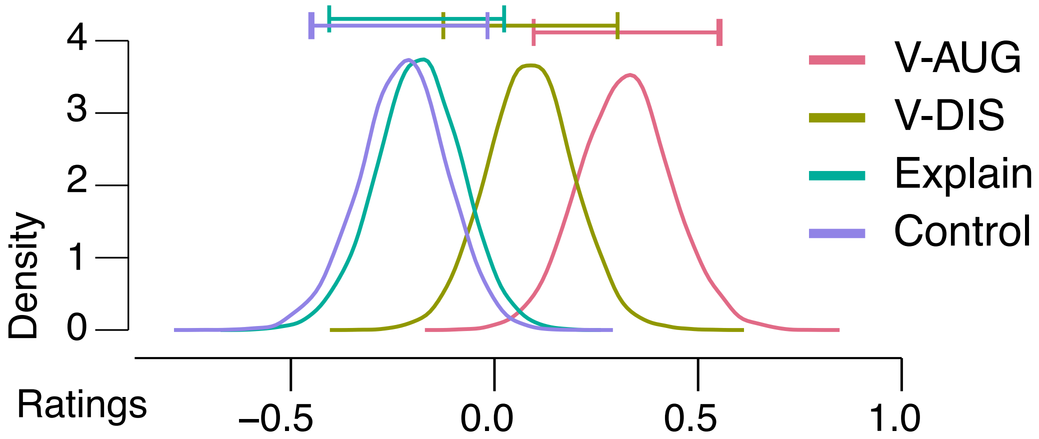


Figure 4.4: Bayesian posterior distributions of the global experience ratings in Block 04. Participants reported a better experience when the performance was augmented with visual augmentations. The *V-DIS* (incongruent visual augmentations) and *Explain* conditions did not differ from the control condition.

A strong effect ($BF_{10} = 19.5$) of the SEATs was found on global experience ($\chi^2(3) = 17.40, p < 0.001$) only in Block 04. As illustrated on Figure 4.4, post-hoc tests showed that ratings in the *V-AUG* condition presented very strong evidence ($BF_{10} = 29.7$) of a superiority with the control condition and a substantial superiority ($BF_{10} = 3.3$) with the *Explain* condition. No difference ($BF_{01} = 1.9$) with the *V-DIS* condition was detected. However, the data showed evidence for the absence of difference between the control condition and the *V-DIS* condition ($BF_{01} = 2$) and the *Explain* condition ($BF_{01} = 3.6$).

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

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

4.2.2.3 Subjective comprehension

The subjective comprehension is assessed in Block 03 (See Figure 4.1) through a within-subjects design (N=21).

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

4.2.2.3.2 Task: Bellotti-Fyans challenges This task is built upon 5 communication design issues _ namely Address, Alignment, Attention, Accident and Action _ introduced by Bellotti et al. [9] in a user-computer interaction context and transposed by Fyans et al. [77] to the perspective of spectators of interactions. In turn, we transpose these challenges from a design perspective to an evaluation perspective by assessing the perception of these design features by observers. We call them the Bellotti-Fyans challenges.

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

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

• "In this video, I can see if the musician is controlling the instrument or if he is not."(Action)

• "In this video, I can see when the instrument is correctly functioning and when it is not."(Alignment)

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

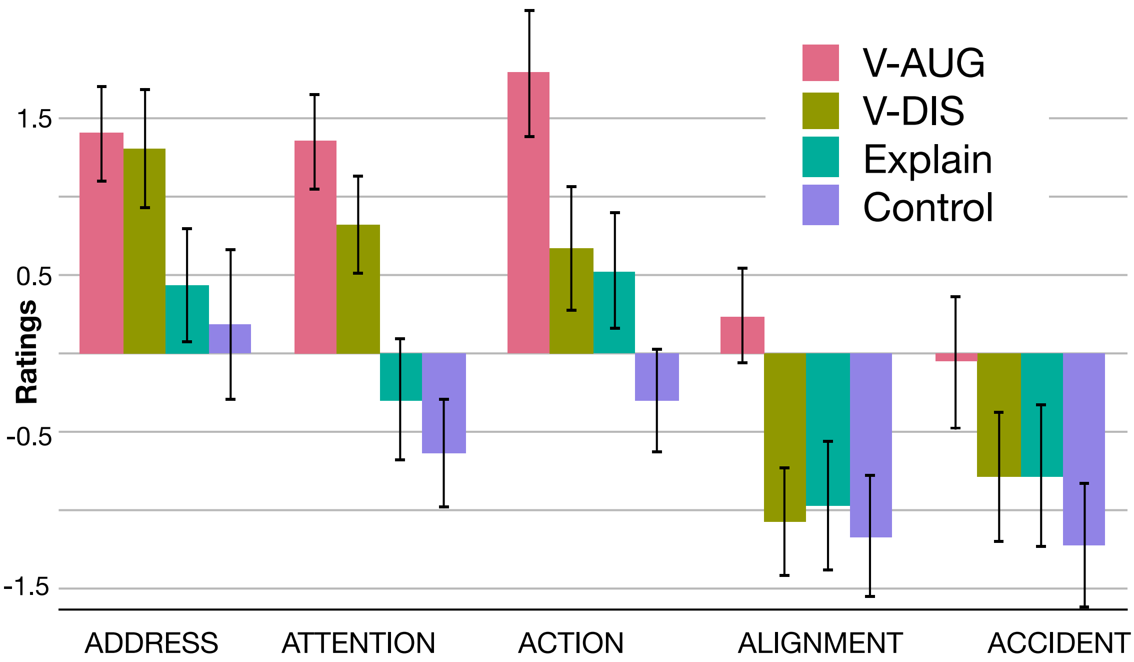


Figure 4.5: The effect of the Spectator Experience Augmentation Techniques (SEATs) on the Bellotti-Fyans challenges.

4.2.2.3.3 Results The augmentation techniques (SEATs) often led to a distinct appreciation of a same interaction (See Figure 4.5). Despite rather subtle differences in the meaning of the questions, like in *Alignment* versus *Attention*, different patterns are clearly observed.

• Address: A substantial effect ($BF_{10} = 3.7$) of the SEATs was detected on the evaluation of the *Address* challenge ($\chi^2(3) = 10.07, p = 0.018$). Bayesian post-hoc tests showed weak to substantial evidence of a better evaluation for the *V-AUG* and the *V-DIS* conditions compared to the other conditions ($BF_{10} = 2$).

• Attention: A very strong effect ($BF_{10} = 160$) of the SEATs was detected on the evaluation of the *Attention* challenge ($\chi^2(3) = 12.95, p = 0.005$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the *Control* condition ($p = 0.02, BF_{10} = 99$) and to the *Explain* condition ($p = 0.051, BF_{10} = 13.8$) and weak evidence ($BF_{10} = 2.5$) for a better evaluation in the *V-DIS* condition compared to control.

• Action: A very strong effect ($BF_{10} = 161$) of the SEATs was detected on the evaluation of the *Action* challenge ($\chi^2(3) = 18.75, p < 0.001$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the control condition ($p = 0.03, BF_{10} = 112$), to the *Explain* condition ($p = 0.02, BF_{10} = 104$) and, with weak evidence, to the *V-DIS* condition ($BF_{10} = 2.7$).

- Alignment: A substantial effect ($BF_{10} = 5.12$) of SEATs was detected on the evaluation of the *Alignment* challenge ($\chi^2(3) = 10.99, p = 0.012$). Post-hoc tests showed a better evaluation in the *V-AUG* condition compared to the control condition ($p = 0.051, BF_{10} = 9$), the *Explain* condition ($BF_{10} = 2.8$) and quite interestingly a substantial difference ($BF_{10} = 4.4$) with the *V-DIS* condition.
- Accident: Data showed no effect ($BF_{01} = 2.4$) of the SEATs on the evaluation of the *Accident* challenge ($\chi^2(3) = 3.66, p = 0.301$).

4.2.2.3.4 Discussion The assessment of the subjective comprehension revealed interesting contrasted results. By analysing the tendency of a spectator to project their ability in a fictive task, we are obviously not measuring their objective abilities but rather the confidence they have in the mental model they built up from the interaction [78]. The SEATs did not significantly influence the participants' subjective feeling of being able to detect errors from the musicians or the machine. From a cognitive point of view, this suggests that neither the visual augmentations nor the pre-performance explanations could reduce the spectator's difficulties in perceiving the performer's intention and predicting the result of an action [77].

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

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

4.2.2.4 Objective comprehension

The objective comprehension was assessed by two tasks, a *modulation attribution task* in Block 02 and a *Global Contribution Task* in Block 03 (See Figure 4.1).

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

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

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

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

Participants reported that the task was quite difficult. The data confirmed that impression with an average score of 42.3% globally (see Table 4.2 for details).

Despite concomitant results in favour of the visual augmentations (Table 4.2), the task did not permit to expose significant contrasts of the SEATs' influence on the participants' scores ($F_{(2,17)} = 3.215, p = 0.184$). This lack of power is mainly due to the rather small size of the groups and the difficulty of the task. The Bayesian analysis showed weak evidence in favour of an effect of the SEATs ($\mathrm{BF}_{10} = 1.6$) and weak superiority of the scores in the *V-AUG* condition. No influence of the SEATs on the reported difficulty ($F_{(2,17)} = 0.257, p = 0.575$, $BF_{01} = 3.1$) was found, neither on the reported confidence ($F_{(2,17)} = 2.952, p = 0.220$ - $\mathrm{BF}_{10} = 1.3$).

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

Table 4.2: Modulation attribution task - Descriptive statistics

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

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

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

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

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

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

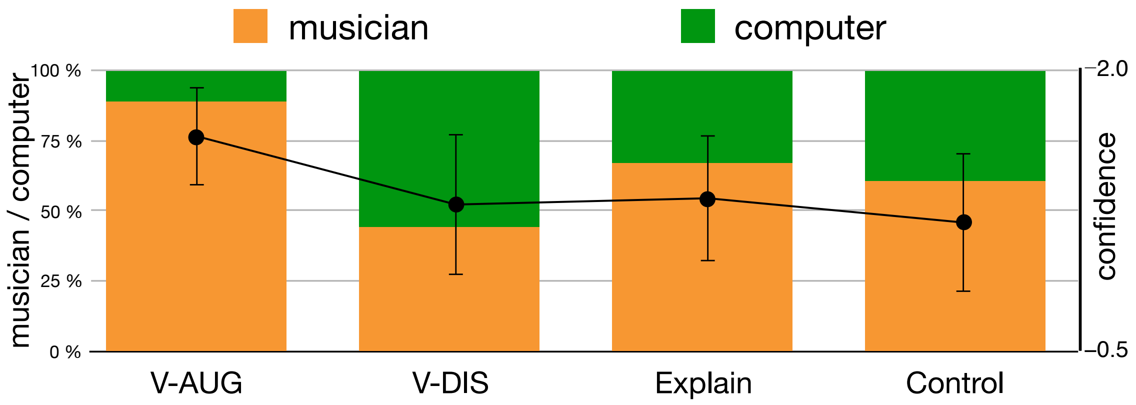


Figure 4.6: Global Contribution Task: the participants largely promoted the greater contribution of the musician (90%, *orange*) and were more confident in their answers *(black dots)* when the performances were augmented with visual augmentations (*V-AUG*). The performances augmented with incongruent visual augmentations (*V-DIS*) led to an under estimation of the musician contribution.

Considering the score of the participants to this task, the data showed a strong effect ($F_{(2,40)} = 7.18, p = 0.002$ - $BF_{10} = 600$) of the contribution level taken alone and a very strong effect ($F_{(6,120)} = 5.98, p < 0.001$ - $BF_{10} > 100000$) of the interaction between the contribution levels and the augmentation techniques. A second analysis was run to understand the details of the interaction. It exposed a very strong link ($BF_{10} > 7000$) between the choice of the participants for the most contributing and the SEATs ($\chi^2_{(4)} = 28.432, p < 0.001$). As illustrated in Figure 4.6, post-hoc tests showed that 90% of the stimuli presented in the *V-AUG* condition were associated to a greater reported contribution of the musician, 67% in the *Explain* condition, 60% for the control condition. Interestingly, the lowest count of reports for the contribution of the musician was measured in the *V-DIS* condition where 56% of the stimuli were associated with a greater contribution of the computer.

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

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

4.2.3 Discussion

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

4.2.3.1 Visual augmentations offer a better experience

The findings of this study expose the positive role of the visual augmentations over multiple aspects of the experience of spectators of performances with digital instruments. They also confirm the weak contribution of the preliminary explanations reported in the literature [16].

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

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

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

4.2.3.2 Restoring the trust in musicians

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

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

4.2.3.2.1 The unexpected benefits of disruptive cues The insertion of uncorrelated visual augmentations led the participants to evaluate differently some aspects of the interactions but not all. The confusion is more visible in the results of the Bellotti-Fyans challenges (See the discussion of the *Subjective comprehension* section). These contrasts in the results reflect the composite nature of the mental models at the origin of the spectator's experience and expectations [101, 62]. They also highlight the inconsistent effect of the augmentations used in the study. Now that the potential efficiency of the visual augmentations is formally supported, design research should compare different types of visual augmentations, graphical material or LODs and explore their influence on identified observer's mental models like the one from the Bellotti-Fyans challenges or the components of the experience we addressed in this work.

4.2.3.2.2 Controlling the level of magic We saw that can lead to the perception of a lower contribution of the musician. They also tend to blur the mental models of the observers. This property could be manipulated by the musicians. By designing the SEATs to increase the amount of disruptive cues in the augmentations, musicians could decide to veil their manipulations at the same time of a music climax and let the audience think that something "magical" is happening. In a sense, if too much doubt about the musician's contribution on the music can reduce the interest of the audience, a parsimonious approach can lead to interesting strategies in the artistic performance, like a navigation in the 2-dimensional space proposed by Reeves et al.[164] (See Figure 4.7).

4.2.3.3 Call for a new design challenge: Association

The Bellotti-Fyans challenges provided insightful data for the evaluation of the subjective experience of the spectators. However, it is not clear if these challenges are sufficient to characterise the ability of a device to expose the contributions of both the user and the system to the audience. To this end, we propose to add another "A" to the five already present in the Bellotti-Fyans challenges with the "Association" challenge. The term *Association* relates to the design decisions that enable a device to expose to the spectators the respective and shared contributions of its user and the system itself. This challenge also relates to the exclusivity dimension of agency defined by Wegner et al. [208], transposed to the audience.

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

4.2.3.4 Limitations

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

As other "in the lab" studies, our work relies on the data analysis methods for controlled experimentation that can be considered as a risk for misleading or overly narrow results [153]. In consequence, we chose to include interviews and to mix objective and subjective trials to limit the bias often met in field studies [29]. Besides, we believe that the extension of the data analysis with Bayesian statistics is particularly helpful in this context of medium sample size (21 participants).

Finally, another limitation of our work naturally lies in the evaluation of a live phenomena in a controlled environment. In this context, important aspects of the experience of live music are missing, from the social environment to the ceremonial of the concert hall. Aware of these limitations, we do not consider this work as an exhaustive approach, much to the contrary as we are very open to the contributions of other disciplines in the study of the relations between humans and artificial systems. Joining the "movement" of mixed-methods [52], we believe that there should be more dialogue between the "in the lab" and "in the wild" paradigms. This enthusiasm is shared by recent discussions in the HCI community [135], which highlights the importance of identifying insights and biases of each paradigm.

4.2.3.4.1 Summary In this section, we presented a controlled study to evaluate the effect of Spectator Experience Augmentation Techniques on the audience experience. We showed that, contrary to pre-concert explanations, visual augmentations increase the global subjective experience. In particular, visual augmentations support the observer's confidence in their mental representations of the interactions, and induce an over estimation of the musician's contribution (Attributed Agency) compared to the machine's and consequently restore spectators' trust in electronic musicians. Finally, we proposed to extend the Bellotti-Fyans challenges

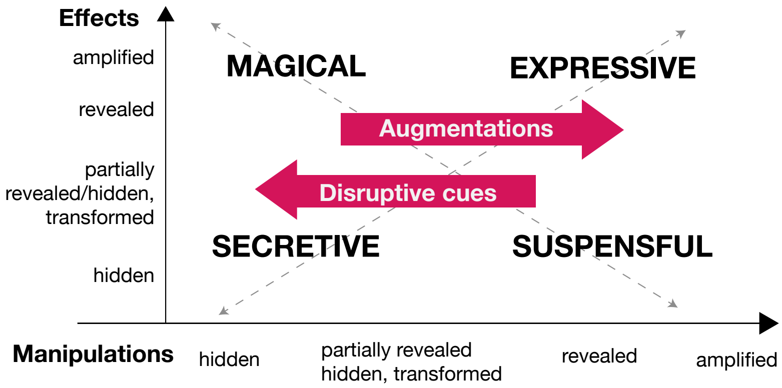


Figure 4.7: When controlled by the musician, the nature and the amount of visual augmentations allow to navigate in the 2-dimensional space proposed by Reeves et al.[164] (Illustration adapted from [164])

with the *Association* challenge, as the ability of a device to expose the contributions of both the user and the system to the audience.

In the next section, we apply the methodology presented here in order to further investigate how visual augmentations influence spectator experience and to explore the role of expertise.

4.3 The role of expertise and levels of detail of visual augmentations

4.3.1 Introduction

In the previous section, we compared two Spectator Experience Augmentation Techniques (SEATs) and showed the efficiency of visual augmentations on multiple aspects of the spectator experience. In this section, we want to refine these first insights about visual augmentations. Indeed, as we saw in 1.3.5, even if techniques explore different modalities (visual, haptic, auditory) and address different aspects of the performance (technical, gestural, intentional), they offer the same fixed level of information to all spectators.

We believe that augmenting the audience experience implies considering spectators from an individual perspective. In fact, the amount and the kind of information needed by each spectator can differ depending on their personal sensitivity and their expertise with Digital Musical Instruments (DMIs). In this section, we propose to evaluate the role of the levels of detail (LODs) in the visual augmentations and the influence of the expertise on spectator experience. Specifically, we will extend the protocol presented in 4.2.1 and evaluate the impact of LODs and expertise on the objective comprehension and the global experience of spectators of DMI performances.

4.3.1.1 Levels of detail

The LOD approach originates from the field of computer graphics where it is used to adapt 3D models and scenes complexity in order to reduce rendering load. It can also be found in the field of information visualisation to adapt quantity of information in order to limit visual overload. In the HCI literature, LODs allow users to access different levels of complexity in the interface, such as with Zoomable User Interfaces [7], or in a musical context to build and manipulate complex musical structures [4].

In our case, LODs allow spectators to adapt the amount and the type of information provided by visual augmentations about the interactions of a musician with a DMI.

[1]All stimuli, illustration videos of the conditions, anonymised raw results, statistical analyses and implementation demos can be found here: `http://o0c.eu/0NA` (under the "All you need is LOD" folder)

4.3.1.1.1 Taxonomy

From a normative perspective with respect to the taxonomy we presented in chapter 3, LODs relate to the *Semantic Density* dimension of visual augmentations and their free modification by users corresponds to the *Full* modality of *Semantic Control*.

4.3.2 Design and implementation

In this subsection, we describe the design and implementation of visual augmentations with controllable LODs for Digital Musical Instruments (DMIs).

4.3.2.1 Visual augmentations

We saw in 4.2.1.4.3 that visual augmentations 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.

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 improve. Spectators might also prefer more or less detailed information for aesthetic reasons and at various times in the performance.

To that extent, we propose to define dedicated LODs for each section (Interface, Processes, Mappings) of the visual augmentations. These *local LODs* can be chosen independently or combined as *global* LODs such as the ones we describe in section 4.3.2.3.

4.3.2.2 Local LODs

The augmentations specifically designed for a section of the instrument are called *local LODs*. We propose 4 LODs for the Interface section, 3 levels for the Mappings section, and 5 levels for the Processes section[1]. Each *local LOD* features a level 0 in which the section is not augmented. If all three sections 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 representations proposed in our implementation are only one of the many possibilities that artists can explore.

In the **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.

In the **Mappings section**

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

Level 2 refines the connection to the parameter level.

Level 3 adds a representation of the operation or series of operations which transform sensor values into parameters values [69], e.g. scaling, inverting, combining and so on.

In the **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.

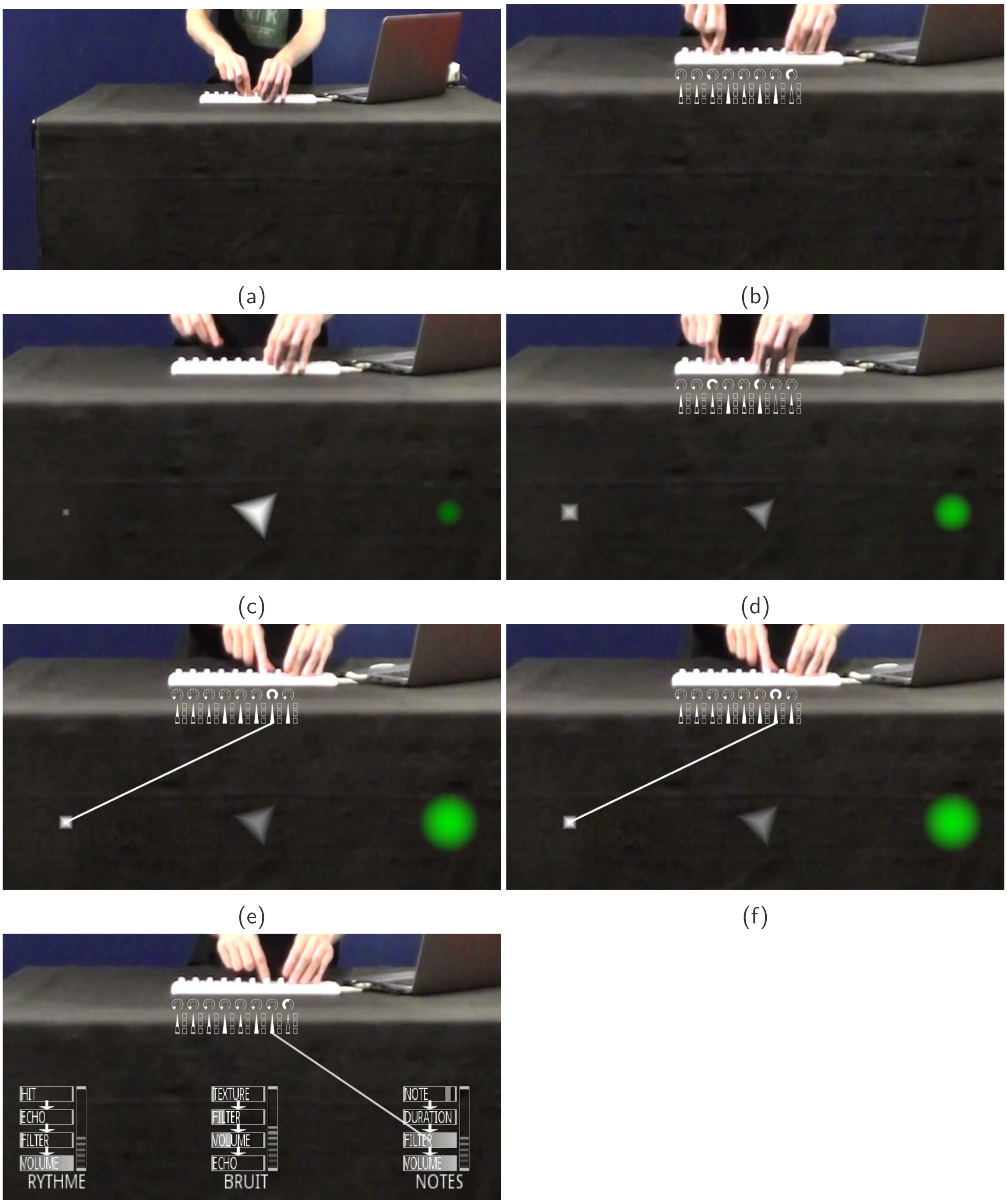
Level 3 adds a dynamic representation of parameters (i.e. inputs) that can be controlled on the processes.

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.

4.3.2.3 Global LODs

Global LODs are a combination of *local LODs*. They provide the spectators with a convenient way to control the LODs by modifying several sections at a time : Interface (I), Mappings (M)



(a) (b) (c) (d) (e) (f) (g)

Figure 4.8: 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 4.3.2.2 and 4.3.2.3).

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 4.8).

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 LODs provided by Turchet and Barthet [195] with haptics, and Perrotin et al. [160] 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 color hue, brightness with color 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. [51].

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 [14].

FULL_COMBINED (I4-M2-P3) refines both the Mappings and Processes sections. It shows which parameter 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. [15].

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.

4.3.2.4 Implementation

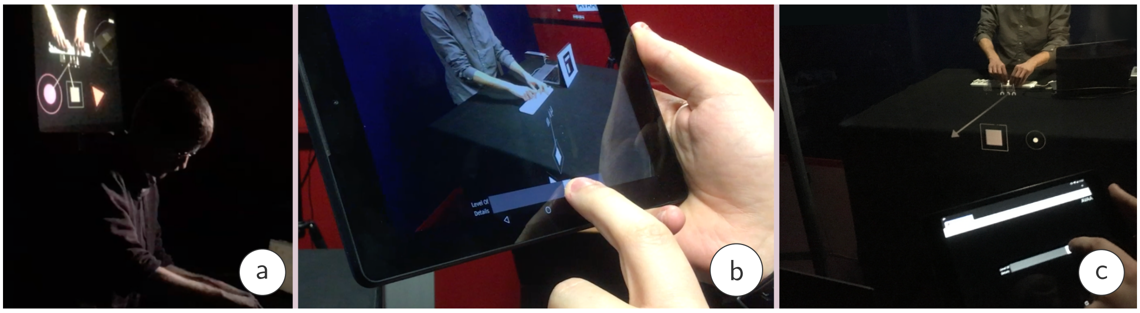


Figure 4.9: Possible implementations[1] of visual augmentations with LODs : a) Shared close-up with Video AR projected behind the musician and artist defined LOD, b) Mobile AR with individual LOD control, c) Spatial AR with an optical combiner shared between all spectators and mobile control to vote for the LOD.

The implementation of visual augmentations with controllable LODs raises questions regarding how to retrieve the information from the instrument and how to give the audience access to the LODs. In this work, the visual augmentations were implemented using the Godot game engine. The information, including activity and mappings, was retrieved in real time from the instrument, implemented with Pure Data, via OpenSoundControl messages. This extraction might however be more difficult with less open software, in which case only the lower LODs might be accessible, i.e. interface and processes activity but not the internal mappings. 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 4.9.b.

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 4.9.c, in which case viewers all perceive the augmentations spatially aligned with the physical instrument. Another possibility is to film and reproject a close-up view of the interface integrating the augmentations, as shown in Figure 4.9.a. This solution however moves the focus away from the physical performer. In these scenarios, only one LOD can be displayed at a time. 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 [214] may be used, in the form of a web interface accessible from their mobile devices, as depicted in Figure 4.9.c. 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) or optical AR (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.

4.3.3 Usage and effects of LODs

We will now present the experiment we designed to evaluate the impact of LODs on audience experience and understanding, and study 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 4.2.3.4 and plan to address social and environmental aspects of public performances in a future work.

4.3.3.1 Procedure

18 participants (16 M, 2 F) took part in the experiment, aged of mean 29 (± 7.3 , min=20, max=43). Before the beginning of the experiment, they 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

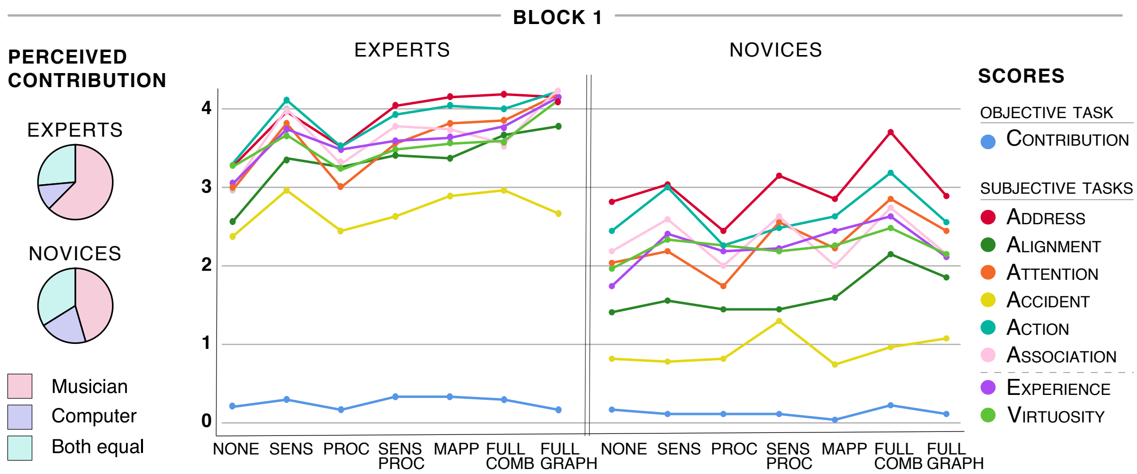


Figure 4.10: *Left* Regardless of the LOD, Experts perceived a higher contribution of the musician than Novices. *Center* LODs did not impact equally the subjective perception of the interactions. Experts reported higher evaluations of the subjective dimensions of the spectator experience.

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 4.8 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 ± 6.4 years (± 6.4, min=10, max=30) and an electronic music practice of 10.7 ± 7.3 years (± 7.3, min=2, max=25) against 1.6 ± 2.6 (± 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 (± 8.3, min=2, max=30) electronic music performances, while for novices the average was 0.6 ± 1.5 (± 1.5, min=0, max=5).

4.3.3.1.1 Dynamic stimuli The stimuli were videos of short performances with a DMI composed of a Korg NanoKontrol (Fig. 1.8) controlling a set of Pure Data patches with three sound processes (melodic, rhythm, granular texture) each with multiple parameters (See Figure 4.8).

We designed three 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

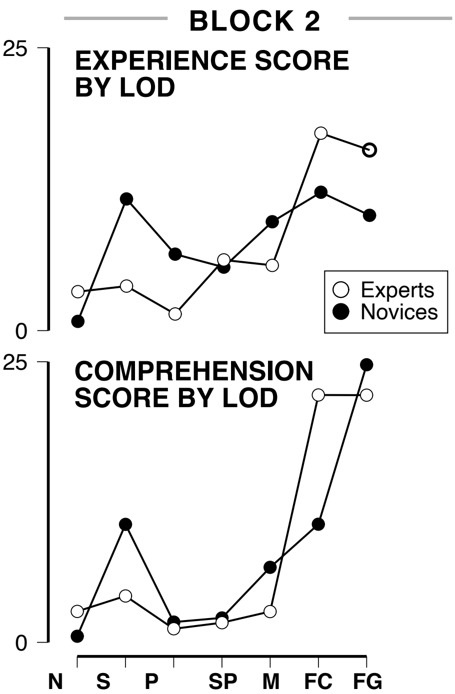


Figure 4.11: When participants could choose their favourite LODs in real time, strategies emerged as illustrated. They were refined in the interviews.

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.

4.3.3.1.2 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-randomized 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 other questions were subjective and were based on 5 communication design issues introduced by Bellotti et al [9] and transposed to the spectator perspective by Gurevitch and Fyans [93]. We complement them with *Association* (4.2.3.3) that targets the capacity to expose to spectators the contributions of the user (musician) and the system (DMI). These design challenges are well adapted to the evaluation of DMIs as they allow for an assessment by components of the subjective experience of spectators. 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 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 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.

4.3.3.1.3 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.

4.3.3.2 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. The analyses were conducted under the common frequentist paradigm and were combined to Bayesian statistics [120]. A Bayes factor is reported as BF_{01} when data better support the null hypothesis and as BF_{10} when data support the alternative hypothesis (note that '01' becomes '10'). For example, the statement $BF_{10} = 2.4$ means that the data are 2.4 times more likely to occur under a model including the corresponding effect. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons. Analyses were performed with SPSS v25, R studio 1.2 and JASP [107].

4.3.3.2.1 Block 1: fixed LODs Analysis did not revealed 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.

Still, 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 4.10 - *Left*).

As depicted in Figure 4.10 - *Center*, experts reported higher evaluations of the subjective questions. 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 graph (4.10 - *Center*) from left to right, compared to *NONE*, the control condition, *SENS*, the LOD 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. In the Experts group, the progression tends to extend to *FULL_GRAPH*. 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 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.

4.3.3.2.2 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 4.11).

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.

4.3.3.3 Discussion

4.3.3.3.1 LODs affect subjective comprehension

. The interviews confirmed and extended the quantitative analyses. Despite the absence of effect of LODs on factual perception, 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 suggests that LODs influence the subjective comprehension of spectators, 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).

4.3.3.3.2 Visual augmentations and expertise

. Our study reveals interesting insights on the nature of expertise in DMI spectators. Results of Block 1 showed that experts perceivea 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.

This contrast is confirmed in Block 2 where only novices favoured the *SENSORS* LOD over no augmentations for a better comprehension and experience (Fig. 4.11), 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.

4.3.3.3.3 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 [93]. A solution to this issue could be inspired by music video games where the virtuosity is materialised by screen indications of 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 on the virtuosity, such as visualisations of input complexity or extra-ordinary values for controls and musical parameters.

4.3.3.3.4 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.

4.3.3.3.5 Summary In this section, we investigated the design of LODs in visual augmentations and their effect on the audience experience. We also showed that experts perceive a greater contribution than novices despite similar scores for objective tasks. Both profiles develop personalised approaches of LODs and we advocate for technology development that include the possibility for spectator to change freely the LOD or turn it off. As future work, we think that augmentations with LODs should be extended to other interfaces, e.g. gestural controllers or graphical interfaces such as live-coding, and that the effect of aesthetic choices should be investigated.

While our results provide useful insights, we believe the controlled experiment approach that we took could be combined with in the wild study of performances.

4.4 Conclusion

In this chapter, we saw that controlled experiments in the lab can provide refined insights on spectator experience. In particular, we showed significant contrasts in the efficiency of pre-concert explanations and visual augmentations. We also contributed to clarify the role of expertise of spectators and proposed the first controlled study, to our knowledge, that addresses the impact of LODs in visual augmentations.

One of the striking aspects revealed by these two related studies is that experience seems to mostly rely on subjective aspects. One might argue that the subjectivity of experience does not require experimental evidence when we all have the intuition of a very personal relation to music. But with these results, we highlight the crucial role of the subjective beliefs in our perception capabilities in the experience of watching a musical performance. Even when objective data show our rather low ability to decrypt musical gestures and the inner mechanisms of DMIs during interactions, visual augmentations and expertise tend to provide cues and background that constitute the subjective support of a greater experience.

If you had a sign above every studio door saying 'This Studio is a Musical Instrument' it would make such a different approach to recording.

Brian Eno

5

Towards personalised augmentations of experience

Main contributions

- Conceptual pipeline for optimal augmentations
- Preliminary insights to support the continuous measurement of Attributed Agency through grip force

> The contribution of this chapter related to the pipeline was published in the proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR 2017) [36].

We saw in the previous chapters that visual augmentations are an effective way to improve the experience of spectators without constraining musicians to modify their interface. We also demonstrated that the levels of detail (LODs) of visual augmentations can impact spectator experience which led us to advocate for customisable setup where spectators can choose their desired type of visual augmentations. To date, the selection of the type of augmentations and the LOD relies on the active choice of the user from a limited range of augmentations.

In this chapter, we first present a conceptual pipeline that addresses the issue of optimal augmentations in live performances. We complete this conceptual contribution with the study of a potential objective marker of Attributed Agency that could integrate one of the modules of the pipeline.

5.1 PEANUT: a pipeline for spectator experience augmentations

In this section, we propose a conceptual pipeline that supports the autonomous generation of optimal visual augmentations based on a continuous evaluation of the familiarity of the audience. The process includes the extraction of data from the musical interactions and the audience, the detection of perceptive gaps and the generation of specific augmentations. To encode and process these information, we introduce a conceptual object: *the correspondence*. While this pipeline remains conceptual, it was designed so that it could be implemented by addressing the challenges described in 5.1.3.

Here is a scenario that we envision with our pipeline:

Patricia attends an electronic music concert. At the entrance, she is given a small device equipped with physiological sensors (a choice of either a bracelet or a special glass that she holds). During the concert, she has trouble understanding what is happening, in particular what the musician's action on the sound is. The device senses a change in a set of physiological signals, that corresponds to a loss in familiarity, and sends the data to a server. Patricia may also directly indicate her loss in familiarity with a graphical slider on an app on her smartphone. Simultaneously, this server has been analysing the musician's gestures, the flow of data inside the instrument and the musical output. When it receives Patricia's familiarity signals, the server, with settings defined by the musician, selects the adequate augmentations to be displayed. They aim at compensating for the familiarity disruption caused by the musical *interaction context. Consequently, visual augmentations are displayed around the musician either for Patricia alone when she watches the performance through her smartphone (e.g. with video augmented-reality), or for the group of people around her using a mixed-reality display. They provide information that improve her degree of familiarity, allowing her for example to perceive the link between the musician's gesture and the resulting sound, and to enjoy the performance to a larger extent.*

5.1.1 Correspondences.

In order to handle the heterogeneous data (physiological, behavioural, musical, visual, mechanical) associated with spectator experience, we introduce the notion of **correspondence**: a conceptual object that stands as a digital multidimensional representation of a musical interaction. It can be viewed as an initiative to imitate the internal model of spectators.

5.1.1.1 Properties

A correspondence associates several *properties* :

- *musical output*: audio
- *gesture data*: time series of 3D coordinates
- *visual data*: video of the movement
- *control data* : time series of sensor values
- *Agent*: author of the interaction. As we saw it in chapter 3, this property specifies the author of the interaction. Indeed, interactions are not only produced by one musician but can be triggered and mastered by another one in a collaborative performance or, more usually, by autonomous prepared processes, e.g. automations or playlists.

5.1.1.2 Elements

Each property is subdivided in three elements: a *raw format element*, a *semantic element*, and a *classification element*.

5.1.1.2.1 Raw format element The *raw format element* is a pointer to a collection of related files. For instance, the raw element of the visual property of a clap correspondence is a pointer to a collection of short videos showing a clap from different angles and velocities and the raw element of the audio property a pointer to a collection of short audio recordings of a clap.

5.1.1.2.2 Semantic element

The *semantic element* is composed of annotated descriptors of the property. For our clap example, the semantic element of the visual property could list obvious descriptors as "hands", "clap", "applause", "brief" but also more precise descriptors relative to specific taxonomies developed in the different analysis of gesture and sound.

The descriptors can be freely provided by the audience or taken from refined existing models such as [33] or [112]. For instance, "excitation gesture" or "effective gesture" could be used to populate the semantic element of the gesture property of a correspondence.

Interestingly, by holding descriptors for each property of a unique interaction (aka correspondence), the semantic element offers the possibility to analyse the links between descriptors from different properties by the analysis of their co-occurrence. Finally, looking forward to the implementation, the processing of meaningful descriptors can rely on the strong foundation of web semantics [12].

5.1.1.2.3 Classification element

The *classification element* is dedicated to the classification of the raw elements through machine learning methods and thus allows for the pairing of close correspondences with respect to their different properties. To compute such a distance between correspondences, multidimensional vectors can be used alongside more specific techniques as neural networks, informed by an efficient semantic analysis thanks to the semantic element of a given property.

5.1.1.3 Familiarity scores

In addition, correspondences hold a *score* for some familiarity dimensions: Attributed Agency, consistency, exclusivity, instrumentality (See chapter 2 - 2.2.1.6 for details).

For example, two correspondences with the same musical result, e.g. the fade-in of an audio loop, can have different gesture and control properties depending on the mapping chosen for the DMI. While a continuous gesture on a fader would have a high score (meaning, the most natural way to fade) for the *consistency* dimension, a discrete tap on a pad would have a low one, since the effect would no be consistent with the cause for a spectator, e.g. discrete input and continuous output.

This score of familiarity held by each correspondence plays a central role in our approach for an increase of familiarity as it allows for sorting correspondences from obvious to abstruse for a given audience. It is computed performance after performance from error estimation between the predicted score and the measures extracted from the audience.

5.1.2 The pipeline

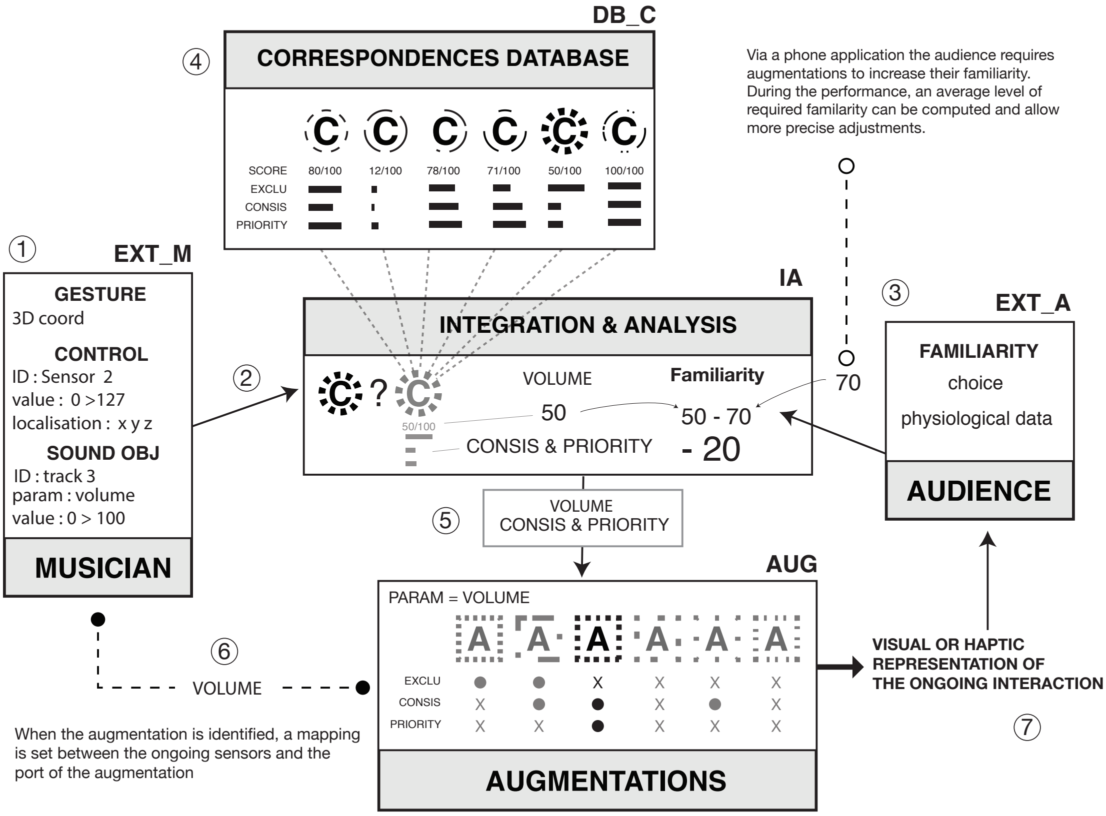


Figure 5.1: The proposed pipeline with its main modules and data flows

Our envisioned pipeline (Fig 5.1) is composed of five modules that handle the extraction of data from the musician and their instrument (*EXT_M*), the extraction of physiological and subjective data from the audience (*EXT_A*), the processing of the data (*IA*) and the selection of fitting augmentations (*AUG*) based on a database of correspondences (*DB_C*).

The pipeline is used at three different moments: before, during and after a performance.

5.1.2.0.1 Before Before the performance, correspondences, coming from a shared online database, or recorded specifically for the instrument, are saved in the *DB_C* module.

5.1.2.0.2 During
① During the performance, *IA* receives musical interaction data extracted by *EXT_M*. This data consists of both dynamic values such as gestural parameters, audio features and control values. Besides, physical (position of gestures and sensors), logical (tracks, effects, synthesizers, ...) and structural information about the musical interaction are directed to the *AUG* module. ② *IA* encode these signals into correspondence objects. ③ Simultaneously, *IA* receives the familiarity evaluation (with the identification of the sensed individual or group) from *EXT_A*. If the familiarity evaluation is low, ④ *IA* finds in *DB_C* correspondences similar to the ongoing one, and select the familiarity dimensions that need to be compensated depending on their scores in these. *IA* then sends *AUG* the data required for the augmentation: live correspondences, associated signals from *EXT_M* with dynamic and structural data and identification of the source of the familiarity evaluation (in order to display the augmentations only to the correct person or group). ⑤ *AUG* creates (or selects if it already exists) the augmentation that matches the received structural data, for example a visual augmented-reality link between the physical position of a sensor and a virtual representation of an audio track.⑥ Augmentations are then connected to *EXT_M* and listen to the signals required to update the augmentation, e.g. control values for the sensor, loudness of the track.

5.1.2.0.3 After
After the performance, the familiarity extracted from the audience can be reused to refine the scores for the familiarity dimensions in each detected correspondence of *DB_C*.

5.1.3 Modules

5.1.3.1 Database of correspondences (DB_C).

DB_C manages all pre-existing correspondences. These can be generic, or specific to an instrument or performance. It receives queries from *IA* to select correspondences matching the ones detected during the performance.

5.1.3.1.1 Web architecture of databases
A very promising approach in dealing with heterogeneous data is the use of databases. Even if the efficiency of machine learning and analysis tools is still evolving, a numerous amount of initiatives, especially in analysis of emotion (DEAP[124], RECOLA[169], EATMINT [42]) but also in music-related actions [91], contribute to shared databases that compile multimodal and synchronized experimental data.

The main goal of these databases is to predict complex and abstract states, e.g. the emotional state of an individual, thanks to the analysis of their physiological and behavioural signals such as face expression, electrodermal activity or heart rate variation. Most of the existing databases are composed of 15 to 30 entries referencing data of diversified nature.

5.1.3.1.2 Web tools to gather data To allow the gathering of a more relevant amount of cross data, we first need to facilitate the indexing thanks to the more and more intuitive and effective front-end technologies of the web. We propose to develop a web interface that could provide the specific tools to aggregate the data required to constitute the properties of a correspondence. A typical correspondence would require a short video footage of a gesture, the motion capture of this gesture, sensor value and audio output of the instrument. Tags could also be manually added for each of these properties.

The interface could be accessible from an open web platform where artists as well as researchers could populate the database, to constitute their own correspondences and therefore optimize their pipeline with more personal choices of gestures, mappings or sound processes.

5.1.3.1.3 Online evaluation sessions The same online platform can then be used for crowdsourced online evaluation sessions in which correspondences are exposed to participants with different levels of expertise. Their task is to indicate their understanding, using a questionnaire along the familiarity dimensions, and tag the correspondences.

5.1.3.2 Extraction of the musician's interactions (EXT_M).

As explained in Section 5.1.2, *EXT_M* extracts data from the instrument and musician's gesture which is then sent to *IA* in order to be aggregated into *live correspondences*.

5.1.3.2.1 Large range of interest variables In the way we extracted data from musical interaction in our studies (Chapter 4), *EXT_M* extracts sensors states, mapping values as well as musical result of the musician interactions. At the musician level, *EXT_M* extracts the control gestures, the body movements and physiological signals. In addition to these signals, structural information needs to be provided for further use in the augmentations, such as the position of the physical sensors of the DMI and the position of the musician's hands, the list of tracks, effects or other sound processes with their names, or the mappings between sensors and sound parameters. While some of these obviously need to be defined by the musician manually, or sensed by devices external to the instrument, others can be extracted through a trans-disciplinary approach.

5.1.3.2.2 Current developments Regarding the instrument input, research on gesture recognition, especially concerning hand gestures as demonstrated by Rautaray & Agrawal [162], can be used to identify the performed musical gestures. Regarding the instrument output, research in music information retrieval (MIR) provides tools for segmenting music from the audio signal only [159] using spectral, tonal, rhythmical descriptors and methods. Finally, the extraction can be facilitated by a multimodal approach, similarly to techniques developed for video indexing [188], as some events might appear clearer on one channel than on others. For example, control data might inform on the temporal boundaries of a change in the sound that can then be analysed

5.1.3.2.3 Challenges We identify two main challenges for the implementation of this module.

The first challenge is the access to the data from the DMIs. In fact, while extraction from the audio signal provides many features that can be used to detect correspondences, it might not be enough for precise analysis. In order to access pre-mappings and post-mappings data, to differentiate between manual and automated changes and to analyse the output of individual tracks or other sound processes, one can not rely solely on the DMIs inputs and outputs, i.e. additional software components will need to be integrated.

In most DMIs, plug-ins can be added at various stages of the instrument. However, the API might not provide enough information on the instrument to a single plugin. For example, one plugin per track might be needed to access and send the audio output features for each separately. The integration of *EXT_M* will be simpler if DMIs are built using patching environments such as PureData or Max/MSP, where the musician has more control over the architecture of the instruments.

The second challenge is to combine detailed but costly and slow analysis of features for the detection of correspondences with maximum accuracy, and fast but less accurate analysis of features for the update of augmentations in *AUG*.

5.1.3.3 Extraction of the audience familiarity (EXT_A).

The role of this module is to extract the audience subjective and objective information in order to inform *IA*.

We saw in chapter chapter 1 that familiarity is one of the key aspects of the experience of live music relies on multiple underlying mechanisms. Therefore, to initiate ecological (i.e. out of the lab, in "real life") measurements of physiological signals, we propose a dual methodology to extract the familiarity of the audience.

5.1.3.3.1 Subjective assessment of familiarity First, we base the subjective assessment on a familiarity application for mobile phone. Its main purpose is to supply *IA* with data about the ongoing familiarity of the audience, from a graphical familiarity slider that spectators activate.

As other authors [6][19], we believe this continuous survey could be a good methodological answer to the reservations we exposed about questionnaires after the performance.

5.1.3.3.2 Objective assessment of familiarity The second part of the extraction relies on physiological signals. Neuroscience studies show that the expertise, a key component of familiarity, influences the perception of action [34]. Those findings, applied to music expertise, may lead to a better understanding of its role in the live music experience. By measuring the peripheral signals, we aim at discovering potential patterns that could correlate with the subjective data we extract. Widely used in the emotion studies, and rather suitable to extract in natural condition, several signals are particularly interesting in our musical context: Heart rate variability [178], electrodermal activity [182], oculometry (eye tracking + pupillometry) and the grip force that we will develop in the next part of this chapter (Section 5.2). The complementary analysis of extracted features of such signals already gave interesting result in the emotion classification by machine learning algorithms [122] and need to be further extended in music experience studies. Moreover, these signals can already be acquired by wearable devices and the quantified self movement [192] will surely provide more accurate and affordable devices in the near future.

5.1.3.4 Integration and Analysis (IA).

IA is the central hub that connects to all the other modules. Its role is to:

- compute the live correspondences with the data extracted
- match the computed live correspondence with a correspondence from the database to calculate the familiarity dimensions that need to be augmented;
- supply the *AUG* module with the information needed for the relevant augmentation selection.

5.1.3.4.1 Classification support To fulfil these tasks, the module can rely on the classification element of each property of correspondences. The classification element is a machine learning model dedicated to the classification of raw data. This model is pre-computed with the raw files registered in the correspondence. Its goal is to discriminate new stimuli and detect those who match with the recorded one. Considering the heterogeneity of the modalities, each property might require a specific machine learning model and specific extracted features.

For example, MIR descriptors for a sound element and a deep convolutional network for picture classification. The main idea is to use the set of models as a global digital representation of the correspondence that can either be projected, depending of the context, on a single and more easy to handle property or be represented as a multidimensional vector that allows similarity comparison of whole correspondences (the matching process).

Considering the variety of data and processes it has to handle, this module needs to be regularly updated with recent findings in signal processing, machine learning or movement models. Without a strongly modular structure of the available tools that the pipeline has at its disposal, the framework may not be able to evolve and thereby join a long list of deprecated initiatives.

5.1.3.5 Augmentation (AUG).

This module manages both a database of available augmentations and a set of active ones.

5.1.3.5.1 Selection When a correspondence needs to be augmented, *AUG* receives the data required to create a new or select an existing augmentation, such as the physical position of the gesture and sensors and logical components of the instruments, the familiarity dimensions that need to be compensated, the destination of the augmentation (individual or group) and the data from *EXT_M* that the augmentation should listen to. The augmentations are selected from a database of augmentations designed to compensate the various dimensions of familiarity.

5.1.3.5.2 Challenges

A first challenge is the creation of a framework that allows one to design augmentations according to the specific dimensions of familiarity that they compensate. Rules will need to be defined so that one can adapt an augmentation to the artistic specificity of a particular performance.

A second challenge is the design of augmentations that provide just enough content to fill the multimodal gap without distracting the audience from the musical performance because of a too heavy cognitive load.

We believe that the development of the *AUG* module could start from the taxonomy we introduced in this thesis (chapter 3).

5.1.4 Synthesis

In this section, we presented a conceptual pipeline for augmenting familiarity of the audience with Digital Musical Instruments and reviewed associated research results and challenges. We introduced the notion of *correspondence* as an initiative to mimic the internal representations spectators have of a musical interaction. Integrating data from musicians, instruments and audience, correspondences allow for the selection of optimal augmentations to target specific perceptive gaps and improve the familiarity of spectators with the ongoing interactions. Among the presented modules and the central role of artificial intelligence, the continuous measurement of electrophysiological markers of familiarity components like Attributed Agency is crucial.

In the next section, we will introduce a first contribution to the practical development of this conceptual pipeline with a study of a potential objective marker of Attributed Agency.

5.2 Measuring the attributed agency through GFv

The measurement of objective markers of Attributed Agency is a great source of innovations in terms of spectator experience. **In this section, we build on the embodiment theory and previous related studies to proposea hypothesis: watching a musical performance triggers unconscious action simulation mechanisms whose amplitude could be linked to the level of Attributed Agency the observers perceive from a musical interaction with Digital Musical Instruments (DMIs).**

We saw in chapter 1 that expectation [101] is one the components of music experience. Furthermore, we know from the work of neuropsychologists like Jeannerod [108] or Grafton [92] that the motor system is involved in the prediction of goal-directed actions. Extending these conclusions, a recent study by Blampain et al. shows that the action simulation can be tracked through the GFv (GFv) exerted by a participant holding a sensor device in his dominant hand while watching or mentally visualising an action [23], in that case a hand crushing a strawberry with a slap (Fig. 5.2).

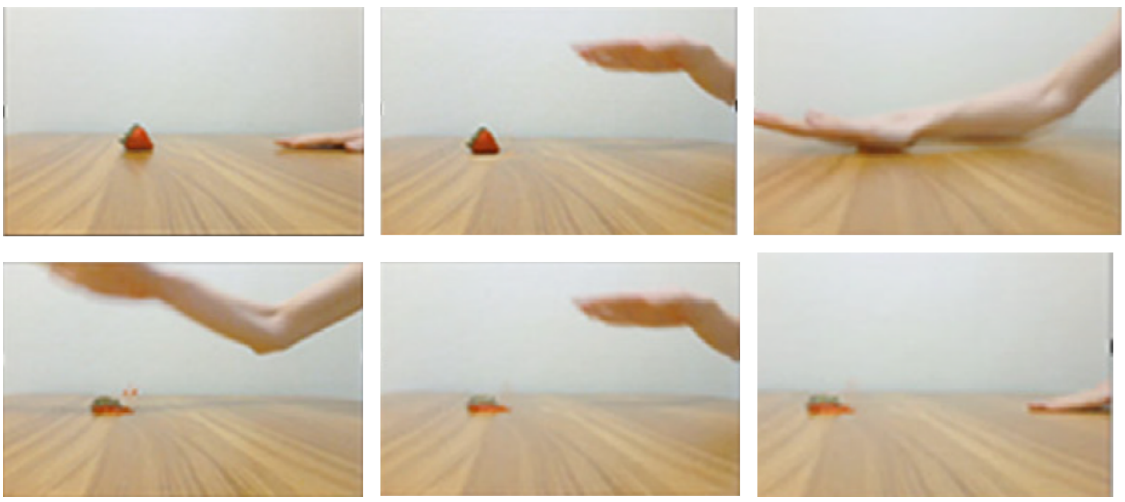


Figure 5.2: Among the interactions studied in Blampain et al. [23], *High Intensity actions* like a hand crushing a strawberry trigger the largest GFv. From [23]

From this background, we propose a grip force experimentation in the context of musical interactions. With this methodological transposition from previous studies to our musical context, we want to explore the efficiency of the setup to study the perception of interaction parameters. The results presented in this section come from a preliminary study on the multimodality of musical interactions and the role of visual augmentations. As our sample size for this first exploration is rather small (n=5), we will not expose nor discuss too subtle aspects.

We invite the reader to consider this material as a first glance to potential methodological benefits in the objective measurement of spectator experience.

5.2.1 The study

5.2.1.1 Participants

Five participants were recruited (3 M, 2 F, 29.9 y.o. ± 6.6). One is a confirmed electronic musician, the others do not have extensive musical education or practice. Before starting the experiment participants were informed about the overall course of the experiment and about their rights and conditions to access and remove their personal data. After having signed an informed consent they answered a short questionnaire about their musical skills.

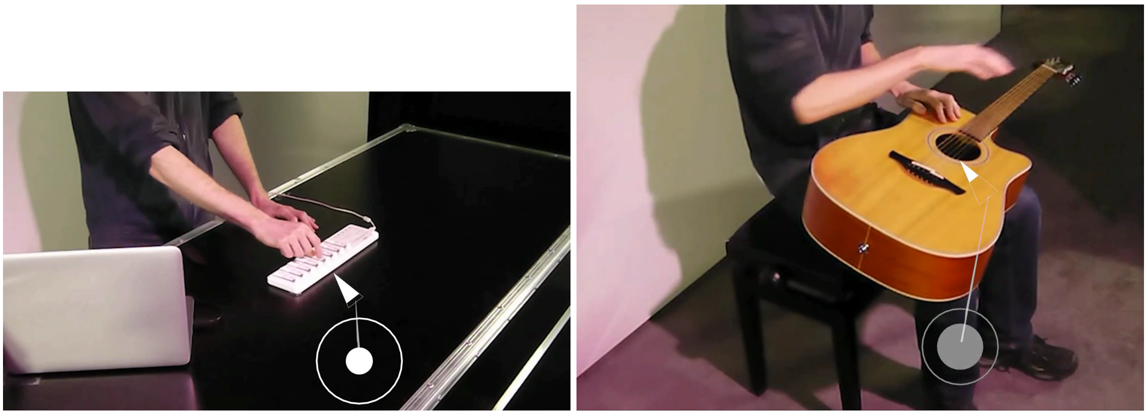
The exclusion criteria were hearing and vision impairments. After control, no participants were excluded.

5.2.1.2 Stimuli

5.2.1.2.1 Unit Perceived Units The stimuli[1]used in this study are brief (4s) single-gesture musical interactions that we call Unit Perceived Interactions (UPIs, see chapter 22.2.2 for more details). To address the grip force response to the multimodality of musical interactions, we used audio only, video only and audiovisual UPIs. The audio and video only stimuli were exported from the audiovisual UPIs. The whole set includes: 18 stimuli, 7 audiovisual interactions (AV), 6 video only interactions (V) and 5 audio footages (A).

5.2.1.2.2 Transparency and Visual augmentations We used stimuli with variable transparency (i.e. "natural" feel of the interaction, see chapter 1 - 1.3.2.2). The most transparent interaction is a single key stroke on a piano keyboard, the least transparent interaction is a circular mid-air gesture over a tilted acoustic guitar (Fig. 5.3b).

To modulate the transparency of an interaction with a same instrument, we modified the parameter controlled by the musician. For instance, in the interaction with the MIDI controller NanoKorg, the musician pushes on of its buttons. In the transparent condition, the gesture triggers a sound with the push and the sound stops when the musician releases the button. In the non transparent condition, a continuous sound can be heard from the beginning of the stimulus (autonomous process). When the musician pushes the button a light parameter



(a) The Nanokorg MIDI controller and a visual augmentation of the interaction.

(b) Mid-air gesture over a guitar with visual augmentation

Figure 5.3

alteration is applied to the autonomous process and stops when the musician releases the button.

To explore the grip force response to augmented interactions, the less transparent interactions could be displayed with or without visual augmentations.

5.2.1.3 Tasks

The tasks mainly consist in watching videos or listening to brief audio sequences while holding a grip force sensor (Fig. 5.4). The experiment is divided in two blocks with the same stimuli in each block.

5.2.1.3.1 Passive and active tasks In the first block, participants just have to passively watch or listen to stimuli while holding a grip force sensor. In the second block, in addition to holding the sensor, they are asked to mentally visualise the ongoing interaction. For audio stimuli, the instruction is to imagine the gesture that could produce the sound. For video stimuli, the instruction is to visualise the interaction without its audio outcome. For audiovisual stimuli, the instruction is to visualise the interaction and its audio outcome.

5.2.1.3.2 Trials Before the beginning of a trial, participants are instructed to take a few seconds to position their arm, holding the grip force sensor until the experimenter calibrates their baseline grip force. A trial consists in the exposition of a stimulus, five times in a row, followed by a series of questions. We repeat the stimuli to compensate the potential surprise of the first exposition and gather cumulative data to reinforce further analyses. After each trial, participants can lay the sensor on the table to relax their hand.

5.2.1.3.3 Blocks A block consists in the successive presentation of all the trials from a same modality. Audio (A) only and video only (V) trials were randomly balanced but the audiovisual trials (AV) always occurred in the third position, leading to two possible sequences, A - V - AV or V - A - AV.

The first block contains the passive tasks, the second block the mental simulation tasks.

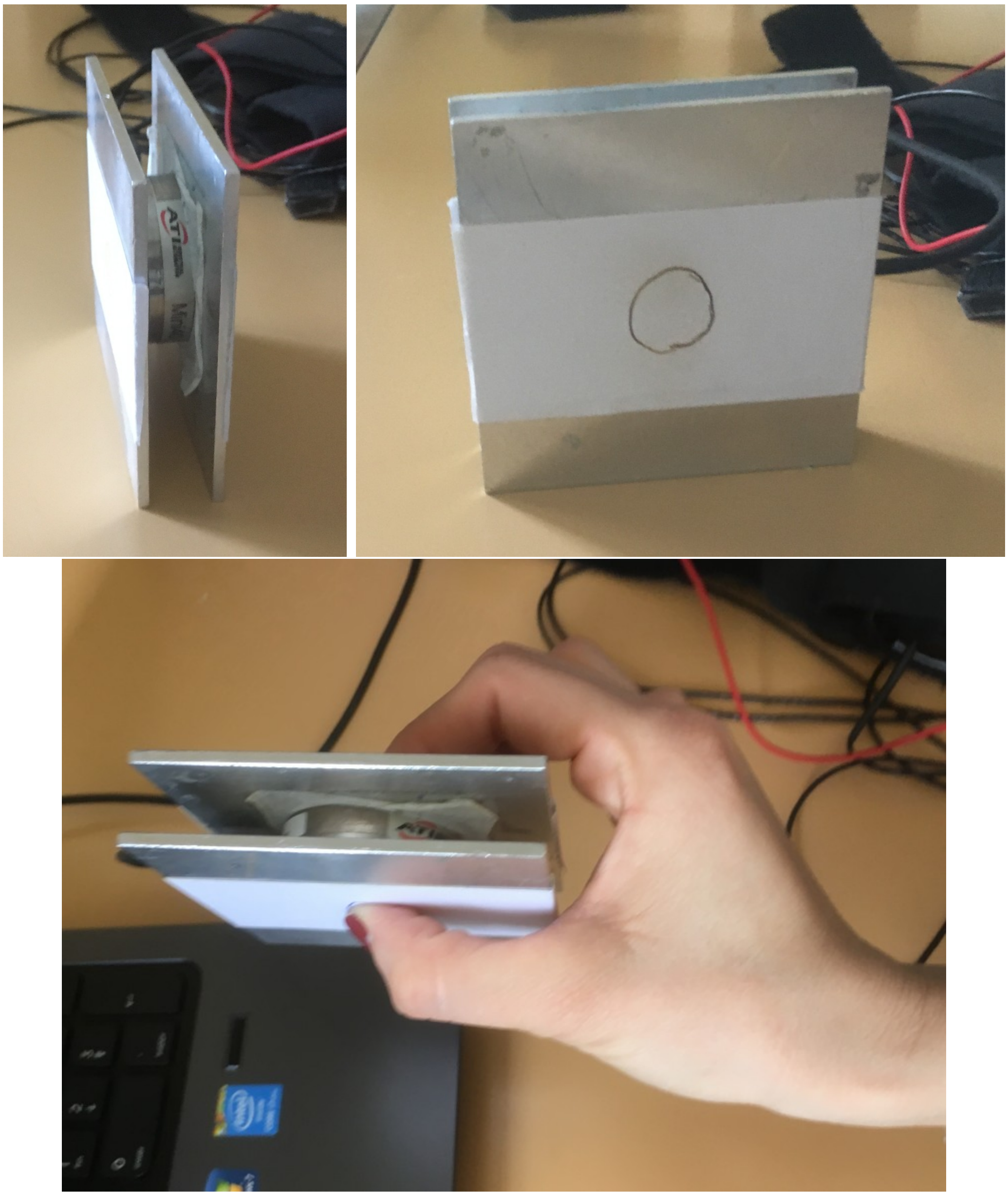


Figure 5.4: *Top left* - The grip force sensor is made of two squared metal surfaces enclosing a pressure sensor. *Top right* - A circle indicating the position of the thumb. *Bottom* - The device is held between the thumb and two fingers (index and middle fingers)

[1] All stimuli available here: `https://o0c.eu/AA0`

5.2.2 Results

5.2.2.1 Signal filter

We followed the data preparation procedure proposed by Blampain et al.[23] and applied the same thresholds and algorithms for artefacts filters. The baseline was tested and no significant difference was observed between modalities (audio-only, video-only, audio-video ; F=.289, p=0.89) indicating a proper calibration and the absence of shifting in the data.

5.2.2.2 Fatigue

We ignored and did not assess the fatigue caused by the experiment as previous and more engaging study showed no particular fatigue. Indeed, in Blampain et al. [23], participants had to hold the sensor 3 min per block before resting. Authors assessed the fatigue score and no variation throughout the whole experiment was detected. In this study, participants had to hold the sensor during a maximum of calibration + 5 repetitions x 4s stimulus $\approx$ 25 seconds in the exact same position than documented before releasing the device.

5.2.2.3 Global overview

As depicted on Fig. 5.5a, the grip force variations (GFv) we measured show an increase pressure shortly after the beginning of the interaction confirming the reproduction of the main outcome observed in Blampain's study (Fig. 5.5b). Please note that on our graph, t=0 ms corresponds to the moment when the musician begins their movement while Blampain's data are aligned on the moment the hand touches the fruit.

5.2.2.4 Time normalisation

Regarding our small sample size (n=5), we do not present too subtle comparisons as they require cumulative data for each condition.

Besides, the UPIs we chose as stimuli do not have the exact same timing sequence. They are all composed of the same successive events but these events do not last exactly the same duration. In order to gather as much signal as possible from these interactions with variable durations, we normalised the data between the events common to each stimulus:

- STIM_START_TIME: beginning of the stimuli, the musician is in still position
- MovStartTime: initiation of the musician's movement

- MusicianTrigStartTime: the musician starts to trigger audio outcomes
- MusicianTrigStartTime: the musician stops to trigger audio outcomes
- MovEndTime: end of the musician's movement, return to still position

5.2.2.5 Impact of transparency

With time normalisation, we have rather sufficient material for a specific analysis that contrasts the GFv between transparent and not transparent interactions.

On Figure 5.6, our preliminary results [2]indicate that during the time window when the musician actually triggers sound [MusicianTrigStartTime - MusicianTrigStopTime], the GFv is higher for transparent interactions. Besides, we observe a delay of the peak of GFv in non transparent interactions that could also be less important. In the last sequence between [MusicianTrigStopTime] and the end of the movement [MovEndTime], the average GFv looks higher for non transparent interactions.

5.2.2.6 Impact of visual augmentations

On Figure 5.7, we compare transparent, non transparent, and augmented non transparent interactions with a MIDI controller (Fig. 5.3a). Our data seem to indicate a gradual incidence on GFv with a highest GFv for transparent interactions, a lowest GFv for non transparent interactions and a intermediate level for non transparent interactions augmented with visual augmentations.

5.2.2.7 Subjective reports

The results of the questionnaires will be addressed in the upcoming complete study.

5.2.3 Discussion

5.2.3.1 Transposition success

With this exploratory study, we confirm the relevancy of a transposition of the grip force paradigm to musical interactions. Our results from a small sample size show that grip force variations (GFv) appear to be an interesting marker for the study of spectator experience. In particular, we observed higher GFv for transparent interactions compared to less transparent

[2] The noisy parts are due to a larger difference of duration between the original sequences. The spiky envelope could be explained by the fact that longer sequences are shrink to fit into small spans. Future development may help smooth the normalised data.

interactions with autonomous processes which tends to indicate a link between Attributed Agency and the GFv. This primary interpretation fits well with the embodied theory of music and the parallel activation of the motor system during the perception of musical interactions. When perceived cues suffice for the observers to build a reliable internal model of the interaction, like in transparent conditions, the motor simulation is amplified and triggers larger GFv.

5.2.3.2 Research leads

5.2.3.2.1 Larger protocol The present study and the exploitable data it provides is a clear call for further research. In particular, the protocol should be now deployed with a larger sample size to gather cumulative data and allow for refined analyses. For instance, contrasts in GFv could be used to analyse the respective impact of interaction parameters (control interface,mapping, ...) or to complete our knowledge on the role of expertise in spectator experience. In particular, it would be insightful to assess the objective consequences of the stronger subjective experience regularly reported by experts in the studies presented in this thesis (See chapter 4).

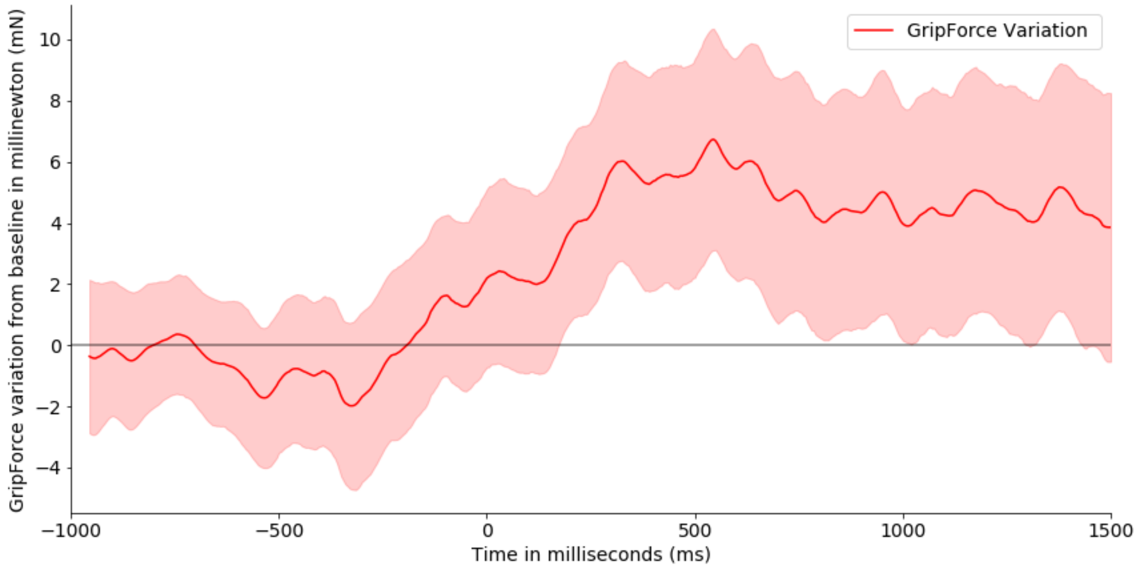
5.2.3.2.2 Ecological data Future technological improvements in sensors can open the way for more ecological protocols based on grip force measurements. For instance, miniaturised sensors could be placed around the body of a reusable glass equipped with a wireless transmitter in the base, like the one used in Canada during Hockey matches. Spectators of a concert could hold their glass during the concert and grip force data could be transmitted to a pipeline like the one we presented in this chapter and be used to adapt visual augmentations or trigger haptic augmentations.

5.2.4 Synthesis

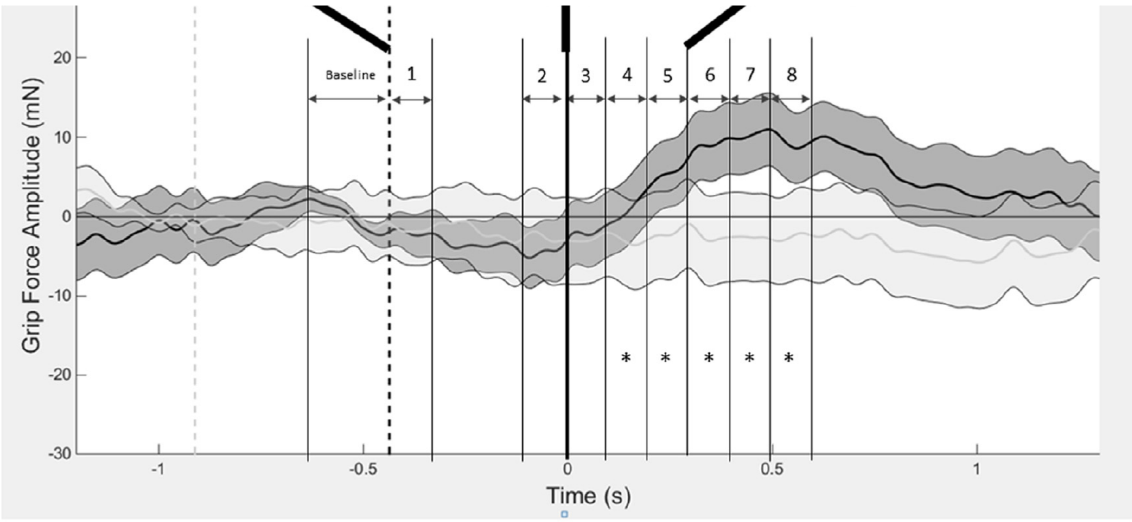
With these preliminary results, we confirmed the potential methodological benefits of the grip force paradigm in the study of the perception of musical interactions. In particular, with greater sample size, the grip force setup can offer methodological opportunities to objectively characterize aspects of the experience of spectators of performances with DMIs and opens the way to more ecological experiments.

5.3 Conclusion

In this chapter, we presented our contributions to improve the experience of spectators of digital music performances. In particular, we exposed a conceptual pipeline that supports the autonomous generation of optimal visual augmentations based on a continuous evaluation of the familiarity of the audience. We presented the correspondences, a conceptual object that aims at imitating the spectators internal model. We also introduced a transposition of the grip force paradigm to the study of musical interactions. This preliminary experiment opens the way for continuous monitoring of Attributed Agency and is a practical contribution to our conceptual pipeline.



(a) Taken globally, our data reproduce the results from Blampain et al. [23]. Here, the data are aligned on the beginning of the musician's movement, so t = 0ms corresponds to the moment the musician initiates their movement.



(b) From Blampain et al [23], the darker shape is the GFv of observers watching *High intensity action videos* like a strawberry crushed by a hand slap. Contrary to 5.5a, 0 is the moment when the hand touches the object.

Figure 5.5: Grip force variations in mN for spectators watching or mentally visualising interactions

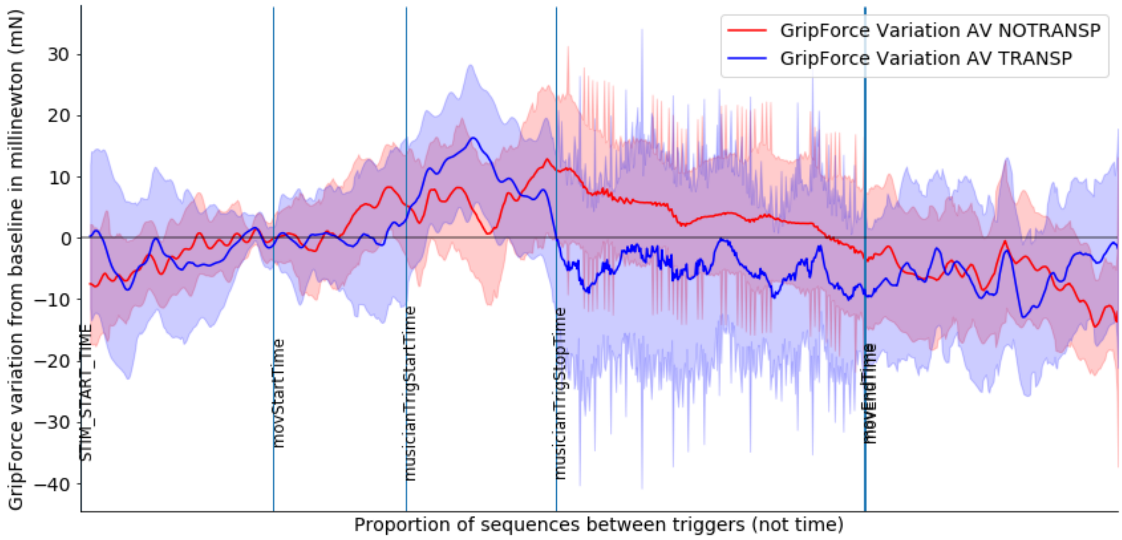


Figure 5.6: Time normalised GFv. To cope with unsynchronised interactions, the data on this graph are normalised along temporal sequences delimited by events common to all interactions.

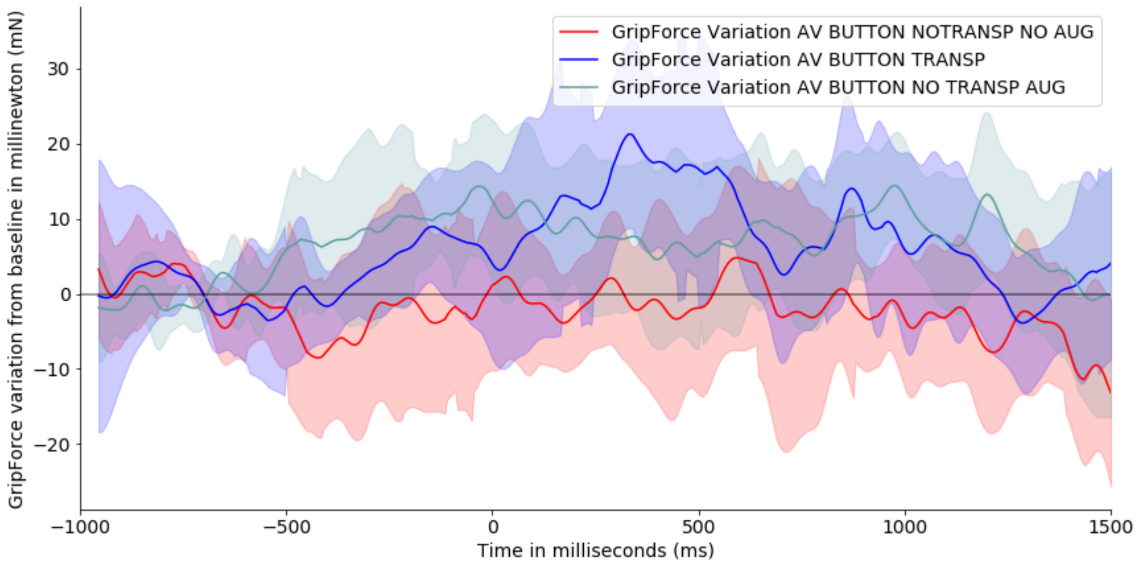


Figure 5.7: GFv for interactions with a NanoKorg MIDI controller. These preliminary data seem to indicate a gradual incidence of transparency with a high level for transparent interactions, a low level for non transparent interactions and an intermediate level for non transparent interactions augmented with visual augmentations.

Within a few years a simple and inexpensive device, readily carried about, will enable one to receive on land or sea the principal news, to hear a speech, a lecture, a song or play of a musical instrument, conveyed from any other region of the globe.

Nikola Tesla (1905)

6 Conclusion & perspectives

6.1 Conclusion

In this work, we presented our research on the experience of spectators of musical performances with digital instruments. The multidisciplinary nature of this problematic led us to adopt methodological and development resources from both computer and cognitive sciences. Specifically, the controlled paradigm often used in cognitive psychology gave us a strong experimental foundation to approach the sensitive study of human experience. Tools and methods from the field of Computer Science and more specifically HCI allowed us to design finely paired stimuli with subtle contrasts to extract details from the global experience of spectators. From our point of view, we could hardly have obtained most of the insights presented in this document without the help of this compound approach. Within this framework, we aimed at contributing to the definition, the measurement and the improvement of spectator experience.

6.1.1 Capturing the experience

Trying to objectivise a subjective experience is not an easy catch. To capture the elusive phenomenon of spectator experience, we proposed two complementary decompositions of a global variable into components easier to handle in experimental protocols. First, we presented and assessed a model of the familiarity spectators have of an instrument and confirmed the major role of *Attributed Agency*. In a second initiative, we decomposed the global experience in three components, the subjective experience, the objective comprehension and the subjective comprehension. We showed that the controlled study of these three different aspects of spectator experience permits to gather complementary data to the one usually reported from field studies.

6.1.1.0.1 Subjective comprehension, internal model and attributed agency In particular, we found that the items targetting the subjective comprehension deliver a good evaluation of the confidence observers have in their internal model of the ongoing interactions. Interestingly, while not correlated with better objective abilities in the understanding of interactions, this confidence is correlated to the perception of a greater contribution of the musician (attributed agency) in musical interactions involving autonomous processes. In other words, spectators do not need to objectively understand the musical interactions to perceive the contribution of the musician.

We confirmed this interpretation by addressing the role of expertise in a controlled study and found that, compared to novices, experts do perceive a higher attributed agency, report higher subjective comprehension evaluations but do not score better than novices in objective comprehension tasks. These concordant results strongly suggest that Attributed Agency in a music performance context is a feeling mostly influenced by subjective considerations like the trust in one's internal model of interactions rather than an objective ability to perceive discreet or fast details of the interactions. These findings are in line with the latest models of agency [193] as they integrate both pre (expertise and personal belief in one's internal model of the interactions) and postdictive (consistency and exclusivity evaluation) components.

Finally, to extend the range of potential factors of the experience, we proposed in section 4.2.3.3 to add a new evaluation criteria to the Bellotti-Fyans challenges: the Association. This property aims at evaluating digital systems with respect to their capacity to give the audience the perception of the contribution of agents (musicians) and the one of autonomous processes.

6.1.2 Improving the experience

Among the Spectator Experience Augmentation Techniques (SEATs) reported in the literature, we chose to confront and detail the efficiency of visual augmentations. In particular, we showed that when DMI performances are presented with visual augmentations, spectators report a greater subjective experience, and a higher subjective comprehension. Visual augmentations also lead participants to think that the musician contributes more than autonomous processes to the global performance, even when objectively autonomous processes contribute more than the musician. We also presented evidence of the deleterious impact of disruptive visual augmentations on the perceived contribution of the musician. Thus, in the context of electronic music performances where doubts about the musician contribution may appear, visual augmentations prove their efficiency at restoring the trust of spectators in the musicians and to reverse the doubt in their favour.

To complete our approach, we documented the impact of the levels of detail (LODs) of visual augmentations as well as the role of the expertise of spectators. Our analyses confirmed the positive role of visual augmentations on the subjective comprehension and showed that the efficiency of LODs was modulated by expertise. When novices can enjoy a relatively low LOD, experts favour more exclusively richer visual representations. Thanks to complementary interviews, we gathered more detailed feedback that make us advocate for a customisable use of visual augmentations to allow spectators to freely modify, during the performance, the level of augmentations they need, if any.

6.1.2.0.1 A pipeline for optimal augmentations In extension to our works on SEATs, we presented a conceptual pipeline that supports the selection and the diffusion of optimal augmentations. The modular architecture of the pipeline allows the integration of multiple factors to select the most adapted augmentations: a module is dedicated to the extraction of information coming from the musician and their instrument, another extracts from the audience electrophysiological data as well as subjective feedbacks while a third module integrates these information into integrative objects that we call correspondences. A correspondence is a conceptual object holding the parameters of an interaction and the consequent audience responses. It is designed to imitate the internal models of spectators and reveal potential perceptive issues in the interactions. Its evaluation allows the pipeline to select the optimal augmentation for a target audience. The implementation of this pipeline requires several advances, especially in terms of measurement device, electrophysiological signals interpretation, computing power and machine learning. In particular, increasing support from both cognitive neurosciences [123, 127] and artificial intelligence [84] communities invites us to consider the Bayesian modelling as a relevant implementation approach for correspondences.

6.1.3 Facilitating future research

Finally, we aimed at facilitating future research by proposing a taxonomy of the SEATs, and by following Open Science recommendations.

From the SEATs reported in the literature, we extracted recurring features and structured them in a taxonomy. Its eleven dimensions and formal descriptors aim at clarifying the reviews for future research and can also be used as a tool to refine the analysis of the efficiency of SEATs. Indeed, identified descriptors can be isolated and individually manipulated to study their respective impact on spectator experience. Also, to permit extra contributions to our first database of SEATs, we built a website and joined discussions in the NIME community.

To complete the methodological contribution for future research, we think that the reproductibility of the experiments and data sharing are worth a particular interest. In this document, we presented the pragmatic benefits of the Open Science recommendations and the use of Bayesian inferences in our statistical analyses. We believe that the generalisation of these mutualisation practices, alongside with initiatives for a greater transparency of the data, are essential to produce better studies and stronger interpretations[1].

[1] At the time those words were written, the scientific world is shaken by a trust crisis in the context of the covid19 pandemic and the Lancet just retracted a paper due to doubts on the "veracity of the primary data sources" by the co-authors themselves.

6.2 Perspectives

6.2.1 Evaluating the experience in ecological contexts

The research conducted in this work was realised in a lab in order to control the different conditions, minimize the random variables related to the environment and gather objectifiable data. Yet, these conditions are far from the natural environment of music performances that are the concert halls. We do integrate in our approach that a not negligible part of the concert experience is missing in too constrained protocols. One solution to keep the best of the two worlds, lab and field, could be the continuous measurement of objective signals like electrophysiological markers directly in the concert hall.

The preliminary results we obtained with the analysis of the grip-force signal (see chapter 5.1.4 for details) are promising in the sense that they seem to reproduce, with musical interaction stimuli, the ones obtained in previous psychophysics studies. More interestingly, we have preliminary data that show a tendency for intermediate grip force variations for non transparent interactions with visual augmentations between highest values for transparent interactions and lowest values for non transparent interactions without visual augmentations. From this basis, there are many interesting questions and the expertise is one of the most intriguing.

Specifically, we saw in chapter 4 that experts do not score higher than novices in objective tasks about interactions with Digital Musical Instruments (DMIs). This indicates that experts and novices both develop an imperfect inner simulation of the interactions. However, we saw that experts do report a better subjective comprehension than novices. What would be the consequences on grip force of similar imperfect representations but supported with a higher confidence ? If a difference between novices and experts could be observed in the grip force pattern, the interpretation of the influence of non physical considerations on premotor signals (at the origin of the grip force variations) could be interesting to confront.

6.2.2 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 on the subjective comprehension. Throughout the thesis, 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 phenomena 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 may have a greater similarity with the ones from the musicians. From our data, it seems that they do not. Expert observers did not score much higher than novices in objective tasks. Rather, they showed a greater confidence in their inner representations and that factor seems to partially explain the better experience that they reported. This point makes us think that the role of SEATs 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. We saw it in our studies with the preference of a rather balanced LOD compared to fully descriptive visual augmentations. In this context, the cognitive overload should also play a modulative role.

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 direct reactions. Such a sensitive role is not a purely technical role any more. It requires to integrate 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 have access to every processes running on stage and could trigger precompiled as well as live visual or haptic augmentations. From the evaluations of the efficiency of the SEATs on identified aspects of the interactions, the augmenter could act as an augmentations conductor and compose with the direct inputs from the musicians instrument and connect them to banks of augmentations to emphasize parts of the interactions. On the contrary the augmenter can 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 performs as a human mediation between digital systems and human agents to reveal the virtuosity of the musicians and the expressiveness of instruments.

6.2.3 Transposition to agency

The Attributed Agency is a cognitive concept that helped us apprehend a facet of the spectator experience. Epistemologically it is also a transposition of the concept of agency from which we also transposed the decomposition (priority, exclusivity and consistency) to initiate our first models of the spectator experience. We now propose to transpose back what we learned from the Attributed Agency to provide insights that may help the comprehension of agency issues, inter alia, for pathological public.

We saw in this work the correlation between high scores in subjective comprehension and the perception of a greater contribution of the musician in interactions where autonomous processes were also engaged. By transposing the perception of the control that someone else has over a device to the perception of one's own control over the environment (agency), we can expose potentially interesting parallels. Thus, with the transposition, the musician of the Attributed Agency model becomes one self (free will), the instrument becomes one's own body and the music the desired outcome of an action. Some of the Bellotti-Fyans challenges can then be adapted to characterise the spread of an agency disorder with questions like:

- Attention: I can feel when my body is responding to my will and when it is not.
- Action: I can feel when I'm controlling my body and when I am not.
- Alignment: I can feel when my body is properly functioning and when it is not.
- Accident: I can feel when I made a bad choice or when my body did something wrong.
- Association: I know the contribution of my will and the one of autonomous processes in my body.

From a refined diagnosis, the use of visual augmentations in wearable device like smart glasses could also be useful to people with agency disorder. How should we augment the environment to reveal the consequences of a behaviour ?

Beyond pathological contexts, do we not, as citizens, need augmentations to be more aware of the consequences of our actions at a local and global scale ? What would be our behaviour if we could better distinguish our individual contribution from the other's as our actions are often dissolved in a noisy effusion of apparently disconnected micro movements. What if art could help us in this lack of perception about the world and our relationship with it?

We believe that artists are part of the crucial augmentations that society needs to embrace essential human questions. By helping musicians to better share with spectators, by revealing to the audience their intense link with the art they produce, hopefully we help the artists give us a better perception of the connections between our behaviour and its consequences on the harmony of the world.

Appendices

A Profile questions

Below are the questions asked before each experiment to gather information about the participants expertise and cultural background. **Note: all questions were expressed in FRENCH.**

A.1 Exclusion controls

Before each experiment, we controlled that participants had no sight or hearing impairment.

A.1.1 Sight

"Do you need a visual correction ?" (y/n)
[if yes] "Do you have it right now ?" (y/n)
[if no] *Exclusion*

A.1.2 Hearing

- On a scale from 1 to 5, what is the quality of your hearing ? (5 is very good)
  (If you are equipped with a hearing aid, indicate 3)"
- *The participant must click a button on the screen and a low tone sound is played*
  "Choose the correct answer. The sound played is rather:" (Low tone/ High tone/ Neutral tone/ No sound)
  [If bad answer or no sound] *Exclusion*

A.2 Digital technology knowledge

- "Do you follow the news about digital technologies ?" (yes regularly/from time to time/never)
  [If yes regularly/from time to time] "What is the most recent innovation you have heard about ?"
- "Do you know theses devices ? Briefly describe their main functionality. Leave an empty answer if you do not know the device." (See figure below)

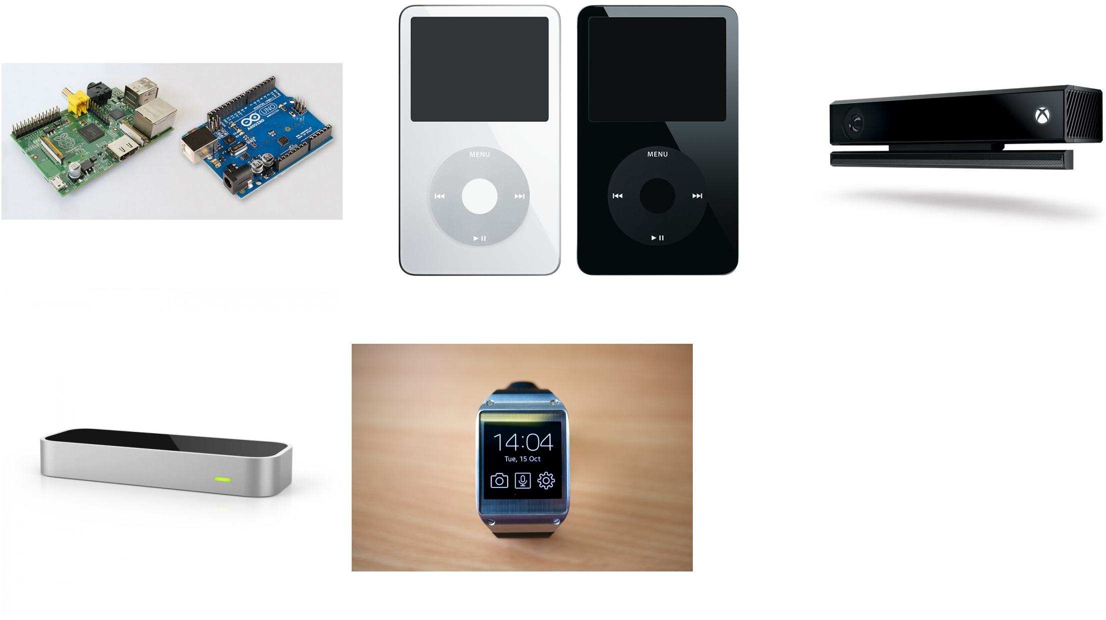


Figure 1: Examples of digital devices for the evaluation of digital technologies knowledge.

A.3 Music and digital performances expertise

- "Do you play music ?" (y/n)
  [If no] *End of music practice questions*
- "How many years of music practice do you have ?"
- "How many years of electronic music practice do you have ?"
- "Do you use graphical interfaces for playing music ?"
- "Do you use control surfaces for playing music ?"
- "In average, how many times a year do you attend a electronic music concert ?"

B Web study survey

B.1 Questions on Familiarity

FAMILIARITY
This musical control is clear, obvisous.
Overall, I find this interaction rather familiar, natural
I have difficulties understanding this video, it's not so clear.
The ongoing action in the video is not clear, not obvious.
CONSISTENCY
This is the right gesture to use for this kind of sonic interaction.
Do you find the sound consistent with the musician's gestures ?
The sound is consistent with the intention of the musician.
The sound is NOT consistent with the intention of the musician.
For the same sonic outcome, the musician could have done something different.
EXPERTISE
Do you know the instrument that the musician is playing ? *No | yes, I have already seen it | Yes, I have already played it*
Have you already seen musicians playing like this, that is with the same gestures but maybe with another instrument ?
Have you already seen the instrument that the musician is playing ?
Have you already played the current instrument (at least once) ?
INSTRUMENTALITY
Without considering the musician nor his gestures, do you find the sound consistent with the instrument ?
Considering the instrument only, do you find the sound consistent with it ?
Only by its physical appearance, I could have predicted the sound of the instrument.
EXCLUSIVITY
In this video, the sound is controlled by : *The musician | A computer | Both | Something else (precise)*
In this video, the musician does nothing, the computer is the only one in charge.
It is clear that in this video, the musician is the only one to impact the sound.
In addition to his instrument instrument, the musician is assisted by a computer to manage to produce the sound
ATTRIBUTED AGENCY
The musician controls the situation.
The musician looks like he knows what he is doing.
The musician does not seem to control anything.

Figure 2: Questions and the *Familiarity* dimensions they target. Affirmations were followed by Likert-style scales of the degree of agreement. See chapter 2

List of Figures

List of Tables

List of Terms

attributed agency The transposition of agency from the feeling of being author of one's own actions to the evaluation of the authorship of someone else's actions. Often in this document, Attributed Agency is the evaluation of the contribution of a musician to the performance compared to the contribution of autonomous processes.
The capitalized form (Attributed Agency), refers to the component term of spectator experience or *Familiarity*. I, 5, 6, 35, 36, 43, 45–50, 55, 57–61, 69, 72, 88, 101, 102, 104, 109, 110, 116, 117, 119, 120, 123, 128, 131

correspondence a conceptual object that stands as a digital multidimensional representation of a musical interaction. It can be viewed as an initiative to imitate the internal model of spectators.. 117, 121

Digital Musical Instrument A musical instrument that includes digital elements like electronic sensors, MIDI controllers or computers for example. 5, 48, 64, 72, 89, 110, 122, 133

disruptive visual augmentations Contrary to visual augmentations, disruptive visual augmentations are not correlated with the ongoing interactions.. 76, 79, 82, 84, 86, 87, 120

DMI performance A musical performance where the setup includes Digital Musical Instruments. 31, 33, 35, 40, 42, 44, 48, 49, 89, 120, 131

expertise In this document, the term *expertise* relates to the artistic and cultural background in electronic music concerts, DMI performances or DMI practice. VI, 4, 5, 25, 26, 28, 46, 48–50, 54, 59, 60, 66, 67, 69, 71, 72, 78, 83, 86, 88, 89, 94, 95, 97–99, 106, 107, 116, 120, 122

familiarity In this document, familiarity refers to the knowledge one has about an instrument. From a spectator point of view, familiarity is the capacity to perceive and understand the expressivness of an instrument and to easily associate gestures and audio outcomes. The capitalize form (Familiarity), refers to the component term of spectator experience. 49, 55–59, 61, 127, 131

grip force The pressure exerted when holding an object. In this document, grip force relates to the measurement of micro pressure variations (in Millinewtons, mN) thanks to a sensor that the participant holds between their fingers.. I, VI, 6, 46, 101, 108, 110–113, 116, 117, 122, 129, 131, 133

level of detail or LOD. The amount and the kind of information provided by a SEAT. In this document, *LOD* is most of the time related to the level of detail of visual augmentations. 133

objective comprehension The objective comprehension refers to the ability of the spectator to integrate factual elements of an interaction, like who from the musician or the machine is the author of a sound modulation. 72–75, 81, 83, 86, 89, 119, 123

spectator experience augmentation technique technical solutions to improve the experience of spectators. e.g.: visual augmentations, pre-concert explanations, mobile application 133

subjective comprehension Midway between subjective experience and objective comprehension, subjective comprehension relates to the inner feeling of being able to perceive an aspect of an interaction. It can be addressed through the Bellotti/Fyans challenges. I, 73–75, 80, 82, 83, 88, 98, 119, 120, 122, 123

subjective experience The subjective experience is a global impression, a rather elusive feeling related to how much we like, or evaluate something we experience, like a performance. I, 73–75, 79, 80, 87, 88, 96, 117, 119, 120

The International Conference on New Interfaces for Musical Expression The International Conference on New Interfaces for Musical Expression gathers researchers and musicians from all over the world to share their knowledge and late-breaking work on new musical interface design. The conference started out as a workshop at the Conference on Human Factors in Computing Systems (CHI) in 2001. Since then, an annual series of international conferences have been held around the world, hosted by research groups dedicated to interface design, human-computer interaction, and computer music. 18, 132

visual augmentations Visual augmentations are graphical representations updated in real time of the musical interactions and the inner mechanisms (controllers, mappings, sound processes) of DMIs. I, 5, 40, 43, 44, 46, 64, 65, 68, 69, 72–76, 79, 80, 82, 84–90, 93, 96, 97, 99, 102, 110, 111, 115, 117, 120, 123, 124

Acronyms

DMI Digital Musical Instrument. 2, 5, 16, 18–20, 22–24, 26–28, 32, 36, 37, 40, 41, 43, 48, 60, 64, 72, 77, 89, 90, 95, 99, 107, 110, 117, 122, 133, *Glossary:* Digital Musical Instrument

GFv grip force variations. VI, 110, 113–116, 129

HCI Human Computer Interaction. I, V, 4, 5, 17, 28, 29, 31, 45, 49, 88, 89, 119

LOD level of detail. 72, 86, 87, 89–94, 97–99, 120, 123, 129, 133, *Glossary:* level of detail

NIME The International Conference on New Interfaces for Musical Expression. 17–19, 23, 29, 32, *Glossary:* The International Conference on New Interfaces for Musical Expression

SEAT spectator experience augmentation technique. I, II, VI, 64, 69, 72–75, 79, 80, 82–87, 120, 121, 123, 133, *Glossary:* spectator experience augmentation technique

Bibliography

[1] J. Armitage. Revealing timelines: Live coding and its gestures. *Proceedings of ICLC*, 2016.

[2] P. Auslander. *Liveness: Performance in a mediatized culture*. Routledge, 2008.

[3] J. Barbosa, F. Calegario, V. Teichrieb, G. Ramalho, and P. McGlynn. Considering audience's view towards an evaluation methodology for digital musical instruments. In *NIME '12: Proceedings of the 2012 International Conference on New Interfaces for Musical Expression*, 2012.

[4] J. Barbosa, F. Calegario, V. Teichrieb, G. Ramalho, and G. Cabral. Illusio: A drawing-based digital music instrument. In *NIME '13: Proceedings of the 2013 International Conference on New Interfaces for Musical Expression*, 2013.

[5] M. Battier. Trends in gestural control of music. Ircam, 2000.

[6] M. A. Baytas, T. Göksun, and O. Özcan. The perception of live-sequenced electronic music via hearing and sight. In *NIME '16: Proceedings of the 2016 International Conference on New Interfaces for Musical Expression*, pages 194–199, 2016.

[7] B. B. Bederson and J. D. Hollan. Pad++: a zooming graphical interface for exploring alternate interface physics. In *Proceedings ACM UIST*. ACM, 1994.

[8] A. M. Belfi, B. Karlan, and D. Tranel. Music evokes vivid autobiographical memories. *Memory*, 24(7):979–989, 2016.

[9] V. Bellotti, M. Back, W. K. Edwards, R. E. Grinter, A. Henderson, and C. Lopes. Making sense of sensing systems: five questions for designers and researchers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM, 2002.

[10] S. Benford, C. Greenhalgh, A. Crabtree, M. Flintham, B. Walker, J. Marshall, B. Koleva, S. Rennick Egglestone, G. Giannachi, M. Adams, et al. Performance-led research in the wild. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(3):14, 2013.

[11] S. Benford, C. Greenhalgh, A. Hazzard, A. Chamberlain, M. Kallionpää, D. M. Weigl, K. R. Page, and M. Lin. Designing the audience journey through repeated experiences. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 568. ACM, 2018.

[12] T. Berners-Lee, J. Hendler, O. Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.

[13] F. Berthaut and M. Desainte-Catherine. Interaction musicale numérique. recréer les signaux du contrôle et de la perception. 2015.

[14] F. Berthaut, M. Marshall, S. Subramanian, and M. Hachet. Rouages: Revealing the mechanisms of digital musical instruments to the audience. In *NIME '13: Proceedings of the 2013 International Conference on New Interfaces for Musical Expression*, 2013.

[15] F. Berthaut, D. Coyle, J. W. Moore, and H. Limerick. Liveness through the lens of agency and causality. In *The 15th International Conference on New Interfaces for Musical Expression, 31 May-3 June, 2015, Louisiana State University, USA*, 2015.

[16] S. Bin, A. MCPHERSON, N. BRYAN-KINNS, et al. Skip the pre-concert demo: How technical familiarity and musical style affect audience response. In *NIME '16: Proceedings of the 2016 International Conference on New Interfaces for Musical Expression*, 2016.

[17] S. A. Bin. The show must go wrong: Towards an understanding of audience perception of error in digital musical instrument performance. 2018.

[18] S. A. Bin, N. Bryan-Kinns, and A. P. McPherson. Skip the pre-concert demo: How technical familiarity and musical style affect audience response. In *NIME '16: Proceedings of the 2016 International Conference on New Interfaces for Musical Expression*, 2016.

[19] S. A. Bin, F. Morreale, N. Bryan-Kinns, and A. P. McPherson. In-the-moment and beyond: Combining post-hoc and real-time data for the study of audience perception of electronic music performance. In *IFIP Conference on Human-Computer Interaction*, pages 263–281. Springer, 2017.

[20] D. Birnbaum, R. Fiebrink, J. Malloch, and M. M. Wanderley. Towards a dimension space for musical devices. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 192–195, 2005.

[21] S.-J. Blakemore and C. Frith. Self-awareness and action. *Current opinion in neurobiology*, 13(2):219–224, 2003.

[22] S.-J. Blakemore, D. M. Wolpert, and C. D. Frith. Abnormalities in the awareness of action. *Trends in cognitive sciences*, 6(6):237–242, 2002.

[23] J. Blampain, L. Ott, and Y. Delevoye-Turrell. Seeing action simulation as it unfolds: The implicit effects of action scenes on muscle contraction evidenced through the use of a grip-force sensor. *Neuropsychologia*, 114:231–242, 2018.

[24] A. J. Blood and R. J. Zatorre. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the national academy of sciences*, 98(20):11818–11823, 2001.

[25] J. Bowers and S. O. Hellström. Simple interfaces to complex sound in improvised music. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 125–126, 2000.

[26] O. Bown, R. Bell, and A. Parkinson. Examining the perception of liveness and activity in laptop music: Listeners' inference about what the performer is doing from the audio alone. 2014.

[27] A. R. Brown, T. Gifford, and B. Voltz. Factors affecting audience perceptions of agency in human computer musical partnerships. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, pages 296–299, 2013.

[28] A. S. Brown and J. L. Novak. *Assessing the intrinsic impacts of a live performance*. WolfBrown San Francisco, CA, 2007.

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

[30] M. Budd. *Music and the emotions: The philosophical theories*. Routledge, 2002.

[31] C. Cadoz. Instrumental gesture and musical composition. 1988.

[32] C. Cadoz. Supra-instrumental interactions and gestures. *Journal of New Music Research*, 38(3):215–230, 2009.

[33] C. Cadoz and M. M. Wanderley. Gesture-music, 2000.

[34] B. Calvo-Merino, D. E. Glaser, J. Grèzes, R. E. Passingham, and P. Haggard. Action observation and acquired motor skills: an fmri study with expert dancers. *Cerebral cortex*, 15(8):1243–1249, 2005.

[35] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, and G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. In *International gesture workshop*, pages 20–39. Springer, 2003.

[36] O. Capra, F. Berthaut, and L. Grisoni. Toward augmented familiarity of the audience with digital musical instruments. In *International Symposium on Computer Music Multidisciplinary Research*, pages 558–573. Springer, 2017.

[37] O. Capra, F. Berthaut, and L. Grisoni. Have a seat on stage: Restoring trust with spectator experience augmentation techniques. In *Designing Interactive Systems (DIS)*, 2020.

[38] O. Capra, F. Berthaut, and L. Grisoni. All you need is lod : Levels of detail in visual augmentations for the audience. In *NIME '20: Proceedings of the 2020 International Conference on New Interfaces for Musical Expression*, 2020.

[39] O. Capra, F. Berthaut, and L. Grisoni. A taxonomy of spectator experience augmentation techniques. In *NIME '20: Proceedings of the 2020 International Conference on New Interfaces for Musical Expression*, 2020.

[40] J. Carifio and R. Perla. Resolving the 50-year debate around using and misusing likert scales. *Medical education*, 42(12):1150–1152, 2008.

[41] A. Chamberlain, A. Crabtree, T. Rodden, M. Jones, and Y. Rogers. Research in the wild: understanding'in the wild'approaches to design and development. In *Proceedings of the Designing Interactive Systems Conference*, pages 795–796, 2012.

[42] G. Chanel, M. Bétrancourt, T. Pun, D. Cereghetti, and G. Molinari. Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 116–122. IEEE, 2013.

[43] C. Chapados and D. J. Levitin. Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition*, 108(3):639–651, 2008.

[44] I. Choi. A component model of gestural primitive throughput. In *NIME '03: Proceedings of the 2003 International Conference on New Interfaces for Musical Expression*, pages 201–204, 2003.

[45] A. Clark and D. Chalmers. The extended mind. *analysis*, 58(1):7–19, 1998.

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

[47] N. J. Conard, M. Malina, and S. C. Münzel. New flutes document the earliest musical tradition in southwestern germany. *Nature*, 460(7256):737–740, 2009.

[48] G. Consonni, D. Fouskakis, B. Liseo, I. Ntzoufras, et al. Prior distributions for objective bayesian analysis. *Bayesian Analysis*, 13(2):627–679, 2018.

[49] P. Cook. 2001: Principles for designing computer music controllers. In *NIME '17: Proceedings of the 2017 International Conference on New Interfaces for Musical Expression*, pages 1–13. Springer, 2017.

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

[51] N. N. Correia, D. Castro, and A. Tanaka. 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*, page 29. ACM, 2017.

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

[53] J. Croft. Theses on liveness. *Organised Sound*, 12(1):59, 2007.

[54] M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*. Harper Perennial Modern Classics. HarperCollins e-books, 2009. ISBN 9780061876721. URL `https://books.google.fr/books?id=epmhVuaaoK0C`.

[55] M. Csikszentmihalyi and R. E. Robinson. *The art of seeing: An interpretation of the aesthetic encounter*. Getty Publications, 1990.

[56] L. Dahl, F. Berthaut, A. Nau, and P. Plenacoste. bf-pd: Enabling mediated communication and cooperation in improvised digital orchestras. In *CMMR*. Springer, 2017.

[57] S. Dahl and A. Friberg. Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24(5):433–454, 2007.

[58] N. David. New frontiers in the neuroscience of the sense of agency. *Frontiers in human neuroscience*, 6:161, 2012.

[59] J. W. Davidson. Visual perception of performance manner in the movements of solo musicians. *Psychology of music*, 21(2):103–113, 1993.

[60] F. Delalande. La gestique de gould: éléments pour une sémiologie du geste musical. *Glenn Gould Pluriel*, pages 85–111, 1988.

[61] M. Desainte-Catherine. Les sciences de la musique: de nouveaux défis dans une société en mutation. URL `https://www.youtube.com/watch?v=ykmAUH3FW8M`.

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

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

[64] T. Eerola and J. K. Vuoskoski. A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, 30(3): 307–340, 2012.

[65] H. Egermann, M. E. Sutherland, O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller. Does music listening in a social context alter experience? a physiological and psychological perspective on emotion. *Musicae Scientiae*, 15(3):307–323, 2011.

[66] H. Egermann, M. T. Pearce, G. A. Wiggins, and S. McAdams. Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, & Behavioral Neuroscience*, 13(3):533–553, 2013.

[67] G. D. Ellis, J. E. Voelkl, and C. Morris. Measurement and analysis issues with explanation of variance in daily experience using the flow model. *Journal of leisure research*, 26(4): 337–356, 1994.

[68] G. Emerson and H. Egermann. Mapping, causality and the perception of instrumentality: Theoretical and empirical approaches to the audience's experience of digital musical instruments. In *Musical instruments in the 21st Century*, pages 363–370. Springer, 2017.

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

[70] R. Fiebrink and L. Sonami. Reflections on eight years of instrument creation with machine learning. 2020.

[71] L. Foglia and R. Wilson. Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 05 2013. doi: 10.1002/wcs.1226.

[72] A. Freed. Open sound control: A new protocol for communicating with sound synthesizers. In *International Computer Music Conference (ICMC)*, 1997.

[73] C. D. Frith, S.-J. Blakemore, and D. M. Wolpert. Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1404):1771–1788, 2000.

[74] A. C. Fyans and M. Gurevich. Perceptions of skill in performances with acoustic and electronic instruments. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 495–498, Oslo, Norway, 2011. URL `http://www.nime.org/proceedings/2011/nime2011_495.pdf`.

[75] A. C. Fyans, M. Gurevich, and P. Stapleton. Where did it all go wrong ? a model of error from the spectator's perspective. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 171–172, Pittsburgh, PA, United States, 2009. URL `http://www.nime.org/proceedings/2009/nime2009_171.pdf`.

[76] A. C. Fyans, M. Gurevich, and P. Stapleton. Where did it all go wrong? a model of error from the spectator's perspective. In *NIME '09: Proceedings of the 2009 International Conference on New Interfaces for Musical Expression*, pages 171–172, 2009.

[77] A. C. Fyans, M. Gurevich, and P. Stapleton. Spectator understanding of error in performance. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3955–3960. 2009.

[78] A. C. Fyans, M. Gurevich, and P. Stapleton. Examining the spectator experience. In *NIME '10: Proceedings of the 2010 International Conference on New Interfaces for Musical Expression*, volume 10, pages 1–4, 2010.

[79] A. Gabrielsson. Emotions in strong experiences with music. 2001.

[80] S. Gallagher. Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, 37(1):8–19, 2004.

[81] S. Gallagher. *How the body shapes the mind*. Clarendon Press, 2006.

[82] V. Gallese and A. Goldman. Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12):493–501, 1998.

[83] V. Gallese and G. Lakoff. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, 22(3-4):455–479, 2005.

[84] Z. Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521 (7553):452–459, 2015.

[85] R. W. Gibbs Jr. *Embodiment and cognitive science*. Cambridge University Press, 2005.

[86] S. Gibet. *Codage, représentation et traitement du geste instrumental. Application à la synthèse de sons musicaux par simulation de mécanismes instrumentaux*. PhD thesis, 1987.

[87] J. Gibson. The ecological approach to visual perception. boston, ma, us, 1979.

[88] J. J. Gibson. The theory of affordances. *Hilldale, USA*, 1(2), 1977.

[89] R. I. Godøy. Motor-mimetic music cognition. *Leonardo*, 36(4):317–319, 2003.

[90] R. I. Godøy. Gestural affordances of musical sound. *Musical gestures: Sound, movement, and meaning*, pages 103–125, 2010.

[91] R. I. Godøy, A. R. Jensenius, A. Voldsund, K. H. Glette, M. E. Høvin, K. Nymoen, S. A. v. D. Skogstad, and J. Tørresen. Classifying music-related actions. In *Proceedings of 12th International Conference on Music Perception and Cognition*, 2012.

[92] S. T. Grafton. Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences*, 1156(1):97–117, 2009.

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

[94] P. Haggard. Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4): 196, 2017.

[95] P. Haggard and V. Chambon. Sense of agency. *Current Biology*, 22(10):R390–R392, 2012.

[96] R. S. Hatten. *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Indiana University Press, 2004.

[97] A. Hazzard, C. Greenhalgh, M. Kallionpaa, S. Benford, A. Veinberg, Z. Kanga, and A. McPherson. Failing with style: designing for aesthetic failure in interactive performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[98] N. Hirose. An ecological approach to embodiment and cognition. *Cognitive Systems Research*, 3(3):289–299, 2002.

[99] A. Hunt and R. Kirk. Mapping strategies for musical performance. *Trends in gestural control of music*, 21(2000):231–258, 2000.

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

[101] D. B. Huron. *Sweet anticipation: Music and the psychology of expectation*. MIT press, 2006.

[102] J. P. Ioannidis. Why most published research findings are false. *PLoS medicine*, 2(8): e124, 2005.

[103] R. J. Jacob, L. E. Sibert, D. C. McFarlane, and M. P. Mullen Jr. Integrality and separability of input devices. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(1):3–26, 1994.

[104] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–210, 2008.

[105] P. Janata. The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19(11):2579–2594, 2009.

[106] P. Janata, S. T. Tomic, and S. K. Rakowski. Characterisation of music-evoked autobiographical memories. *Memory*, 15(8):845–860, 2007.

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

[108] M. Jeannerod. *Motor cognition: What actions tell the self*. Number 42. Oxford University Press, 2006.

[109] M. Jeannerod. The sense of agency and its disturbances in schizophrenia: a reappraisal. *Experimental Brain Research*, 192(3):527, 2009.

[110] A. R. Jensenius and M. J. Lyons. *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*, volume 3. Springer, 2017.

[111] A. R. Jensenius and M. M. Wanderley. Musical gestures: Concepts and methods in research. In *Musical Gestures*, pages 24–47. Routledge, 2010.

[112] A. R. Jensenius, M. M. Wanderley, R. I. Godøy, and M. Leman. Musical gestures. *Musical gestures: Sound, movement, and meaning*, 12, 2009.

[113] S. Jordà. 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*, pages 6–9, 2003.

[114] P. N. Juslin. Communicating emotion in music performance: A review and a theoretical framework. 2001.

[115] P. N. Juslin and J. A. Sloboda. *Music and emotion: Theory and research.* Oxford University Press, 2001.

[116] P. N. Juslin and D. Vastfjall. Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, 31(5):559, 2008.

[117] C. I. Karageorghis and D.-L. Priest. Music in the exercise domain: a review and synthesis (part i). *International review of sport and exercise psychology*, 5(1):44–66, 2012.

[118] C. I. Karageorghis and P. C. Terry. The psychophysical effects of music in sport and exercise: A review. *Journal of Sport Behavior*, 20(1):54, 1997.

[119] C. I. Karageorghis, P. C. Terry, and A. M. Lane. Development and initial validation of an instrument to assess the motivational qualities of music in exercise and sport: The brunel music rating inventory. *Journal of sports sciences*, 17(9):713–724, 1999.

[120] M. Kay, G. L. Nelson, and E. B. Hekler. 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*, pages 4521–4532. ACM, 2016.

[121] P. Kennedy, D. B. Miele, and J. Metcalfe. The cognitive antecedents and motivational consequences of the feeling of being in the zone. *Consciousness and cognition*, 30:48–61, 2014.

[122] J. Kim and E. André. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence*, 30(12):2067–2083, 2008.

[123] D. C. Knill and A. Pouget. The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12):712–719, 2004.

[124] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.

[125] C. L. Krumhansl. Music: A link between cognition and emotion. *Current directions in psychological science*, 11(2):45–50, 2002.

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

[127] T. L Griffiths, C. Kemp, and J. B Tenenbaum. Bayesian models of cognition. 2008.

[128] C.-H. Lai and T. Bovermann. Audience experience in sound performance. In *NIME '13: Proceedings of the 2013 International Conference on New Interfaces for Musical Expression*, pages 170–173, 2013.

[129] C. Latulipe, E. A. Carroll, and D. Lottridge. Love, hate, arousal and engagement: exploring audience responses to performing arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1845–1854. ACM, 2011.

[130] M. Leman. Musical gestures and embodied cognition. In *Journées d'informatique musicale (JIM-2012)*, pages 5–7. Université de Mons, 2012.

[131] M. Leman and P.-J. Maes. The role of embodiment in the perception of music. *Empirical Musicology Review*, 9(3-4):236–246, 2014. ISSN 1559-5749. doi: 10.18061/emr.v9i3-4.4498.

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

[133] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger. Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481): 410–423, 2008.

[134] M. Liggins II, D. Hall, and J. Llinas. *Handbook of multisensor data fusion: theory and practice*. CRC press, 2017.

[135] C. Liu, H. S. Alavi, E. Costanza, S. Zhai, W. E. Mackay, and W. Moncur. Rigor, relevance and impact: The tensions and trade-offs between research in the lab and in the wild. In *CHI Extended Abstracts*, 2019.

[136] E. F. Loftus and J. C. Palmer. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, 13(5):585–589, 1974.

[137] L. Lordier, S. Loukas, F. Grouiller, A. Vollenweider, L. Vasung, D.-E. Meskaldij, F. Lejeune, M. P. Pittet, C. Borradori-Tolsa, F. Lazeyras, et al. Music processing in preterm and full-term newborns: a psychophysiological interaction (ppi) approach in neonatal fmri. *Neuroimage*, 185:857–864, 2019.

[138] R. Lyon and S. Shamma. Auditory representations of timbre and pitch. In *Auditory computation*, pages 221–270. Springer, 1996.

[139] C. K. Madsen and W. E. Fredrickson. The experience of musical tension: A replication of nielsen's research using the continuous response digital interface. *Journal of Music Therapy*, 30(1):46–63, 1993.

[140] J. Malloch and M. M. Wanderley. The t-stick: From musical interface to musical instrument. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 66–70, 2007.

[141] J. R. Matyja and A. Schiavio. Enactive music cognition: background and research themes. *Constructivist foundations*, 8(3), 2013.

[142] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588): 746–748, 1976.

[143] C. B. Medeiros and M. M. Wanderley. A comprehensive review of sensors and instrumentation methods in devices for musical expression. *Sensors*, 14(8):13556–13591, 2014.

[144] V. Menon and D. J. Levitin. The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage*, 28(1):175–184, 2005.

[145] M. Merleau-Ponty. *Phenomenology of perception*. Routledge, 1982.

[146] L. B. Meyer. *Emotion and meaning in music*. University of chicago Press, 2008.

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

[148] I. Molnar-Szakacs and K. Overy. Music and mirror neurons: from motion to'e'motion. *Social cognitive and affective neuroscience*, 1(3):235–241, 2006.

[149] A. Momeni and D. Wessel. Characterizing and controlling musical material intuitively with geometric models. In *Proceedings of the 2003 conference on New interfaces for musical expression*, pages 54–62, 2003.

[150] F. R. Moore. The dysfunctions of midi. *Computer music journal*, 12(1):19–28, 1988.

[151] R. Morey and J. Rouder. Bayesfactor (version 0.9. 10-2). *Computer software*, 2015.

[152] T. Murray-Browne, D. Mainstone, N. Bryan-Kinns, and M. D. Plumbley. The medium is the message: Composing instruments and performing mappings. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 56–59, 2011.

[153] J. Nielsen. Risks of quantitative studies. *Jakob Nielsen's Alertbox*, 2004.

[154] L. Nijs, M. Lesaffre, and M. Leman. The musical instrument as a natural extension of the musician. In *the 5th Conference of Interdisciplinary Musicology*, pages 132–133. LAM-Institut jean Le Rond d'Alembert, 2009.

[155] L. Nummenmaa, E. Glerean, R. Hari, and J. K. Hietanen. Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, 111(2):646–651, 2014.

[156] K. Overy and I. Molnar-Szakacs. Being together in time: Musical experience and the mirror neuron system. *Music Perception*, 26(5):489–504, 2009.

[157] D. Páez, B. Rimé, N. Basabe, A. Wlodarczyc, and L. Zumeta. Collective emotional gatherings. five studies of identity fusion, shared beliefs and social integration. *Manuscript in preparation*, 2012.

[158] A. D. Patel. Language, music, syntax and the brain. *Nature neuroscience*, 6(7):674–681, 2003.

[159] J. Paulus, M. Müller, and A. Klapuri. State of the art report: Audio-based music structure analysis. In *Proceedings of ISMIR*, 2010.

[160] O. Perrotin and C. d'Alessandro. Visualizing gestures in the control of a digital musical instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 605–608, London, United Kingdom, June 2014. Goldsmiths, University of London. URL `http://www.nime.org/proceedings/2014/nime2014_406.pdf`.

[161] O. Perrotin and C. d'Alessandro. Visualizing gestures in the control of a digital musical instrument. 2014.

[162] S. S. Rautaray and A. Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.

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

[164] S. Reeves, S. Benford, C. O'Malley, and M. Fraser. Designing the spectator experience. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 741–750. ACM, 2005.

[165] S. Reeves, S. Sherwood, and B. Brown. Designing for crowds. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 393–402, 2010.

[166] M. Reybrouck. Biological roots of musical epistemology: functional cycles, umwelt, and enactive listening. 2001.

[167] M. Reybrouck. Music cognition and the bodily approach: Musical instruments as tools for musical semantics. *Contemporary Music Review*, 25(1-2):59–68, 2006.

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

[169] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2013.

[170] G. Rizzolatti and L. Craighero. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27: 169–192, 2004.

[171] A. Roudaut, A. Karnik, M. Löchtefeld, and S. Subramanian. Morphees: toward high" shape resolution" in self-actuated flexible mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 593–602, 2013.

[172] J. N. Rouder, R. D. Morey, P. L. Speckman, and J. M. Province. Default bayes factors for anova designs. *Journal of Mathematical Psychology*, 56(5):356–374, 2012.

[173] J. N. Rouder, C. R. Engelhardt, S. McCabe, and R. D. Morey. Model comparison in anova. *Psychonomic bulletin & review*, 23(6):1779–1786, 2016.

[174] J. N. Rouder, R. D. Morey, J. Verhagen, A. R. Swagman, and E.-J. Wagenmakers. Bayesian analysis of factorial designs. *Psychological Methods*, 22(2):304, 2017.

[175] J. B. Rovan, M. M. Wanderley, S. Dubnov, and P. Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Kansei, The Technology of Emotion. Proceedings of the AIMI International Workshop*, pages 68–73. Citeseer, 1997.

[176] S. Saarikallio and J. Erkkilä. The role of music in adolescents' mood regulation. *Psychology of music*, 35(1):88–109, 2007.

[177] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh. Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, 19 (2):86–91, 2015.

[178] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology*, 44(2):293–304, 2007.

[179] A. Schiavio, D. Menin, and J. Matyja. Music in the flesh: Embodied simulation in musical understanding. *Psychomusicology: Music, Mind, and Brain*, 24(4):340, 2014.

[180] A. Schiavio, D. van der Schyff, S. Kruse-Weber, and R. Timmers. When the sound becomes the goal. 4e cognition and teleomusicality in early infancy. *Frontiers in Psychology*, 8:1585, 2017.

[181] W. A. Schloss. Using contemporary technology in live performance: The dilemma of the performer. *Journal of New Music Research*, 32(3):239–242, 2003.

[182] H. Sequeira, P. Hot, L. Silvert, and S. Delplanque. Electrical autonomic correlates of emotion. *International journal of psychophysiology*, 71(1):50–56, 2009.

[183] L. Shapiro. *Embodied cognition*. Routledge, 2019.

[184] S. M. Shell. *The embodiment of reason: Kant on spirit, generation, and community*. University of Chicago Press, 1996.

[185] N. Shortway. "this one goes to eleven...": A methodological study of the recording and evaluation of emotional response to music. 2007.

[186] J. A. Sloboda. Music structure and emotional response: Some empirical findings. *Psychology of music*, 19(2):110–120, 1991.

[187] J. A. Sloboda. Music in everyday life: The role of emotions. 2010.

[188] C. G. Snoek and M. Worring. Multimodal video indexing: A review of the state-of-the-art. *Multimedia tools and applications*, 25(1):5–35, 2005.

[189] G. L. Stephens and G. Graham. *When self-consciousness breaks: Alien voices and inserted thoughts.* The MIT press, 2000.

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

[191] G. M. Sullivan and A. R. Artino Jr. Analyzing and interpreting data from likert-type scales. *Journal of graduate medical education*, 5(4):541–542, 2013.

[192] M. Swan. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, 1(2):85–99, 2013.

[193] M. Synofzik, G. Vosgerau, and M. Voss. The experience of agency: an interplay between prediction and postdiction. *Frontiers in psychology*, 4:127, 2013.

[194] A. Tanaka. Interaction, experience and the future of music. In *Consuming music together*, pages 267–288. Springer, 2006.

[195] L. Turchet and M. Barthet. Haptification of performer's control gestures in live electronic music performance. In *Proceedings of Audio Mostly*, pages 244–247. ACM, 2019.

[196] L. Turchet, T. West, and M. M. Wanderley. Smart mandolin and musical haptic gilet: effects of vibro-tactile stimuli during live music performance. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, pages 168–175, 2019.

[197] A. Van Goethem and J. Sloboda. The functions of music for affect regulation. *Musicae scientiae*, 15(2):208–228, 2011.

[198] F. J. Varela, E. Thompson, and E. Rosch. The embodied mind: Cognitive science and human experience. 1991.

[199] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, and D. J. Levitin. Cross-modal interactions in the perception of musical performance. *Cognition*, 101(1):80–113, 2006.

[200] B. W. Vines, C. L. Krumhansl, M. M. Wanderley, I. M. Dalca, and D. J. Levitin. Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2):157–170, 2011.

[201] F. Visi, R. Schramm, and E. Miranda. Gesture in performance with traditional musical instruments and electronics: Use of embodied music cognition and multimodal motion capture to design gestural mapping strategies. In *Proceedings of the 2014 International Workshop on Movement and Computing*, pages 100–105, 2014.

[202] J. K. Vuoskoski and T. Eerola. Can sad music really make you sad? indirect measures of affective states induced by music and autobiographical memories. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3):204, 2012.

[203] E.-J. Wagenmakers, M. Marsman, T. Jamil, A. Ly, J. Verhagen, J. Love, R. Selker, Q. F. Gronau, M. Šmíra, S. Epskamp, et al. Bayesian inference for psychology. part i: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, 25 (1):35–57, 2018.

[204] M. M. Wanderley. Gestural control of music. In *International Workshop Human Supervision and Control in Engineering and Music*, pages 632–644. Citeseer, 2001.

[205] M. M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004.

[206] C. Wang and P. Cesar. The play is a hit: But how can you tell? In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pages 336–347. ACM, 2017.

[207] D. M. Wegner. The mind's best trick: How we experience conscious will. *Trends in cognitive sciences*, 7(2):65–69, 2003.

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

[209] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. *Computer music journal*, 26(3):11–22, 2002.

[210] P. H. Westfall, W. O. Johnson, and J. M. Utts. A bayesian perspective on the bonferroni adjustment. *Biometrika*, 84(2):419–427, 1997.

[211] D. M. Wolpert and J. R. Flanagan. Motor prediction. *Current biology*, 11(18):R729–R732, 2001.

[212] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–1882, 1995.

[213] J. C. Wu, M. Huberth, Y. H. Yeh, and M. Wright. Evaluating the audience's perception of real-time gestural control and mapping mechanisms in electroacoustic vocal performance. In *NIME '16: Proceedings of the 2016 International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 206–211, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University. ISBN 978-1-925455-13-7. URL `http://www.nime.org/proceedings/2016/nime2016_paper0042.pdf`.

[214] Y. Wu, L. Zhang, N. Bryan-Kinns, and M. Barthet. Open symphony: Creative participation for audiences of live music performances. *IEEE MultiMedia*, 24(1):48–62, 2017.

[215] R. J. Zatorre and V. N. Salimpoor. From perception to pleasure: music and its neural substrates. *Proceedings of the National Academy of Sciences*, 110(Supplement 2): 10430–10437, 2013.

[216] R. J. Zatorre, J. L. Chen, and V. B. Penhune. When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, 8 (7):547–558, 2007.

[217] L. M. Zbikowski and L. M. Zbikowski. *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford University Press on Demand, 2002.

[218] M. Zentner, D. Grandjean, and K. R. Scherer. Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, 8(4):494, 2008.