

Perspective

An Introduction to Musical Interactions

Insook Choi ^{1,2}¹ School of Arts and Media, University of Salford, Salford M5 4WT, UK; insook@insookchoi.com² Wowsome XR Limited, 24 Old Bond St, London W1S 4AP, UK

Abstract: The article presents a contextual survey of eight contributions in the special issue Musical Interactions (Volume I) in *Multimodal Technologies and Interaction*. The presentation includes (1) a critical examination of what it means to be *musical*, to devise the concept of music proper to MTI as well as multicultural proximity, and (2) a conceptual framework for instrumentation, design, and assessment of musical interaction research through five enabling dimensions: Affordance; Design Alignment; Adaptive Learning; Second-Order Feedback; Temporal Integration. Each dimension is discussed and applied in the survey. The results demonstrate how the framework provides an interdisciplinary scope required for musical interaction, and how this approach may offer a coherent way to describe and assess approaches to research and design as well as implementations of interactive musical systems. Musical interaction stipulates musical liveness for experiencing both music and technologies. While music may be considered ontologically incomplete without a listener, musical interaction is defined as ontological completion of a state of music and listening through a listener's active engagement with musical resources in multimodal information flow.

Keywords: musical interaction; multimodal technology; multimodal interaction; HCI; action perception cycle; affordance; mental model; design; adaptive learning; second-order feedback; temporal integration; Common Practice Period

Citation: Choi, I. An Introduction to Musical Interactions. *Multimodal Technol. Interact.* **2022**, *6*, 4. <https://doi.org/10.3390/mti6010004>

Academic Editor: Kamran Sedig

Received: 16 November 2021

Accepted: 24 December 2021

Published: 2 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Music is a structured sonic event for listening. This description is inclusive of a listener who is an actor in musical interaction. Music without a listener is ontologically incomplete. Composers and performers model listening experiences by being listeners themselves in planned or on-the-fly production of musical events. In ways, music models a meaning that precedes logos. While there are shared neural resources in both music and language processing as shown in [1–4], a musical experience is not readily and effectively describable in words. Therefore, we can state that music, in its initial encounters with listeners, is relatively free from, or defers the kind of syntactic and semantic probing required for language processing. At the same time, music appeals to language when a listener wanders through a musical landscape exploring how to describe their listening experiences, to convey meaning to themselves and others by ‘figuring things out’. The ‘things’ here are perceived musical elements or features from a background event, and to ‘figure...out’ means to draw distinctions among musical elements. This process engages neurophysiological pathways from low level sensory motor mechanisms to cognition as well as social and cultural reference frameworks. Music inherently evokes multimodal interaction and musical interaction research can potentially catalyze a community of practice when applied to the field of multimodal technologies and interaction (MTI).

This MTI special issue on Musical Interactions (Volume I) is a response to an emerging research opportunity among multiple disciplines that share a growing history of conversations around three foundational topics: music, technologies, and interaction. For musical interaction, the research agenda is yet to be refined, and calls for understanding the

complexity of coordinated dynamics that form a circuit between technologies and human action perception in musical liveness. Presented as a research theme, “musical” is the active qualifier and “interaction” is the designated subject. This is inverse to the research perspective where “interaction” is a qualifier, such as “interactive music” or “interactive technologies”. For a quarter century, increased computing power has opened many possibilities for musical information encoding and retrieval, applying real-time digital signal processing (DSP) and interactive input-output processing, which are also utilised for conditioning the state of interaction. This era has welcomed novel applications with interactive music systems engineering which exploit multiple forms of technology appropriation in musical practice. Prolific publication has trended with human-centric or machine-centric investigations, including audience-user studies producing data that is both self-reported and psychologically measured. User Interface and User Experience (UIUX), real-time DSP signal flow, interactive system architecture and prototyping and related system performance demonstration and evaluation are also relevant for musical interaction. Among these, two MTI-centric phenomena remain underexplored: (1) Musical experiences are ever contextualized by crossmodal perception and cognition, and by listeners’ situated encounters. How do the interactions with multimodal technologies impact people’s multimodal perceptual processing and subsequent musical experience? (2) Many situated interactions arise in the context of technology applications in daily life, where musical experiences may have broader impact on the quality of interaction. What kind of musical interaction is desirable with technologies in daily life and how do we recognize when there is one? For musical interaction, it seems to be clear that the concept of music can also be broadened and devised with respect to fundamental aspects of human perception and cognition, beyond music specialisation, and this presents challenges as well as opportunities in the current research landscape. Whether applied to scientific testing or aesthetic experience, musical interaction with multimodal technologies will always require an instrumentation, a process for connecting human subjects and devices to enable information exchange and data flow.

With multimodal technologies, it is essential the meaning of music and interaction be mutually modified to satisfy emerging requirements and diversity of human experiences with evolving technological capacity. Multimodality of music has been recognised in research across modelling [5,6], music information retrieval [7,8], music therapy [9–12], and multisensory and crossmodal interaction [13,14], many of these informing the context of embodiment and mediation [15]. Music Supported Therapy (MST) especially in conjunction with neurological data has demonstrated positive impact on brain recovery [16,17] and on motor recovery as is evidenced by clinical measures. Wan et al. [18] and Ghai et al. [19] inform MST by investigating underlying auditory motor connectivity and coupling [18,19]. Neuroscientific findings in conjunction with brain imaging also inform the impacts of music on brain development through a structural adaptation by long-term training [20–22], and the impact of musical multisensory and motor experience on neural plasticity [23–25].

Scientific findings from the above literature bring an increasing awareness of the multimodal nature of music. The implications from research design in this literature also point to the importance of instrumentation. There is a promising space with both depth and breadth in the topic of musical interaction. The goal of this issue is to be inclusive and mindful to cultivate coherent breadth for that depth. Identifying the loci of musical interaction poses significant challenges, which may require new approaches to how we behold methods and evaluate both music and systems for studying purposeful, planned, intended, or even straying actions emerging through a musical experience, in ways that are inclusive of interactive and listening experiences of both human and technological actors. In this regard, it is valuable to see the range of original research in this collection. This special issue on Musical Interaction includes topics wide-ranging yet connected, from engaging mobile devices to supercomputers, from prototyping pedagogy to production of musical events, from tactile music sensing to communicating musical control by eye gaze,

from literature survey to development perspectives. Leveraging the diverse paths demonstrated by the authors, we can expand horizons in musical interaction research and application areas with more inclusive multidisciplinary teams. New horizons will also deepen our perspective on how to anchor musical interaction as a research theme, to influence ways of engaging devices and technologies for mediating human activities and experiences in daily life. The purpose of this article is twofold: (1) To provide an exposure to the depth of discourse undertaken in a community of practice evolved around music as a central focus, and (2) to consider a conceptual framework to discuss and assess musical interaction in MTI context whereby the contributing articles are contextually surveyed.

Organization of this Article and Rationale

This article is organized in two parts for a twofold purpose. Part one includes Sections 2 and 3 where a context of musical interaction and theoretical framework are discussed with respect to a community of practice. Part two includes Sections 4 and 5, which introduce and discuss each of the contributing articles and present a summary and conclusion. For a deeper perspective on musical practice, readers may consult Appendix A which provides a succinct survey of the European historical Common Practice Period (c.1650–1900). Appendix A is developed to foreshadow our emerging community of practice by illustrating a handful of examples and referential frameworks to minimally disambiguate what it means to musically interact, whereby to motivate (1) imagination of what it is like to form an emerging practice with the qualifying criteria of music, and (2) strategic thinking towards future poles and markers proper to MTI agenda. It is important to be cognizant that musical origins which we share are likely to belong to a past Western classical music tuned to certain ethnic and social groups, therefore our informed reflection is in order, on our assumptions of what it means to be music.

Section 3 considers a conceptual framework for research and design with five enabling dimensions for musical interaction. The 1st dimension, Affordance, adopted from Gibson's theory of ecology [26,27], denotes (inter) action possibilities perceived by an agent through the relationship between the agent and its environment. The 2nd dimension, Design Alignment denotes the process of identifying design criteria for gathering requirements tuned to musical criteria so that the design resources and music resources in an interactive system can be structurally coupled. The 3rd dimension, Adaptive Learning denotes the system capacity to facilitate a user's learning pathway as well as their learning capacity. The 4th dimension, Second-Order Feedback is a term adopted from general systems theory (general control theory) [28,29] and second-order cybernetics [30]. Here, it is applied to multisensory feedback processed with top-down auditory attention (see Section 3.4) by listeners to observe and guide their own sensory-motor performances. The 5th dimension, Temporal Integration denotes the requirement for time critical data transmission in a multimodal system's architectural flow between users and system components to facilitate musical interaction. In Section 3, each dimension is presented as subsections and discussion focuses on MTI context. Section 4 introduces, surveys, and discusses the contributing articles with respect to the five dimensions. Section 5 presents a summary and conclusion.

2. Musical Interaction as a Community of Practice

A community reflects a referential framework comprised of a shared skill set and expertise, methodologies, assumptions informed by literature and theories, and a set of criteria for evaluating outputs, all together characterizing an ecosystem of the field [31–33]. Accordingly, music of the Common Practice Period (Appendix A) evolved around literature (in the form of musical scores with implicit performance practice and music theory), performance repertoire (a canon of compositions written for an instrument or ensemble) and instrument design aligned with literature and performance idioms (for example viola da gamba vs. violin). In current music research involving technologies, a topic relatively unexplored is what may constitute an ecosystem, meaning a system of

interconnected ideas or elements shared in a community of practice, that would catalyze concurrent and mutually supporting developments across literature, repertoire, instrument design, and their equivalents, through which a coherent research pathway is foreseeable. Musical interaction research in the context of MTI requires a contextual shift with an interdisciplinary research team beyond music specializations. In that sense, a community of practice for MTI musical interaction takes a different turn: literature will include scientific discoveries as referenced in Section 1; repertoire will include neuroscientific findings on multimodality in music production and perception, demonstration of case studies, data, interaction patterns, and use cases where instrument design can be analogous to designing DMI (Digital Music Instruments) and will extend to designing interactive systems and prototypes.

What constitutes musical interaction? When is an interaction musical and how does it come about? In recent Eurocentric tradition, music is identified as a persistent memory in a notated form so that a musical event can be reconstructed by enacting that memory. In the Eurocentric tradition, music is often approached as synonymous to its notated musical score. In a crude digital analogy, a musical score is more akin to a Memory Address Register than to musical data stored in that memory. Music notation is a placeholder aiding performers to execute tone, which requires interpreting a represented code of musical information. This case holds for music performed from a memorized score and music improvised from a simplified score such as a commercial music lead sheet. Horsley defines a score as a visual representation of musical coordination [34]. Musical coordination includes temporal execution for both sequence and synchrony of tone production, managing speed and regularity of meters and beats, applying expressive dynamics to tones and phrases, and teamwork in an ensemble. A score is a kind of performance manual for executing and coordinating musical information, but the instruction is implicit because the musical score is representational, and interpretation refers to a tradition in a community of practice. This means, even with our highly developed notation system, performance is implicitly aware of the oral tradition specific to Eurocentric culture. To facilitate the persistent memory from generation to generation and from instrument to instrument, the European notation system evolved over 1000 years before its use in a fully developed form during the Common Practice Period. This era recognizes a widespread Western art music tradition from the mid-17th to early 20th centuries, centered around a shared tuning system, a system of tonality, a common notation, and a theory of harmony. Together these constitute an ecosystem where a theoretical framework for common practice matured, as illustrated in Appendix A.

Establishing a common practice of European art music, reinforced music as a specialized domain requiring highly tuned skills to compose—which is to register the representation of memory, and to perform—which is to enact the memory for artful retrieval. The expertise of a composer or virtuoso cannot be acquired in any measurable magnitude of investment. Music virtuosi dedicate their entire life to one and only one thing, performing instruments. In this tradition, the word “music” carries a specific meaning requiring a finely tuned tonal acculturation [35] for a listener to apprehend the idiomatic expressions. It is noteworthy that the Common Practice system of tonality affords the musical idioms built around the stability and instability in tonal relationships, which are directly linked to perceptual consequences in listening. Composers of the Common Practice Period artfully played with this link. Wagner’s well-known Tristan chord marks an extreme boundary by fully exploiting and exhausting the use of this link with prolonged tension-building musical schema, without violating the rules of play. This tradition portrays a well-developed and established ecosystem, which provides an affordance for a common reference framework for diverse idioms and expressions. These references suggest the long-term relevance of an alternative and informed concept of a music applied to MTI.

3. Musical Interaction Research and Design: A Conceptual Framework with Five Enabling Dimensions

The complex ecology of the Common Practice Period constituted a system of references, notation, theory of harmony and musical forms, which evolved through a long discourse primarily concerned with ratios in sonic events that are directly linked to human senses and cognition. Today, we are advanced with new tools and techniques, and we are no longer bound to the Common Practice paradigm. Whatever assumptions we may have for what it means to be musical, we can step far back in an attempt to look at the very foundational conditions of music, ‘going back to the drawing board’ if you will.

First, the concept and idea of music is sensorially concrete but linguistically elusive. The potential semantic space for the word “music” is vast, perhaps much larger than many other words. The concept of music is highly subject to variance according to culture and epoch as well as individual memory. For example, we cannot experience the phenomenon of the Greek chorus of its time; not only is it nearly impossible to reconstruct the music from remaining papyrus fragments [36], but also modern people hold no clue to decode the musical information registered in the papyrus that can be only interpreted with a priori oral tradition [37]. In terms of musical experience, we live in a world of connectivity through technological devices and infrastructure where multicultural or diachronic proximity cannot be reduced into the sociological or historical concept of “cultural diversity” or an “origin” [38], because every culture generates its own lived experience [39,40], and individuals’ memory formations are also influenced by their cultural backgrounds [41–44]. Altogether, these implicate the need to understand the relationship between people and music in radically different ways with multicultural perspectives, thinking through what it means to be musical in the context of MTI.

Second, music is an ephemeral phenomenon. No live musical instance can be exactly repeatable due to the emergent behaviours among all interacting bodies involved in the phenomena. This aspect of music has been underrepresented historically but can be highly relevant for MTI. Effectively, MTI presents a multitude of affordances to explore musical liveness. Understanding the interaction dynamics is critical whether the interaction is for composing, performing, listening and especially for *experiencing technologies with musical interaction*. This implication of technological interaction upon music can be traced to the early 1950’s and the history of the experimental marriage between music and computing machinery. Early experimenters include Lejaren Hiller with the ILLIAC 1 [45], Max Mathews with the IBM 704 & 7094 [46], and Alan Turing with the Mark II [47]. They endured non-real-time workflows with laborious programming and waiting hours or days for the results. Most of the experimenters were not professional musicians and some were not accepted as legitimate musicians in recognized fields. Nonetheless, their experimental outcomes catalysed a new discipline called computer music, which became an agency of the results and techniques used in new millennium platforms such as web, games, and mobile applications. These widespread outcomes of early investigations suggest that it is not prudent to assume that ‘music is music’, and it makes more sense to ask, “What constitutes musical interaction?” and “When is interaction musical and how does it come about?”. Let’s also address what the early experimenters did with same questions. The following discussion is based on the hypothesis that, unless we do so, we may not be able to articulate the deeper relationship between music and technologies in terms of any form of interaction reflecting human input, choice, cognition, machine processing and mediation, outputs, and experiences, which altogether constitute a contextual adaptation in musical interaction, not a causal link, yet informed of causality.

Third, musical interactions are designed, whether for a concert, gameplay, pedagogy, therapy, exercise, or psychological testing. This research collective suggests an agenda to conceive of a coherent path for characterising musical interaction for MTI with a frame of reference that we can consult for designing and assessing a project. The following discussion identifies five enabling dimensions to consider for musical interaction on which a theoretical framework can be explored:

- Affordance
- Design Alignment
- Adaptive Learning
- Second-Order Feedback
- Temporal Integration

The choice of these terms is informed by design practice, Human Computer Interaction (HCI), AI, music composition and engineering.

3.1. Affordance

Multimodal technologies afford musical interaction by situating expectation. An actor anticipates how an action will produce a musical outcome, and the way this expectation is resolved contributes to the formation of a mental model of musical interaction. Researchers and practitioners investigate how people utilize their senses when interacting with technologies, asking what properties appeal to people suggesting repertoires of actions or set expectations for experience potential. The role of expectation in musical interaction can be approached by investigating affordance and mental model.

The concept of affordance was conceived by J.J. Gibson [26,27,48] for describing an ecological relationship between an organism and an environment, where the organism perceives its environment as to offer potential resources. In 1988, D. Norman introduced the term to the design community [49,50]. In 1991, W. Gaver introduced the term to the HCI community, proposing that affordance be considered for designing perceptible objects or features to offer information how they may be acted upon or explored for complex actions [51]. Reybrouck [52] exercises two terms, “music users” and “sonic environment”, the former denoting the observers and the latter standing for the broadened concept of a music. Music users explore musical affordance in sonic environment. Menin and Shiavio [53] propose to investigate musical affordance around the sensory motor experience that is pre-linguistic, involving intrinsically-motor-based intentionality. Consistent to this line of thinking, Krueger [54] describes listeners as active perceivers and solicitors of musical affordances. In this issue, Rowe [55] discusses an affordance in relation to representation. The term, affordance, resists a simple conceptual model. Interpretation and adaptation of its concept differs across research communities and can be confusing, making it hard to see how it can be applied to research and practice. Affordance is also deeply related to the concept of mental model, which is another term often used with over-simplifications or over-complications. The following discusses these concepts as interrelated by referring to their original sources.

Gibson constructed the word affordance from the verb “afford” and he “... coined this word as a substitute for values ...” to “mean simply what things furnish, for good or ill. What they afford the observer, after all, depends on their properties.” [48] p. 285. Therefore, affordance is what an environment offers. At the time of prevalent behavioural science and the beginnings of cognitive psychology, Gibson authored *The Senses Considered as Perceptual System* [48]. In the preface, he expresses his attempt to reformulate old theories such as stimulus-response theory, Gestalt phenomenology and psychophysics to extract new theorems. Gibson’s theory of senses denotes senses as active and outreaching perceptual systems to acquire perceptual information about the world, therefore senses are “to detect something” (active) rather than “to have a sensation” (passive). Note the information is perceptual, meaning it is internal to the observer. In this formulation, affordance is something that is perceived and explored by an observer. Sensing an affordance is preceded by another state of observation that occurs “When the constant properties of constant objects are perceived (the shape, size, color, texture, composition, motion, animation and position relative to other objects), the observer can go on to detect their affordances.” [48] (p. 285); “The properties of perceived... are nutritive values or affordances.” [48] (p. 139). As our senses are active perceptual systems and that activeness coordinates movement for exploring affordances, to perceive is to obtain information

about an environment, what affordances (possibly value propositions) it may offer. In that sense the perceptual information about an affordance is a perceived value potential. By using the information, an organism orients itself to action, whether to explore or exploit the detected affordance. Based on the outcome, an organism's perceptual information as well as its own action can be assessed and updated with respect to the previous state of its perceptual information. In this regard, an observer's actions performed upon the perceived affordances assume a perceived set of invariants from an environment and that becomes a basis for anticipations or expectations.

Gibsonian affordance characterises the nontrivial circularity built in the relationship between an observer and an environment in which the observer is part of their environment. Belonging to the environment provides an essential dimension for musical interaction, for characterizing the relationship between people and systems. The term is also intended to bypass the dichotomic division between subjectivity and objectivity, which Gibson considers to be inadequate for describing the (entity) relationships in an ecology [48]. The Gibsonian perspective relates to *What the Frog's Eye Tells the Frog's Brain* [56]; what senses are registered in an organism's perceptual mechanism determines how the environment is perceived and what affordance it may present. For design orientation, affordance leads us to attend to the conditions of interaction, not to the system features and properties or to an affordance directly, but to the thoughts for how the features and properties can contribute to the relationship between systems and users. One cannot design an interaction nor an affordance without attending to the conditions from which the affordances can be perceived by users for action-interaction potential. For musical interaction, an affordance can be envisioned by creating a set of propositions that a system may appeal to users as meaningful, meaning how the system is designed to suggest users what they may gain or experience. A problem is that affordance may not be assessable in a quantitative measure when prototyping and testing user experiences. Here, mental model comes as a pragmatic tool, not as means of representation of, or for modeling users cognitive states, but as a design tool to work with users on what value potentials they recognized, what expectations they had and what action-interaction they thought as possible when they encountered and interacted with the system. Both qualitative and quantitative assessments become possible by measuring the discrepancy between design intents and users' descriptions, and the discrepancy between users' descriptions (what users understood they were doing) and system data (what the system recorded users actually did).

The concept of mental model in its original conception can be traced to Kenneth Craik in 1943 [57]. The Artificial Neural Network (Warren McCulloch and William Pitts) was also introduced in that year, inspired by biological neural mechanisms, especially the associative nature and causal links, which also had an impact on Craik's thinking in the relationship between human operators and technologies. Craik describes a kind of thought model simulated in an organism's head, a 'small-scale model' of external reality and possible actions in it (and with it). In this conception, the model means a system, whether physical or chemical, that "... has a relation-structure similar to (or parallel to) that of *the process it imitates*" [57] (p. 51). The similarity is in the structural relationship and does not hinge on a pictorial resemblance. Craik uses an example of tide predictor that has no resemblance to tides, but it produces oscillatory patterns to imitate the variations in tide level. An implication for designing an artificial system is to inquire of the structure and processes for implementing the system that can imitate and predict external processes, the process external to the machine, that is, the state changes of input signals from people or anticipating what users may do. For people to know how to input signals or what to do next, they use mental models. Therefore, a system needs to be presented to facilitate the (in)formation of a mental model analogue or compatible to what it can offer.

Johnson-Laird describes the process of constructing and using mental models as different from formal reasoning based on a set of beliefs: mental models represent distinct possibilities, or a kinematic sequence unfolding in time, whereby we base our conclusions [58]. The formative process of a mental model allows people to handle uncertainties such

as what their actions may entail or what to do next when first time working with a system of representation standing for a set of functional premises coded in the system. Consistent to Johnson-Laird's description [58], the process such as working with interactive systems, especially learning to play/work with novel systems, is very different from formal reasoning with a set of beliefs due to the constant adaptation process through an exploration of affordances. It also means making use of working memory to perform tasks and handle counterexamples. Through experiences, people can modify, improve, or refine their mental models. Appendix B illustrates the relationship between mental model and affordance.

3.2. Design Alignment

Aligning the variety of the resources between music and design determines the quality of musical interaction. Design tasks tend to involve identifying problems and solutions whereas music is more about generating structured sound events. While both require an aptitude of creativity to complete the tasks, the orientation of the tasks in each field consults different criteria. While musical instruments evolved over centuries in the musically specific ecosystem, musical interaction in MTI needs to consider diverse applications and translating the principles from music to a system of interaction through design implementation. A multimodal interaction requires an instrumentation with multimodal technologies for engineering a site for an actor, which introduces the investigation of interface and affordances. The general HCI human in the loop and gameplay are good references but insufficient for music. Musical interaction requires musical propositions. Musical interaction with MTI requires technological propositions. The values musical interaction may offer are not obvious largely due to the lack of existing mental models, while the values music may offer are more familiar with the mental model of a listener being seated, physically inactive, regardless of the state of listeners' mental activity in perceptual and cognitive engagement. With MTI, audiences are invited to generate musical experiences by proactively experiencing technologies, which means, MTI musical interaction proposes one interaction, that is musical, for two experiences, that are of both music and technology. Designers and researchers for musical interaction cannot take such things for granted. First, user interface components (UI) and system interfaces need to be temporally aligned to support time critical interaction. Second, musical agenda and design directives need to be well aligned with a shared vision of what kind of relationship to people a system may afford, what kind of active participation a system suggests, and what exposure does a system present to access control, all together mobilizing the formation of a mental model. Third, the system implementation needs to be guided by a structural relationship between musical agenda and design directives, so that task domains for generating music resources and design resources are compatible with the target range of musical outcomes and design solutions.

Three priorities are (1) Defining design tasks for a system of interaction to be musical; (2) defining the repertoire, variety, and features of musical resources so that musical interaction can grow alongside the formative process of a mental model; (3) aligning musical resources and design resources so that interactivity can be conveyed to people as action possibilities, with the intended ranges of action perception cycle and the degree of controllability. Musical interaction in MTI requires good control strategies for generating sounds through UIs that embody an affordance; both system and user interface define the variety of control, the interaction repertoire, and the access to hierarchy of information granularity.

For musical interaction, an affordance can be envisioned by creating a set of design propositions. Aiming for or assessing affordance in design can be simplified by a target-set of constructs measured against hypothesis and outcomes. Affordance clues in users to orienting for an interaction possibility, therefore affordance is related to but not the same as mental model. What values users may gain, and how to achieve them can be only suggestive. Beyond this phase it is a mental model that leads to more specific design directives. In a design process, mental models are consulted to determine what kind of

interaction pattern is suitable for which task, or an overall workflow for a complex task. The compatibility between system and people's mental models is assessable from a user's perspective by soliciting a set of use cases and workflows in their models, then measuring against the design intent, implementation, and system performance (see Section 3.1 for related discussion).

The term, *requisite variety*, is a useful concept applicable to the structural relationship between design and music: Ashby's law of requisite variety states, "Only variety can control variety" [59]. The design tasks and music tasks need to be mutually informed of the possible variety, and at some point, define the scope within which the two tasks optimize the alignment. For example, in terms of system interface, musical interaction is, by definition, time critical, which requires highly synchronized system feedback to support the user's action perception cycle (see Sections 3.4 and 3.5). Both UI design and music tasks can provide and receive information of how system architecture and instrumentation are to be aligned to account for which modality is synchronized with which channels of interaction data. Accordingly, system designers will allocate resources for concurrent signal processing and scheduling including asynchronous data buffer management to optimize the real-time data flows among various components interfaces. These processes ensure that the variety in the domain of system control is compatible to the desired variety in the range of musical outcomes.

3.3. Adaptive Learning

We are all novices in the first conception of digital technologies. Musical instruments were designed for virtuosi. Digital interfaces are designed for novices. A musical instrument is an interface between a performer and sound events. In recent trends, some HCI researchers in digital interfaces have been inspired by an analogy between musical instruments and new interfaces. While the analogy may evoke a compelling intuition, it has been a source of misconception in the field of digital music interfaces and interactive music software. Historically, musical instruments have been evolved with virtuosi, the masters of certain musical instruments in certain era. J.S. Bach was a driving force behind the development of the keyboard instrument known as Well-Tempered Klavier applying finely tuned string ratios known as Well-Temperament, a tuning system, which enabled Bach to play on 24 different keys. These are precursors for the piano as we know now, and the modern tuning system called Equal Temperament (see Appendix A). Virtuosi were the forces and users for whom the modifications and improvements were made as instrument makers tailored and tested their product. The process of perfecting a musical instrument was like a prolonged physical coding exercise, an iteration upon eliciting expert knowledge, changing the instrument's physical structure, and testing with mature skills.

Unless for an expert system, a virtuoso consultation may risk a bias for conceiving digital technologies. We are all novices for digital interfaces and for interaction in its first conception even for professional digital tools. Most musical instruments offer few playful entry points below a certain skill level. In contrast, MTI with musical interaction is open to consider a system that affords (1) a playful entry point for all skill levels, and (2) a trajectory or capacity to evolve along maturing skills. To advance a musical interaction paradigm with MTI, we can envision not only facilitating novices' playful interaction but also designing a system of interaction that can afford a mutually supportive learning between people and machines with AI and machine learning techniques. A playful entry point can facilitate a user's orientation to the system from the very beginning by forming a crude mental model. Beyond that point, it is desirable for a system to learn to sense and keep pace with a user's skill level to support an adaptive learning pathway, and to facilitate users to advance by dynamically maturing and refining their mental models. Otherwise, a progressive level design can suffice, as is common in game design.

3.4. Second-Order Feedback

We observe our own actions by observing the result of our actions. Music performance is a sensorimotor coordination flow guided by auditory perception to achieve a musical goal. In second-order cybernetics, von Foerster describes an observer as an observing system that observes itself while accounting itself as a part of its observation, thereby observation affects the observed [60]. This is consistent with music performance where a performer's state is constantly affected by the performer observing her own performance. In this context, Second-Order Feedback sensitivity refers to the physiological and psychological condition of a performer engaged in time critical action perception cycle interacting with an instrument or a device in a certain environmental setting. Fuster describes perception action cycle as the circular flow of information between the environment and an organism's sensory structures, in which an organism is engaged in sequences of sensory guided actions with a goal directed behaviour [61]. What is implicit here is the environmental feedback that influences an organism's sensory guided actions which in turn influence the environmental changes or responses. For musical interaction, the circular flow is an important concept, and it is more proper to call it action perception cycle because a performer always initiates an action to perceive her own action and its outcomes. The idea of circular flow was already seeded through the well-known 1940s and 1950s Macy Foundation conferences on an interdisciplinary topic "Circular Causal and Feedback Mechanisms in Biological and Social Systems", chaired by Warren McCulloch. Notable participants included Gregory Bateson, Norbert Wiener, Margaret Mead, and W. Ross Ashby, catalysing the meta-discipline known as Cybernetics. This class of cyclical model for embodied action is reflected in a broader context of discussions in cognitive science and philosophy of mind [62–65].

To illustrate this dimension, let us take a familiar example. A violinist performs an action of bowing on strings and this introduces an excitatory energy into the resonating body. The subsequent sound quality is its resonance response to the input patterns (excitatory signal patterns) further shaped by the violinist's left-hand control. By engaging an external body, which is a violin in this case, the interactive feedback cycle involves the second-order circular flows extending from the performer's body to the instrumental body. In effect, a performer acts and perceives her own sensory motor coordination, which constitutes the first-order circular flow, and she also assesses how her action entails the resonating responses of the external body, which constitutes the second-order circular flow. For the performer to anticipate and to project a following action, it is critical that the second-order circular flow feeds back multisensory information including auditory feedback, a series of audible complex waveforms that conveys the quality of sounds. For a violinist, the auditory feedback follows the immediate tactile vibratory sensation through her chin from the violin as a resonating body as well as the friction sensed from bowing, whereby she perceives and confirms her own action assisted by proprioceptive feedback. Krueger describes this as "ongoing mutually regulatory integration" involving motor entrainment [54]. Particularly, music performers are trained to use auditory feedback to assess their own performances engaging auditory attention, attention regulation, and expectancy, which are higher-cognitive processes in auditory perception as explained in [66–68], the processes engaged in the second-order circular flow. Further, an acute situational awareness is required for a performer to assess complex environmental responses such as resonant frequency and amplification characteristics of a concert hall so that she can fine tune her performance to achieve an optimal projection of sounds to bloom in a particular hall. In sum, what determines the quality of a performance is an artful execution of sensory motor coordination in an effective and timely manner mastering the circular flow in musical interaction with an instrument and her own situational awareness.

To account for the time critical nature of musical interaction, the term, Second-Order Feedback is adopted from the general control theory [28,29] and cybernetics [30,69,70]. To respect a performer's second-order feedback sensitivity, the foremost important factor is timing. For musical interaction in MTI, the timely coordination between the temporal

granularity of computing and performer's multimodal action perception cycle is the most challenging problem. It requires a careful temporal alignment between DSP units for computing sounds, and desired composition of interactive signal granularities applied both to control sounds and to communicate with various interface units. Musical interaction, by definition, is driven by auditory feedback along which performers (users) transition from one state to the next by assessing multisensory feedback with directed attention to sounds, thereby influencing their actions. Perceiving music involves a multilayer temporal structure for processing pitch (related to frequency), timbre (related to frequency spectra), and rhythms (comprised of unit durations of note relationship and the patterns of disposition). Accordingly, musical interaction applied to MTI requires a multilayer DSP compatible to human perception and real-time feedback capacity, to support a user's fluent action perception cycle. In the author's multimodal musical interaction practices, the timing represented in the LIDA model (Learning Intelligent Distribution Agent) [71,72] offers a useful temporal framework compatible with the cascading flow of action perception cycles in multimodal music performances [73,74]. Figure 1 illustrates Second-Order Feedback and the LIDA model related to temporal integration.

3.5. Temporal Integration

Any system of musical interaction is a complex system that requires a temporal integration in its architecture. Musical interaction agenda in MTI may note a special emphasis on temporal requirements for designing and engineering real-time signal and information flow among all components in a complex system with respect to end users' second-order feedback sensitivity. The choice of parallelization to compute concurrent real-time processing, for the wholistic signal flow through all multimodal components, may result in down sampling some perceptual features, and this choice is prioritized by second-order feedback sensitivity. People's perception affords a contextual adaptation within an acceptable temporal range. Three main properties subjected to temporal integration are (1) the unit definition of human performance gesture, (2) signal mapping and navigation responses of designed components, across UI and system interfaces and (3) DSP component responses. Requirements for temporal integration are contingent to the design alignment in terms of the definition of temporal granularity and mapping between control gestures and musical outcomes (through wherever modalities the control inputs are channeled), as well as the degree of indirection from action to perceived musical results.

Further details can be found in [73,74] with multimodal performance examples applying temporal integration among multimodal components and an architecture for musical interaction. Figure 1 (adapted and modified from [74]) illustrates a general architecture for temporal integration, based upon implementations of multimodal performance systems. Parallel streams of multimodal interaction generate first-order and second-order feedback through image/video and sound/music streams. Digital processing generating interactive media streams maintains temporal symmetry with micro, meso and macro timescales of user perception and cognition, parsing the user's continuous actions into the three temporal control bands. Temporal integration anticipates the user's mental model of governing the pacing of action.

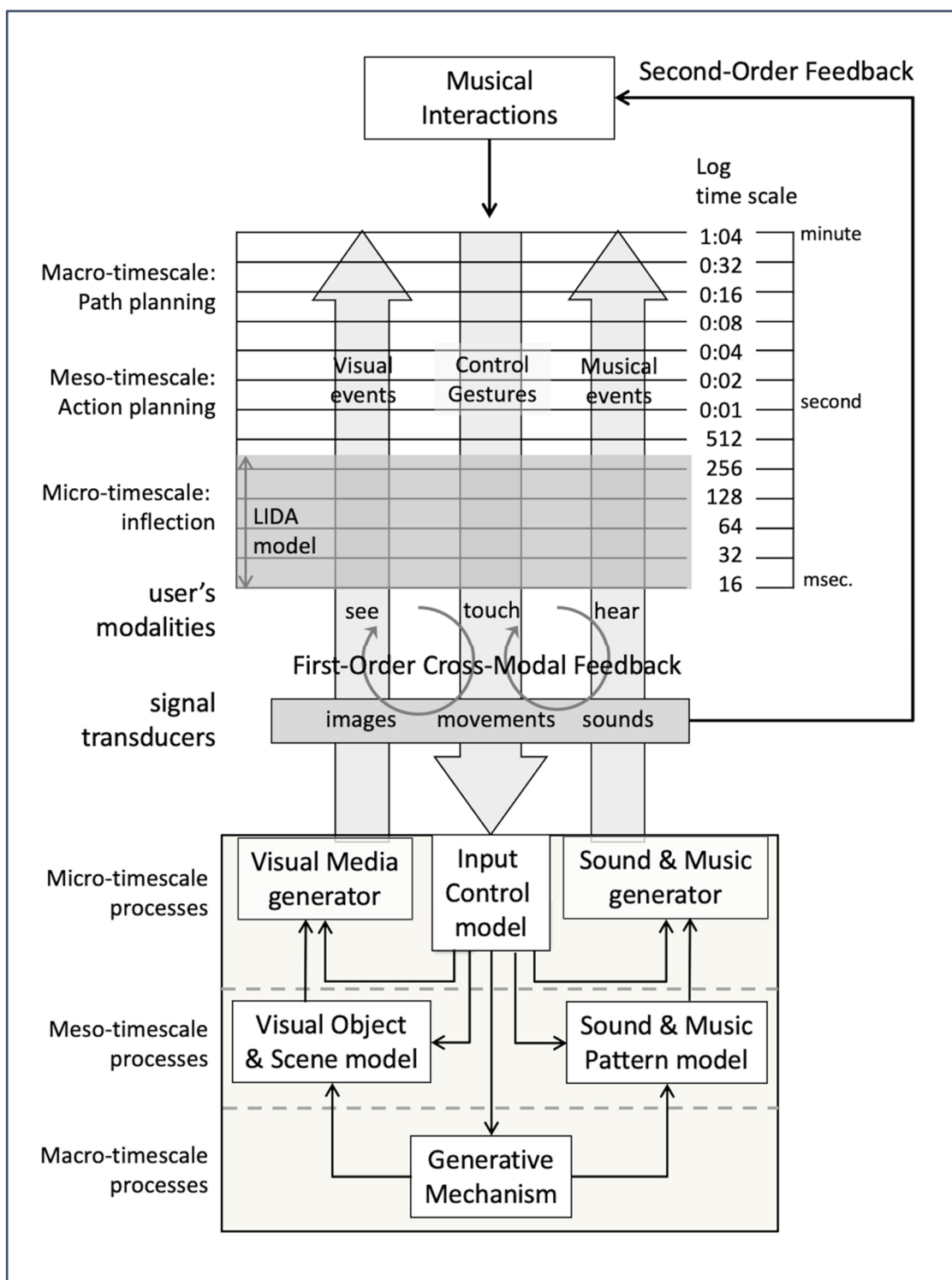


Figure 1. Schematic temporal integration architecture for musical interaction with multimodal and multimedia system: the parallel vertical data flows illustrate the interactive processes between user and system. A user’s action perception cycle is represented in the temporal regions of a cognitive cycle following the LIDA model [71,72]: a cycle constitutes the temporal windows of sensing, attending, and acting, engaging both pre-cognitive and cognitive processes. In this diagram, user input signals are converted into control data through a Generative Mechanism—a background visual media model—which routes control signals to parallel subsystems for generating sound and visual media. Users experience multisensory First-Order feedback. Users engage Second-Order feedback to observe multisensory outcomes with directed attention to auditory outcomes for action planning. In musical interaction, both levels of feedback cycles cascade in a continuous flow for sensing, attending, and acting. Adapted and modified from [74].

Referring again to Craik, his original conception of mental model describes a kind of servo-control mechanism with built in prediction of when the necessity arises so that "... sensory-feedback must take the form of delayed modification of the amplitude of subsequent movement" where "... the sensory control can alter the amplification of the operator with a time lag and determines whether subsequent corrective movements will be made [75] (p. 87)." While such servo-control mechanism has been advanced in many mechanical systems, especially for sensory-driven multimodal coordination in robotics, this is still a challenging problem in time-critical multimodal music performances due to a performer's second-order feedback sensitivity. In highly time critical interactive music performance, even with powerful CPUs and parallel computing, 10 ms. delay can cause a disruption. This is critical because **the mental model for musical interaction is also a model of time**, requiring consistent support for anticipation, meaning predicting in time, with clear presentation of accessibility and variability of unit control and temporal granularity. Some time lag is prone to happen in experimental real-time multimodal performances that require intensive computing resources. A system affordance in such case is to secure tolerable temporal variations within a predictable range so that a performer can develop tolerance and recovery skill to counteract small inconsistencies. When a performer is committed to such experimental systems, especially for prototyping, a flexible mental model is helpful. When performing with a multimedia or multimodal system, by trusting the reliability of synchronization between input and output signals, a performer may explore the available affordance of performance gestural repertoire.

4. The Contributing Articles

This section is dedicated to introducing and surveying the eight contributing articles, detailing and positioning them as future propositions. It is noteworthy the authors who contributed to this issue perform highly interdisciplinary research and found their ways to contribute from many different perspectives. The presentation of their work is ordered in recognition of the articles' diverse orientations. The first and second articles, authored by creative practitioners, are discussed in sequence to illuminate similarities and differences on aims and approaches, artistic dimension, maturity of detail, ways of engaging technologies, and implications for MTI. The third article is a literature survey proposing a framework for classifying creative outputs of interactive sound installations. From this trio of articles, readers may draw further implications on the relationship between documentation, representation, and creative practices. The fourth article is a perspective essay which pivots between the initial trio of creative investigations and a following set of scientific investigations. The latter offer innovative inquiries, presenting research questions and proposing solutions for musical instrument online learning, mobile computing, communicative gestures in music performance and cross-modal musical interaction. For each of these articles, Section 5 shows the distribution of research in each article as aligned with the five enabling dimensions of musical interaction.

4.1. Using High-Performance Computers to Enable Collaborative and Interactive Composition with DISSCO, by S. Tipei, A. B. Craig, and P. F. Rodriguez

By engaging High Performance Computing (HPC) in music composition, the first author, Sever Tipei [76] sustains the Illinois School of experimental music, active since 1956 when Lejaren Hiller produced the first algorithmic composition (*Quartet No. 4 for Strings 'Illiac Suite'*, 1956 [45]) utilizing the early Illiac supercomputer. Hiller, originally trained as a chemist, pioneered a kind of non-real-time musical interaction using computational processes for generating musical instances, which gave rise to what we call algorithmic compositions. Given a set of rules and instructions as inputs, computation returns instances of outputs, which Hiller translated into musical notation. Tipei et al. [76] addresses how 21st century HPC may be engaged in the algorithmic processes applied to both musical structure and sound synthesis. The rules of the interaction, or the rules of the game played with a computer if you like, stipulate that to preserve algorithmic

integrity the composers should not change the output to create arbitrary musical effects. Situating a machine process in the middle of a creative workflow entails an alternative human behaviour for creative pursuit in the presence of large-scale computation.

A kind of discourse, “from musical ideas to computers and back” as eloquently expressed by Herbert Brun [77] is necessitated by the constraints and integrity imposed by time-bound state of the art HPC. Often, the discourse works in parallel for both optimising use cases and defining the next generation of HPC tools and methods. Perhaps, those with a fine taste of traditional music must be invited to experience a kind of “music that I don’t like, at least not yet”, meaning that such musical outputs may need time to mature in the listener’s expectations. This approach implies a kind of musical interaction where a listener has awareness of the input actions that were applied to produce musical outputs. **A participant’s awareness of input actions is a foundation of all musical interactions.** In Tipei et al. musical interaction extends from sound production into the domain of composition. For those who work with an algorithmic process, this consideration is in part philosophical and in part ethical. What “I don’t like yet” is the result of what I did, meaning the choice can be made: Either change what I did as input, rather than change the output to imitate something I did not do, or I learn to understand the system and my inputs and keep discovering what the result may offer.

Tipei is a highly acclaimed composer and pianist who is committed to algorithmic compositions. His use of HPC is due to the intense computing power and parallel processing required to compute the massive number of oscillators for additive sound synthesis, granular synthesis, and stochastic processes to achieve coherent structure from sound grains to musical form. Note that a high level of computing power brings musical interaction to confront the vast number of instructions required to generate 20,000 samples per second for each monophonic sound source, in addition to instructions for rendering the musical events from notes to rhythms, from voices to harmonies, from phrases to sections and large forms.

Due to the experimental conditions interacting with HPC, for composers like Tipei, there is no comfort zone for the habit of falling back to established styles and accepted aesthetics. Interacting with computers to create algorithmic compositions forces composers to focus on a creative attention to inquiries, not to styles. Tipei et al. represents a long history of working towards more sustainable ways of managing the compositional processes, in terms of generating massive instruction sets without relinquishing creative control, also streamlining the time it takes, modularizing computing resources as reusable assets, and being able to work in teams. The result is an HPC collaborative platform applied to music. Due to the multiple skill requirements working with HPC, the collaboration is an important feature of the platform.

The article presents two main topics: (1) the platform called DISSCO (Digital Instrument for Sound Synthesis and Composition) which runs on the Comet supercomputer: the authors describe the components and technical details of implementation, and (2) the collaboration: the authors describe teamwork and collaboration management. Often, sound synthesis and musical forms are processed using different tools, incorporating either structured random functions or deterministic means. DISSCO is an integrated system that handles both sound synthesis and compositional structure. The latter is “... the implementation of an acyclic directed graph, a rooted tree, whose vertices or nodes represent “Events” at different structural levels.” Then the events from DISSCO need to be translated into the sequentially playable sound events, and the authors state, “... facilitating this translation for multiple users is what makes this platform implementation very different than organizing an online group of musicians with physical instruments.” Three modules constitute the architecture of DISSCO: a library of sound synthesis instruments, a composition module, and a Graphic User Interface (GUI). DISSCO runs nearly in real-time while affording multiple users working concurrently. Significant contributions of this paper include a benchmarking and sorting solution with optimal window size to solve the problem between parallel computing and the serial (temporal) nature of sounds.

Here the project investigates a fundamental affordance of HPC. The design tasks involve asynchronous parallelization that is required to return signal outputs applying timely 'hold and release' for synchronization for musical requirements.

A further contribution is the explication of collaboration in composition necessitated by skill requirements, which touches upon another dimension of interaction dynamics for compositional activities. Studies of collaborative composition in cases, articulations, and knowledge are rare and beneficial for musical interaction research. A human-to-human interaction demands behavioral adaptation to sustain a collaboration until a mutually satisfactory musical output is achieved. While these topics are not fully developed in this article, it serves as an overture for widening a scope of discourse over decision-making processes. By using a platform like DISSCO, we can observe more systematically why and how choices of computing techniques are made and how those choices are based on musical intents. It also involves optimizing computational resources (parallel processes and multicore distribution) to support the complexity of instruction sets for sound synthesis, which requires high-definition signal processing to produce detailed musical qualities. Tipei et al. increases the relevance of musical practice migrating from the experimental field of computer music to contemporary technological practices, serving as a conduit to deepen collaborative insight, to spin an evolving aesthetics with respect to computing intervention and to question the values musical practice may offer.

4.2. Promoting Contemplative Culture through Media Arts, by J. Wu

Jiayue Wu's article [78] is an invitation to view the landscape of the author's creative practice engaging multimodal technologies. It can be considered as an individual practitioner's autoethnographic case study how MTI can facilitate what can be described as an experience transfer, through technological appropriation for a cultural practice, in Wu's case by transferring sensorial resources from the Tibetan spiritual practice. The article compiles three creative outputs produced over several years under a common theme of contemplative cultural practice. Presented as case studies, each project is elaborated by the aims, artistic goals, techniques, process descriptions, collaboration, and results. Due to the complexity of combined technologies and ethno-musical dimensions, the three projects involved various collaborators with a common goal as the author states, "... to address the questions of how media arts technology and new artistic expressions can expand the human repertoire, and how to promote underrepresented culture and cross-cultural communication through these new expressions."

The first case study presents the multimedia performance piece, *The Virtual Mandala*. The interactive piece utilizes both live voice and electronic instruments, particle simulation for visualising the sand-like formation of Mandala, motion tracking, physical to virtual space mapping, MIDI activated real-time control input to sound synthesis module and interactive 3D object files for creating atmospheric ambience. The form follows the traditional Mandala process of construction, climax, and deconstruction. The second case study presents *Tibetan Singing Prayer Wheel*, an "haptic-audio system" engaging a physical controller for a multimedia experience emulating the circular motion inputs to a singing bowl. Three input channels pass signals from the voice, gestures from the prayer wheel, and a set of trigger onsets to activate a virtual singing bowl, voice processing, and synthesis modules. The result is a sound installation where an audience can ease into an interactive exploration leveraging the interface shapes and gesture inputs, intuitively mapped into the resulting visual and sound experiences. The third case study presents *Resonance of the Hearts*, which utilizes a pattern recognition system for a set of hand forms called "Mudra" which are used as control gestures to trigger and manipulate sounds. For continuous gesture to continuous sound generation mapping, a special fractal rendering technique is appropriated in real-time, with machine learning technique to anticipate the next state of hand trajectories while recognising the current state and sensors to accommodate the barehand mobility to allow the flexibility to convey "the beauty of the ancient form"

in unobtrusive ways. The result is multifaceted as it was used for teaching and learning in classroom situations, real-time performance and an interactive installation.

Wu's work touches on the enabling dimensions of integrated design and music as well as second-order feedback. These are discussed in informal ways describing technical challenges. Her work appeals to general audiences for its artistic presentation of alternative cultural experience with easy-to-engage interface experiences. Through the three case studies, the author states that her goal is to create "Embodied Sonic Meditation". In terms of scholarly presentation, the writing style of the article combines the style of artist statements regarding the source of inspirations, artistic goals, an aesthetic motivation, and the style of describing the work that was completed, the dissemination channel, and the audience responses. The inclusion of Wu's article in this issue provides for a bidirectional discourse between an authentic voice of a creative practitioner and a scientific norm of a scholarly assessment. Rather than formal user studies, Wu adopts artist-centered subjective observations and informal descriptions of audience responses. The paper eschews technical details and methodologies, focusing on personal discoveries rather than foundational relationships to prior art.

As a final remark, it is worthwhile to discuss a creative impulse and motivation with respect to the wisdom and anticipated cost working with technologies. Wu states, "... from these case studies, I also discovered that sometimes even a cutting-edge technology may not achieve the original goal that an artist planned." Often a cutting-edge technology presents more challenges than solutions it offers, and this is what mobilizes the creative impulse for those experimental composers and artists who address challenges more than opportunities and niches. Perhaps, have they known from the very beginning what may come out of the creative tunnel and the costs through the process, some creative outputs might have never seen the light of day. Then often, those practitioners engage the challenges one after another, not because what came out of the tunnel from the previous project was recognized and rewarded by others, but because what came out of the tunnel not only differs in its kinds of reward but also for its further possibilities. How a discourse unfolds between and within the process of creation and the culture that hosts it, is an open question. To begin with, creative practitioners need to be critical about their own artistic goals: what are they pursuing, an aesthetic effect or a creative cause?

4.3. Comprehensive Framework for Describing Interactive Sound Installations: Highlighting Trends through a Systematic Review, by V. Fraisse, M. Wanderley, and C. Guastavino

In terms of a creative engagement of sounds, we can consider two broad modes: sound as an exclusive medium and sound as a primary medium alongside other modalities. In Western European tradition, the former is more familiar to the contemporary audience who will likely turn on a music channel or go to concerts to hear music. For the latter, human societies have been always engaging sounds in various activities such as farming, rituals, hunting and social play. However, in the Western European tradition, the origin of the latter can be traced to the early 20th Century Dada and Avant Garde movements as a manifestation of breaking out of the concert tradition. This backdrop is foregrounded to suggest that sitting in a concert hall to listen to sounds offers less certain engagement compared to actively doing something with or along with sounds or making sounds. Concert-goers may immerse themselves in hours and hours watching how kinesi-
thetic patterns and coordination among orchestra members play out in time and how such interactions constitute sounds. At the same time, we should not overlook the consideration that active and participatory ways of engaging sounds may be more natural for people. The last two decades saw prolific practices of sound as a primary medium under the label of "public art", "interactive sounds", "sound installation" or "site-specific audio art". Those art forms often deploy sensors with varying degree of intelligence to encourage and process audience's participatory behaviors. Given the prevalence of sound installations, there have been relatively few systemic inquiries on how these practices come together, what technologies and methodologies are used, what the artists think they are doing and

whether their intentions resonate with audiences. Previous research provides useful contributions mostly focusing on interfaces, yet often these accounts are anecdotal without structural analyses linking systems and aesthetic outcomes. In that regard, this article is highly relevant for this issue because it is one of the first aiming at an overarching framework, inclusive of prior works with specificity, that is extensible for describing musical interaction.

Fraisse et al. [79] explore systematic approaches for describing interactive sound installation, regardless of the purpose either engineering or artistic. The researchers used literature review methodology adopting the curatorial protocol called PRISMA, and indirectly investigated 195 interactive sound installations for extracting descriptors from 181 publications where the installations were discussed. Clearly this methodology will result in exclusion of all sound installation works that were not discussed in the 181 sources. However, the authors are very clear about inclusion and exclusion criteria, search processes, and further curatorial processes involving qualitative assessment as well as manual coding. There are also consequences that more relevant work in the Scopus database would have been excluded from the search process, simply because the publications lacked the search terms. There are also several merits: First, the corpus is a collection of publications that went through peer review where obscure terms and jargon have been scrutinised. Second, by limiting the corpus to research-oriented documents, the process is likely less overburdened by artists' statements that have different communication protocols and goals. Third, since the methodology is relatively transparent, the limitations and omissions are clear; what is omitted and why are obvious, not a shortcoming of the methodology nor of the authors' research design. If any, the last point speaks to how and why works of art will benefit by being informed of and aligned with a kind of literacy, as compiled by Fraisse's team as an example, encouraging artists' responsibility of authenticating their descriptions.

The result of the literature survey is a taxonomy with maximum four layers in the hierarchy across 111 taxa. The root level consists of three nodes: artistic intention, interaction, and system design, which are described as the three complementary perspectives. Nodes in the middle and leaf layers are organized according to the three perspectives and constitute a conceptual framework for describing interactive installations. The authors present the resulting taxonomy as the proposed framework, which can be considered more than a taxonomy because it is framed to enable further insights, encouraging readers to explore an interactive data visualization on their website. For exploration and interpretation of the data, one should always bear in mind that the data is corpus bound, not the representation of the sound installation practice at large, which is noted by the authors. The organization of the article is effective and includes peripheral data such as bibliometrics showing a stiff rise of publications between 2000 and 2006, which may indicate the increased accessibility of technologies and prototyping opportunities with integrated circuits, sensors and actuators and LAN bitrates. Other informative data includes the landscape of research fields around the topic with diverse focus and motivation: music and computer science applications are equal top contributors followed by software fields. While expected, this can be interpreted as a concentration concerning implementation of prototypes or artworks exploiting technological and application opportunities, with less focus on user studies and little on explanatory frameworks. In that context, Fraisse et al. is a timely contribution, especially with its systematic method eliciting a set of clean terminology. Implicit in their framework is an ecology of the musical interaction research community, various constituents in interdisciplinarity, and diverse profiles in terms of project motivations. Throughout the article, authors meticulously present their findings, and it is worthwhile to stress that some findings are valuable indicators of limitations and challenges directed to practitioners, primarily in how they purpose the documentation about their creative practices, and secondarily in how they choose the descriptors compatible to the semantics they wish to associate to their creative practices.

Despite or perhaps because of simplicity, the framework's three organizing perspectives are both complementary and comprehensive. It seems unnecessary to arrive at additions or alternative perspectives even with a larger corpus of literature and bespoke vocabularies. An adaptive framework may be considered by increasing the affordance of the framework to avoid a closed reinforcement cycle between the inclusion criteria and the existing taxa. At the same time, it may be challenging to align the arts- and humanities-oriented language with physical science and engineering oriented language. Perhaps, a clearly defined scope of language may be a choice that needs to be respected.

4.4. Representations, Affordances, and Interactive Systems, by R. Rowe

Robert Rowe's article [55] provides a deep perspective on the relationship between system and musical practice, through the concept of representation and how it effects musical activities from conceiving an interactive system to composing and performing. Underlying Rowe's perspective is the conjecture: creative outputs are not separable from the systems with which they are produced. While this may be obvious for some readers, in music practice, systems cannot be taken for granted as means and tools for some unarticulated higher priority that one desires, or for a mere pursuit of a musical effect to evoke an affective state, without closely examining the compatibility between the musical information one wishes to encode and the system to encode it. The article ends with an inspirational note indicating that, by thinking through "the issues of representation, abstraction, and computation", artists are positioned to make a central contribution now (more than ever before), for "Artificial intelligence has great utility and has made rapid progress, but still has a long way to go."

For discussing symbolic and sub-symbolic representation, Rowe presents music notation (from the Common Practice Period) as an example of symbolic representation, as compared to raw samples from audio recording as an example of sub-symbolic representation. The discussion of MIDI is nuanced considering how much it dominated the computer music community with its "standardized representation" as a communication protocol between keyboard and sound synthesizer. For electro-acoustic music, a progenitor of MIDI, the relationship between control and output signals is at the heart of music creation. Before MIDI, readers may imagine there were two primary ways of coding the relationship between control signals and an output signal: either electrical, through wiring complex patterns of patch cables in analog synthesizers such as Buchla or Moog; or digital, through classical software programming languages (such as Fortran or C). Compared to these precedents, MIDI offered an efficient and convenient way of setting the relationship between control signals and musical output signals. With the use of MIDI however, Rowe systematically draws an implication that users in exchange for convenience may unwittingly commit to accepting the hidden layers of processing with little control. For example, consider the hidden processing where the system registers time stamps from MIDI messages and organizes them into Western style musical time units. This resonates with the discourse between French and Italian schools of the 13th and 14th century, over encoding the medieval rhythmic modes into the system of notation (Appendix A). Rowe is referring to the age of MIDI when he states, "The wild success and proliferation of the MIDI standard engendered an explosion of applications and systems that are still based today on a 35-year-old conception of music", but the Western keyboard paradigm that informs the discrete control and symbolic levels of representation in the MIDI protocol, dates far earlier than the Common Practice Period.

As Rowe critically examines, the determination of "what information is sufficient to encode" hinges on a system of representation. Young composers in our digital age will benefit by deeply internalizing this reality; although the acquisition of musical expertise requires guidance and support with well-established traditions, there is no obvious tradition for musical tasks engaging modern technologies. As discussed earlier, the tradition of common practice symbolizes the long history of evolution from practice to theory, then in turn the theory provides prescriptive functions for the practice, and this cycle continues

until the system is exhaustively exploited. Better or not, in the absence of such tradition with the use of computation in the late 20th century, we are in a complex landscape of many systems and musical practices taking idiosyncratic paths, where no coherent ecology can be seen other than the exemplary traces of thought and activity of composers like Koenig, Hiller, Xenakis, and Eno as described in Rowe's article. In this regard, the article carries us further by introducing the current challenges in Artificial Neural Networks (ANNs), their black-box quality and difficulties in representations of multimodal signals. Despite ANN's sub-symbolic capacity encoding training sets, their internal representation of learning process as weight functions is far from grounded in ways we understand how knowledge formation may occur. Certainly, this concern is shared by the explainable AI community to improve the transparency of the internal processes of black boxes and explanation of deep network representation [80–82]. With Rowe's train of thought and articulation, it is logical that he arrives at Gibson's ecological perception and the concept of affordance. Rowe presents affordance in relation to accessibility of system control, as he states "... the exposed control parameters present an explicit set of affordances". This makes sense if an affordance is perceivable by artists as to expose its representation of available spaces for making artistic choices, and if that explicitness in the set of affordances is engineered by a system designer through a system of representation. Therefore, which one comes first, affordance or representation, is subject to further thoughts.

Rowe's article raises three considerations, imperative for future discourse on musical interaction: mapping, navigation, and representation. Referring to user control of computer-generated sound, Rowe presents the difference between mapping and navigation, stating "The difference appears as a change in orientation toward the underlying representations: mapping creates point-to-point correspondences between input features, or groups of features, and output behaviours. Navigation (or sailing) suggests the exploration of a high-dimensional space of possibilities whose complex interactions will emerge as we move through them." Here, Rowe's use of "mapping" is implicitly aligned with a specific use in the computer music community. Rowe cites the work of Chadabe [83] who introduces a focused use of "mapping" to refer to the creation of a type of audible relationship between control signal and sound. Musical output is sometimes criticised when a control mapping produces an invariant audible signature in the musical flow. Whereas the general use of "mapping" is synonymous to implementing a transfer function for scaling a control signal to a range of synthesis parameter values. It is unlikely that a navigation paradigm can be implemented without the use of a general transfer function. Mapping in this sense is a useful technique for defining choices and constraints to discover emerging control spaces in a continuous and multi-dimensional exploration. The general concept describes a necessary condition for information encoding and resolution capacity.

In terms of navigation, if navigation also engages a mode of exploration, the very nature of exploration does not "aimlessly circle through undifferentiated choices" because an exploration is inherently based on a what-if scenario. Human perception and cognition can only defer aimlessness in attention even at the level of simple awareness. To resist limitations imposed on compositional orientation is understandable, but the general function of mapping is also exploration-enabling.

For representation, we may further investigate the relationship and the order of representation and affordance as Rowe has put forward. Recent neuroscientific findings indicate that a genome does not encode representation or strategies based on ANN-type optimization principles. Genomes encode wiring rules and patterns from which instances of behaviors and representations are generated [84]. This ties well with Gibson's theory of senses as active and outreaching perceptual systems to acquire perceptual information about the world (see Section 3 above: "to detect something" rather than "to have a sensation"). As our perceptual systems are highly interrelated and their activity level coordinates movement and sensing, this is an area we need to examine further on two sides; what affordance musical interaction may bring to MTI and what affordance MTI may bring to musical interaction, in ways the two sides mutually expand the margins. Perhaps,

this activity contributes to the plasticity of neuronal wiring, which creates further affordances or alternative ones, and this should inspire the future perspectives of musical interaction involving AI.

In sum, Rowe's perspective article provides both critical insight and new orientation for young generations of composers, with highly informed interdisciplinary concepts. Deeply committed to the future of interactive systems informed by computer music literacy, Rowe brings interdisciplinary connections to refresh the foundational inquiries for musical information encoding and representation. The above discussion is much in debt to the authenticity of Rowe's article and the maturity of his enduring creative practices.

4.5. What Early User Involvement Could Look Like—Developing Technology Applications for Piano Teaching and Learning, by T. Bobbe, L. Oppici, L.-M. Lüneburg, O. Münzberg, S.-C. Li, S. Narciss, K.-H. Simon, J. Krzywinski, and E. Muschter

Bobbe et al. [85] propose an application of Tactile Internet with Human-in-the-Loop (TaHIL) to piano lessons and present their studies exploring online teaching and learning scenarios. One can argue that the concept of TaHIL [86,87] is not new considering many elements housed in that idea have been around for a long time, especially from the era of telepresence and teleoperation [88,89]. However, as this article is titled, the authors take the first step in testing TaHIL-inspired scenarios with potential end users. This research emphasizes the value of early user involvement to ensure the users' participation in decision making, when the definition and scope of design tasks are "fuzzy" at front end of a design process. While this kind of approach has been long practiced in Human Centered Design (HCD) and Participatory Design, the methodology is still loosely defined, examples are rare involving nuanced control of musical instruments and instruction, and the process is difficult to execute due to the different social skills required for the researchers.

The position of this article is informative to foreground by examining two challenges for this kind of project: origin of the conception and mental model incompatibility. First, the conception meets its application area, which is music. The contemporary music field has been historically active in pioneering and inventing various technologies such as digital instruments, music controllers, composition and notation systems, performance analysis, real-time performance system with machine learning and score following, and multimodal representation of feedback signals, to mention a few. In sum, musicians are adventurous. However, the concept of design is not well aligned with the concept of music, in part due to differences of professional culture. Traditionally musical instruments evolved with their virtuosos, and this is not the case with digital technologies. Digital technologies identify end users as novices in first conception. Often the music technologies' inventors are users themselves, so their inventions mature along an individual's use and testing trajectories, which can be idiosyncratic. A well appreciated playful device or production application may attract small communities of users. It is often the case that new devices applied to musical applications go through a time-consuming design process involving end users before implementing a first prototype. As the authors assert, many technological inventions in music do not make their ways to end users' hands.

Second, end users' mental models can be deeply ingrained and generate reluctance and skepticism if not hostility when meeting new propositions of potential use cases of technologies. Piano lessons, the authors' application area, are considered a time-honored highly personal teacher-student relationship requiring intimate observations and assessments through physical execution of performance. Equally important as interpreting musical scores are certain postures and forces interacting with the musical instruments for producing proper intonations and articulatory gestures in resulting sounds. Any technology intervention applied to piano lessons, especially online and remote interaction can be considered highly disruptive and depersonalizing, therefore incompatible with users' familiar mental models. Nonetheless, further investigations are relevant especially in the context of the COVID-19 pandemic, where educators and conservatories have struggled to sustain music programs using distance learning technologies. Here we may re-examine

a fundamental requirement in musical interaction: managing one's own body movement, the second-order feedback cycle sensitivity extended from the first-order feedback to the instruments in the circuit of music performance. Musical interaction includes a performer monitoring body movements interacting with an instrument body and its mechanism and resonating behaviors, while monitoring the arriving sound, involving microsecond intervals of overlapping action perception feedback cycles.

With these challenges, this paper's contributions are timely. It is certainly plausible to undertake a design process starting from scenario testing considering the user community as described above. The authors' scenarios may appear technologically naïve lacking some technical details. Setting aside technical plausibility, they were able to gather very useful comparative datasets by recruiting the participants both from learners and teachers. Many statements in their ethnographic notes may sound expected, nonetheless, are most applicable when recorded as data. The four scenarios presented are well touched in terms of the core issues of musical interaction required for playing piano, and consistent with second-order feedback learning. The first scenario tests offering visual assistance by capturing body movement focused on upper limbs and torso by means of wearables as well as finger movement using a data glove. Their data confirms visual feedback as distraction but yields insights to employing personalized machine learning techniques applied to tactile feedback for performers to prevent injuries caused by bad postures and movements. The second scenario tests whether performance analysis offered along with the musical score can be useful for piano practice. The study shows that, since this feedback accompanies a musical score, it is less visually distracting, and it may help students' self-practice as long as it allows variations of daily practices and does not interfere with diverse interpretations of the music. The ethnographic notes also show teachers' interests in monitoring students' self-practice to tailor assignments and feedback. The third scenario tests attention assistance, and possibly stage fright alleviation by means of music visualization with selected parameters to provide visual feedback in addition to aural. The results indicate some enthusiasm deviating from the original intent for the possibility of motivating students to practice. The fourth scenario tests the idea of instrumented gloves worn by both student and teacher to exchange tactile and haptic information. to supplement the missing intimacy during online lessons. Their report shows split responses with a possible use case applied to beginners.

While this study can be considered as a pilot phase, it is a significant and important step due to the challenges described above. Readers are encouraged to survey the narratives in their scenarios and resulting data with a depth of implications. This work represents an essential design step that is not always practiced in the development of musical interaction systems and opens the doors to further problem identification in terms of technological offering as well as deeper insights to the user community.

4.6. Musical Control Gestures in Mobile Handheld Devices: Design Guidelines Informed by Daily User Experience, by A. Clément, L. Moreira, M. Rosa, and G. Bernardes

Clementi et al.'s project [90] can be best described as data informed design practice anchored on users' daily interface experiences. The paper is a significant contribution to musical interaction design literacy for Mobile Handheld Devices (MHD). The authors present experimental results from observing how participants use MHD for controlling music stimuli. Two groups of participants were engaged, musicians and non-musicians, both with one common profile of being everyday MHD users. The research design combines qualitative and quantitative methodologies to capture and study the participants' task-oriented data, which may lead to design insights by inquiring what kind of interaction patterns emerge as a central tendency of intuitive use patterns, purely based on their experiences from using MHD. The authors state the aim of the article is "to provide the first steps towards defining guidelines for optimal sonic interaction design practices in mobile music applications". Readers will appreciate the trajectory and processes taken with "the

first steps” and how the authors elaborate the experimental design, tooling, procedure, and results.

One may expect to find a body of literature and prior art when pursuing research for mobile devices, but the authors encountered the situation where they had to invent and implement an experimental system by filling gaps in existing methodologies. This is a case as discussed in Section 3 regarding the shortfall of coherent methodologies in combined music and design practice. Here the authors’ teamwork merits high originality in the research landscape of musical interaction. For streamlining data coherently from capture to analysis and evaluation, the researchers required a protocol which they had to invent for their experiential setup. The authors propose the protocol for analysing and validating parameter mapping between music and gestures in MHD applications and instrumentation. The protocol is a significant contribution which other researchers can use and tailor for their own purposes.

For eliciting how people adapt familiar interaction patterns from their daily use of mobile phones to reproduce sound stimuli, experimental tasks were designed to reproduce music stimuli by controlling musical parameters. The device operation to execute tasks leverages MDH’s built-in capacity for detecting touch screen finger movement trajectories and device orientation using six degrees of freedom. Data acquisition was divided into two phases for comparative data analysis but from participants’ point of view, the experimental setup was divided into three progressive stages moving from naïve execution, to reflection, to informed task execution. This is an excellent strategy akin to the typical rehearsal scenario in music practice. The phased elicitation method was intended to acquire observational data regarding participants’ tendency to act intuitively for executing the task, first with no prior instruction, then to reflect upon their own performances, and repeat the task so that their performance rationale will be elicited with questionnaire-based verbal descriptions. Five musical stimuli were presented, and for each stimulus, the participants were asked to reproduce them so that the system could collect the data of how they were associating musical parameters with device operational parameters for determining possible trends in exploiting device affordances.

The result shows the differences in operational mapping trends with musical parameters; clear trends were found for controlling pitch and duration and less obvious for amplitude. There are also notable differences between musicians’ and non-musicians’ overall performances and in the comparative performances in two stages in terms of changing performance behaviors. From this, one can infer a strong learnability for both groups of participants but with a varying degree of parameter-aware performances between the two groups, indicating the prior musical skills present either desirable trends or bias. For interpreting the data, the authors are thoughtful to note that some natural gravitations, meaning tendencies in mapping behaviors indicate culture specific influences. The authors specifically point out the pervasiveness of the mental model of piano in Western European music reflected in the user interface representation of digital musical instruments and controllers. This opens future research opportunities for user studies in non-Western European cultures, especially for the ethnic groups who are not biased by the Common Practice Period tradition.

In addition to the data presented in the main body of the paper, the appendices include 29 tables of data as well as the experimental questionnaire for other researchers. The collection of data present varying degrees of ambiguity and difficulties to distil conclusiveness in certain trends, partly due to the necessity to present the stimuli in parametrically distinctive ways and the idiosyncratic nature of music stimuli design. Protocols for validating the design of music stimuli are not widely recognized, beyond rationales for simplicity and systemic variation for parameterized presentations, to heuristically target the best possible combinations. Nonetheless, these data are rich resources and subject to different interpretations, therefore they are sources for drawing further design insights. The immediate future is promising by enabling automated reproduction of gestures using the captured interaction data from tracking the user and event ID, so that further

dimensions can be studied by correlating a participant's performance data and their verbally expressed rationale. The design guidance at this phase of the research confirms our common knowledge regarding note onset, duration, and pitch mapping associations with a device's operational parameters, and extends to interesting discussions on amplitude mapping and future possibilities contingent on technological advances. Since the project leverages an affordance of MHD's self-contained sensors and actuators as well as the computing power equivalent to personal computers, the article has a broad impact towards a possible scalability for enabling musical interactions for a larger audience.

4.7. The Power of Gaze in Music. Leonard Bernstein's Conducting Eyes, by I. Poggi, L. Ranieri, Y. Leone, and A. Ansani

Poggi et al.'s research [91] is inspired by an extraordinary question: "...for the gaze communication system we use in everyday life, it is possible to write down a lexicon and an "optology", therefore, why not write down a lexicon of the conductor's gaze?" Such "why not" leads to the conception of experimental studies reported in this article. As discussed in Appendix A, a major legacy of common practice in Western music is the formation of an orchestra, entitling the emblematic position of a conductor as a symbolic figure whose tiniest gesture could not be ignored. However, the main function of a conductor rests on pragmatic labour for shaping musical expressions by ways of coordinating the large-scale ensemble. Historically, an authentic signature of a conductor can be recognized by sounds alone, and skilled listeners can reliably differentiate one signature from another because of the different traces in rendering tempo, handling transitions, guiding the harmonic progressions, timber shaping, expressive loudness differentiation over instrumental groups and particular emphasis of articulatory gestures on musical phrases. One of the fundamental techniques is the line of connections (of the baton), which refer to a set of choreographic patterns for arm movements, mostly the right arm, corresponding to the meters and beat patterns specified in compositions. Beyond meters and beats, Maestros deploy this technique as an interplay of time and space, not only for managing their own embodied space but also for coordinating spatially distributed players in instrumental groups, drawing differentiated musical expressions and synchronizations.

While conductors' gestures can be considered as non-tangible actions, the gestures of conducting can be perceivable through the ensembles' corresponding sounds, therefore gestures are physically tangible as a medium of expression. This aspect is called out upfront in the beginning of the article "... the ways in which musicians make music in an ensemble is influenced by its participants and, if there is one, by the conductor's body", and this is what makes the topic highly relevant for musical interaction research. Conductors' movements have been attaining various research interests with the increasing availability of motion analysis technologies. In the field of conducting, there are abundant teaching and learning textbooks and technical handbooks. These are excellent sources for MTI researchers, providing many technical details on multilevel engagement of a conductor for managing social and situational dynamics of performances, in addition to mastering their own conducting techniques. Involving technologies for studying conductors' gestures, the article offers an excellent body of citations that are relevant to MTI researchers covering a range of foci including movement patterns, facial expressions, semantics of body parts associated to musical instructions, gesture lexicons, possible common features across conductors, intent, communicative or musical effects and other associated topics.

One of the main challenges in studying conductors' gestures is the dynamically situated nature such that quantitative data acquisition is difficult, due to the complexity of the phenomena that resists a common framework for parametrization coupled with musical outcomes. For this reason, the authors' research design combines three studies: (1) ethnographic interviews engaging five choir conductors, (2) qualitative coding and analysis on the video corpus capturing the gaze performance of one conductor and (3) focus groups for testing and evaluating a comprehensibility of gaze functions in an experimental set up with stimuli clips from the video corpus. The first study shows mixed

results, and it is difficult to draw systematic interpretations due to several factors: the study was rather informal, there are no supporting materials to yield details on interview data, and the sample size was small. However, a set of nine questions were well composed and presented to the study participants, progressing from eliciting individual traits with respect to musical parameters, towards finding out what awareness the participants attribute to eye gaze for conducting. The overall conclusion is drawn that choir conductors are not deeply aware of the potential of eye gaze even though there is consensus on the eye gaze functions such as calling attention or giving feedback. The second study is based on a bold eye gaze performance of Leonard Bernstein captured on video. Coding and analysis with annotation applied the eye gaze lexicon from the previous study. The result includes the highly informative distribution of gaze functions ([91] Table 4), which provided a basis for testing and evaluation in the third perception study. In terms of a corpus, the choice of a deeper qualitative analysis on one sample is justifiable because the synergy between a conductor and orchestra members reflects highly adapted and learned communicative behaviors tailored for that specific relationship. Involving multiple conductors' samples will require different observational frameworks to account for the shifting dynamics of the ensemble members' interactions. The primary original contribution of the article comes with the third study where systematic experimentation engaged the sample size of 186 participants composed of three profiles, non-expert, expert, and amateur. For the perception study of gaze, the participants are presented with three classes of stimuli conveying three different gaze functions, "Start", "Pay Attention" and "Crescendo and Acceleration", each functioning in two modes, video only and audio-visual. The stimuli were extracted from the video corpus of Bernstein's eye gaze conducting. The study reports comparative data analyses and interpretations on (1) how the mode of presentation effects comprehensibility, and (2) what kind of different data profiles can be seen with respect to the three participants' profiles, in terms of the degree of recognition of the three gaze function stimuli. The details of comparative data based on these variables are potentially useful resources for future researchers to confirm results or to present counter cases.

The overall interpretation of the data is that the communicative function of eye gaze is comprehensible across the participants, despite variations in gaze recognition effected by their level of expertise, mode of stimuli presentation, and the attributed meanings. It seems premature at this stage to determine whether this confirmation is indicative of possibilities of a systematic and sharable lexicon across all types of conducting. There is a notable discrepancy in trends when comparing the interview responses from choir conductors with the qualitative analysis on Bernstein's gaze performance. As the article reports, the choir conductors show relatively low investment in the function of gaze while Bernstein evidently demonstrates diverse gaze functions for conducting an orchestra. The discrepancy may be suggestive of the data contingency to the character of an ensemble, e.g., choir conducting vs. orchestra conducting, specifically the contingency to the fact that, the two ensembles utilize two different spatialization profiles for positioning members associated with the two different musically functional distributions. This may well influence the repertoire of gestures or eye gaze as well as the degree of awareness or deliberation on the use of gaze. While the article achieves a comparative synthesis of the differences and similarities, this spatiotemporal functional engagement of gaze in musical interaction is an open door to future studies.

This research is pertinent to the topic of multimodal perception and cognition at large, specifically investigating the role and effects of a conductor's eye gaze. Leveraging Poggi's previous research on lexicons and effects of conductors' gestures, this article extends the research repertoire by investigating the comprehensibility of gaze function through perception study. The art of conducting can be defined as an expressive management of an intricate circuit of musical interactions embedded in a social interaction with an ensemble. In this context, this article has a significant place for making a unique contribution to knowledge of gaze functions and the chain of effects, from the domain of individuals' perception to social cognition in daily communications, through explorations

of how gaze can be a constituent in multimodal communication for musical interaction and what impacts it may have on musical processing.

4.8. FeelMusic: Enriching Our Emotive Experience of Music through Audio-Tactile Mappings, by A. Haynes, J. Lawry, C. Kent, and J. Rossiter

Haynes et al. [92] propose a sensory augmentation possibility of musical experience with the sense of touch. The project FeelMusic involves an experimental setup designed to investigate an affective correlation between the two modalities, tactile pressure, and sound. The authors state "... to project the emotive elements of music into the haptic domain", their research design also explores an affordance of tactile domain for musical experience. The setup allows applying tactile stimulation, auditory stimulation, and combined stimulation to participants, then acquiring self-report data from a user interface engineered to collect the affective responses. For tactile stimulation, the researchers implemented the device called "Pump-and-Vibe", a wearable tactile interface for the upper and lower arm. The wearable ensures the signal input pathway for receiving auditory stimuli carrying music information is properly engineered for controlling pressure modalities. The paper presents the experimental design, methodologies for system implementation, stimuli design, data acquisition, and the experimental results with comparative data analysis.

While the authors state their research purpose is to enhance musical experience with tactile sensation, the paper also contributes to understanding the integrated affective responses given multimodal presentation of touch and music, contributing to affective neuroscience and crossmodal perception research at large. In terms of methodology, a significant contribution comes with the presentation of their research ecosystem: how each component was engineered, the conceptual framework and models, consideration of usability and sensibility, temporal resolution, and the aspiration to design the stimuli through multimodal parameter mapping. As an early phase of research there is room for improvement, yet FeelMusic is an exemplar of creative research involving engineering practice for demonstrating feasibility of propositions with working systems, wherein the system prototypes and wholistic design of information/signal pathways are required to inform an ambitious research agenda.

Unlike a project with an incremental research agenda, which carries its own merits, a project like FeelMusic entails entirely different ways of contributing to the field of musical interactions and opens the doors to a wealth of research questions and curiosity, which are very much welcomed in this special issue. The authors take great responsibility balancing their novel approach with literature review and references to prior art, which is critical for venturing with new ways of investigation. Novelties are sometimes exercised by "just doing and reporting" where often there is no risk for doing so, and little consequence. Novelties are other times exercised by necessity and there are always risks of shortcomings, nonetheless, such shortcomings always merit further inquiry. FeelMusic belongs to the latter by inspiring questions in a thoughtful reading with specific directions rather than vague implications.

To anchor their novel approach, the FeelMusic researchers employed well-established and proven models, Hevner's adjective circle, and the circumplex model of emotion with valence and arousal dimensions. Circumplex model is known to be consistent with recent findings that affective states involve cognitive mechanisms for integrating and interpreting neural senses from lower mechanisms [93]. Hevner's adjective circle has been widely adapted in psychological experiments for studying various modalities as well as in fields such as AI and robotics. It is noteworthy that Hevner's adjective circle was originally conceived as eight clusters of emotional descriptors for music, targeting Western common practice music in particular [94]. In reference to the Common Practice Period in music as discussed in Appendix A, it is notable that the precursor of the Hevner's cycle was the adjective checklist that Hevner used to investigate whether listeners can really perceive the historically affirmed characteristics of music, as embodied in major and

minor modes of common practice harmony, and if listeners perceive this, to what extent training plays a role for them to recognize the characteristics. Many of Hevner's experimental results are still largely upheld, especially pitch and rhythm as primary musical attributes consistently checked in her adjective clusters. FeelMusic researchers also chose to utilize pitch and rhythm in their experiments. The music they chose for auditory and tactile stimulation also belongs to the Western common practice tradition.

Here are two examples of the questions ignited by the FeelMusic presentation with directed specificity. One outstanding question is whether harmony can be represented as touch and properly projected through tactile stimulation. The question relates to Hevner's conclusion that melody (the contour of a single sound source producing a pitch series variation) is less important than harmony (a complex timbre with multiple concurrent recognized pitches) for conveying emotional states to affect responses in the adjective circle. The other question is more complex. It is prompted by the authors' choice for designing tactile stimuli using music parameter mapping rather than the simple frequency of sound to convey musical information. This opens the door to many challenging inquiries. As reflected thoughtfully in their stimuli design, music is not a simple matter of triggering a "ping" that can be translated into equivalent information by applying pressure to the skin through a vibrotactile pathway. The main challenge for this kind of investigation has to do with the potential incompatibility between arousal response of cutaneous tactile sensation by end organs through skin stimulation, and the resolution in audible frequency range for perceivable differentiation with sound modality. Specifically, musical information is ambiguous because in addition to hearing it may also be perceived by proprioceptive mechanisms, as listeners' musical experience may not be independent from their kinesthetic sense, which is suggested by cross modal binding especially in lip reading research [95]. In this sense, much can be said regarding the many opportunities for FeelMusic in terms of further advances, especially in stimuli design and interpreting data.

In sum, FeelMusic demonstrates a great potential for developing sensory augmentation devices to cultivate embodied musical experience as well as investigating the affective quality of vibro-tactile stimulation itself. It is highly relevant for musical interaction in which emotions are experienced and communicated through cognitive interpretations by identifying the neurophysiological changes (movements or dispositions) in the valence and arousal systems. Regardless of ambiguities in early-stage implementation, the authors advance a critical approach to prototyping a wholistic ecosystem, exemplified in FeelMusic for undertaking novel experimentation towards further knowledge and insights, and for generating greater opportunities for new directions of research.

5. Summary and Conclusions

For summarising this special issue survey with a conceptual framework, Table 1 shows how each dimension in the framework relates to the contributing articles. Most contributing articles' research are in explorative phases, therefore when described as *not discussed* or *not applicable* in Table 1, it does not mean the research is not relevant.

Table 1. Comparative summary of the contributing articles with respect to the proposed framework of five enabling dimensions.

	Affordance & Mental Model	Design Alignment	Adaptive Learning	Second-Order Feedback	Temporal Integration
Tipei et al.	HPC affordance for creative practice	Platform design aligned to musical tasks	Not discussed	Implicit when applied to interactive composition	From synthesis to musical form
Wu	Technology affordance for cultural practice	Tangible User Interface design aligned to music control	Not discussed	Implicit in the discussion of real-time computing	Applicable but not discussed
Fraisse et al.	Affordance of literature survey methodology	Requires extending the framework	Potentially applicable to the framework	Requires extending the framework	Requires extending the framework
Rowe	Affordance defined in representation and exposure to access control	Design aligned to the choice of representation	Potentially related to Artificial Neural Networks	Not discussed	Not discussed
Bobbe et al.	Affordance of online technologies and explores mental model	Exploratory phase but highly applicable	Applicable but not discussed	Exploratory phase but highly applicable	Exploratory phase but highly applicable
Clementi et al.	Affordance of mobile device	Design aligned to gesture-to-sounds	Not discussed	Exploratory phase but highly applicable	Exploratory phase but highly applicable
Poggi et al.	Affordance in gaze applied to musical communication	Not applicable	Not applicable	Implicit in the discussions of delay effects in gaze vs. hands	Not applicable
Haynes et al.	Affordance in tactile perception for musical appreciation	Design aligned in mapping two stimuli cross modalities	Not applicable	Highly applicable and discussed in device implementation	Highly applicable and discussed in system implementation

For example, Adaptive Learning can be highly relevant for future research in Bobbe et al., Clementi et al., and Haynes. An interesting case is Fraisse et al.: Adaptive Learning can be applied to their framework for ingesting new descriptors, organising and configuring adaptable semantic space while other dimensions can be considered for searching if the researchers substantially extend their framework. This suggests, with respect to the five enabling dimensions, that a new space of semantics will emerge with new descriptors from musical interaction research. Second-Order Feedback is highly applicable for all. Rowes's work on interactive systems has been concerned with this dimension as well as with Temporal Integration. The topics are not discussed in his article but implicitly present when he discusses navigation and mapping. Tipei et al. applies temporal integration to synthesize to composition output, not to a system of interaction. Poggi et al. is another case where three dimensions are noted as not applicable, but they will become applicable when the findings contribute to gaze communication system for musical interaction.

Across the board, all researchers are cognizant of the Affordance dimension. It is either implicit or a main purpose in their investigations. Tipei et al. report the affordance of HPC application for music and Wu explores musical experience as a cultural practice with new technological affordance. Fraisse et al. explore a systemic framework affordance for describing interactive sound installations by exploiting the affordance of literature survey methodology. Rowe's perspective explicitly addresses an importance of affordance in musical interaction in terms of control accessibility determined by whether a system exposes symbolic or sub-symbolic levels of control, and related system representation issues. Bobbe et al. explore an affordance of online technologies for music learning and teaching by building scenarios to investigate several mental models by engaging different participants. Clementi et al. investigate an affordance of mobile devices for musical interaction, for which they provide a protocol from their experimental design. They had to implement a temporal integration to synchronise multichannel signal flows for data acquisition. The central focus of Poggi et al. is to explore the affordance of eye gaze as an effective communicative device. Haynes et al. investigate an affordance of tactile modality for sensing musical stimuli through translating the vibratory patterns of music to the tactile domain.

Most researchers are implicitly aware of these dimensions; either they use their own domain specific terms related to one or more dimensions, or they use descriptive sentences effectively describing one or more dimensions. This indicates how a conceptual framework may offer a coherent way of describing research and design approaches to musical interaction, in system implementations as well as assessments.

The present conceptual framework is introduced in response to the formative inquiry of Sections 2 and 3: "What constitutes musical interaction?". Section 2 emphasizes the meaning of both music and interaction will be mutually modified, (1) to account for a potentially broad impact on human experiences with multimodal technologies situated in daily life, and (2) to satisfy emerging requirements with evolving technological capacity. Adjacent research communities are introduced to expand the literature inclusive of neuroscience and therapy. Appendix A accompanies Section 2 to introduce an orientation to a depth of discourse undertaken over centuries in Western European music during the Common Practice Period. This period is an inescapable influence for multiple reasons, among which, most music repertoire played in modern concerts are from that era, and much music stimuli used in current therapy practice or psychological experiment are based on the common practice. More importantly, the story of the Common Practice Period exemplifies a formative process through a contextual adaptation enforced by a rigorous discourse in a community of practice deeply concerned with phenomenological and scientific understanding of music, with respect to human perception and cognition.

For example, equal temperament tuning of common practice exploits the human perceptual affordance for dominant frequency resolution. The 12 notes of the scale are equally tempered within an octave. This achieves an audible compromise between pure and artificial sonority and has become a social norm across diverse instruments, because it affords both keyboards and large ensembles maximum mobility across 24 keys. This history provides an excellent basis to inform a perspective for adopting music as a qualifying criterion in the context of MTL, and devising beyond Western European culture the concept of music for multimodal interaction with modern technologies.

Section 3 presents two foundational conditions of music: (1) the elusiveness of the concept of music in diverse culture and individuals, and how we may address the differences in what it means to be musical across regions where multimodal technologies impact people's daily lives; (2) musical experience as an ephemeral phenomenon of experiential liveness, and how we may account for the ephemeral nature in a system of interaction. By going back to the drawing board, these two conditions provide a neutral ground for a research agenda, bypassing a potential cultural bias and the cost of reproduction. To question the elusive nature of a musical concept: will Mozart have a therapeutic effect for patients who have never been exposed to common practice music? We can also question

the ephemeral nature of musical experience: what alternatives to a live concert can afford the liveness in musical experiences?

Perhaps, the most important thesis in musical interaction research is that the musical proposition for multimodal interaction dynamics is critical for *situating technology in human activity whereby musical liveness is experienced*. For musical interaction, a musical proposition is at the core of utility, or value potential for both music and multimodal technologies, for which the proposition needs to be properly reflected in instrumentation design. The elusive and ephemeral nature of music discussed above are in favor of this agenda to conceive a radically different frame of reference. In this context, thinking towards future poles and markers, the conceptual framework with five enabling dimensions is developed. To summarize, the 1st dimension, Affordance enables an ecological relationship between systems and people. A Mental Model of musical interaction includes a predictive time model, which is related to the 4th and 5th dimensions. The 2nd dimension, Design Alignment yields instrumentation requirements and system specifications, and enables an optimization between music and design resources for users' and systems' utility, optimizing system implementation tasks by structurally coupling music and design components. The 3rd dimension, Adaptive Learning enables a mutual growth of people and system, connected by the formation of a Mental Model from users' end and machine intelligence from systems' end. The 4th dimension, Second-Order Feedback accounts for performance sensitivity in time critical interaction by enabling users to reflect on their own performances in cascading flows of action perception cycles, which constitutes the experiential musical and technological liveness. The 5th dimension, Temporal Integration in multimodal system architecture anticipates the users' mental model for governing their interaction with pacing and predictions.

Conclusion

The interdisciplinary nature of musical interaction suggests the value of a conceptual framework that can be shared among researchers, the framework not too narrow or prescriptive. The contextual survey in Section 4 demonstrates that, based on their domains of expertise, researchers may differently define and describe the overlapping research inquiries and approaches to musical interaction, applying their own domain specific terms. The five enabling dimensions offered a functionally comprehensive set of interdisciplinary concepts for surveying the works of the researchers in this issue, who come from many different disciplines. This indicates that the framework may offer a coherent way of describing and assessing research and design approaches to musical interaction as well as system implementations. Future research may elicit different ways of framing an agenda and/or substantiating this framework.

We have yet to see how a community of practice around musical interaction may evolve. While the first three dimensions point to foundational and common-sense approaches to future direction, the 4th and 5th are highlighted as necessary conditions for MTI systems for musical interaction. In terms of motivations, the future perspectives implied in the framework include (1) building a dependable relationship and trust between everyday people and systems in the circuit of time critical interaction, and (2) envisioning multimodal music systems for highly skilled performers whose expertise can venture into generating alternative musical senses and experiences with MTI systems supported by robust 4th and 5th dimensions. This article concludes with a working definition: *Musical interaction is an ontological completion of a state of music and listening, through a listener's active engagement with musical resources in multimodal information flow*. In this definition, "music" and "listen" each appear in two contexts: music as an objectified entity and musical resources as encoded values; listening as an action and listener as an actor.

Funding: This research received no external funding.

Acknowledgments: The reviewers of this special issue played a significant role in improving and substantiating the body of bibliographic references, for which all contributing authors were

extremely grateful. The discourse conducted through the review process was both rigorous and cordial, and I believe the process was enlightening for both authors and reviewers. All articles in this issue informed the perspectives of this introduction.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. The Common Practice Period: Two Contributions that Enabled and Defined Musical Interaction

Though differing to some degree, musicians in general implicitly explore and exploit musical affordances. The exploitation may be subconscious and not explicitly analytic, but it employs assessment and judgement on interaction involving instruments, environment, the musician's own physiology, and musical interpretations. When the relationship among these constituents is optimally exploited and manifested flawlessly in circular flows of interaction, we can say that an optimal performance is achieved. This exploitation is routinely exercised during musicians' rehearsal and is critical for learning and developing musical skills and expertise. Then, what is the musician's community of practice and what frame of reference does she consult in the implicit exploitation of affordances? Here we turn to the case of the Common Practice Period.

Mozart, Beethoven, and the Romantic composers belonged to the Common Practice Period of Western European art music, and the period influences persist today demonstrating strength as a paradigm still actively underlying jazz, popular music, film music, and new "classical" repertoire. The Common Practice Period represents a contextual adaptation case through conflicts and resolutions between new and old ideas, techniques, systems, and theories, directly anchored on human experiences of musical sensations and epistemology.

Two critical conditions for catalysing common practice emerged from developments during the medieval era. Both conditions helped establish a sharable reference framework for aligning the production of musical signals:

1. A fixed unit duration applied to time division to enable performers to execute rhythmic patterns in a unified framework; the patterns were codified in a notation system with reference to a fixed unit duration.
2. A common framework for tuning frequencies in pairs of notes called *intervals*; this was codified in *temperament systems* which coevolved with theories of consonance and dissonance and developing musical practice.

The following two sections survey the development of these two reference systems.

Appendix A.1. Definition of Temporal Resolution: Unit Duration and Rhythmic Patterns

It is difficult to imagine the medieval sense of music without extensive scholarly experience decoding medieval notation into modern notation. For processing musical information, modern DMI utilizes parametrization such as pitch and duration, and many more attributes. But from antiquity through the medieval era, pitch and duration were not separable in representations of melody. Perhaps this was because a melody meant an inflection pattern over a text string where the sense of duration was in the oral tradition, the 'know how' for concatenating phonemes and syllables to manage vowels and consonant as sung. In that regard, a community of practice meant an oral tradition. The system of bars, meters, and measures as we know today is the system of the Common Practice Period. It descended from the 13th century mensural notation system, which was used to codify a set of repetitive temporal patterns called rhythmic modes, commonly used in the oral tradition. The mensural notation system used two basic elements: (1) the staff system invented by Guido D'Arezzo around 1000 to codify the pitch direction of melodies, and (2) the note system of *neumes* (from the Greek "neuma" meaning "a sign"). Leo Treitler [96] describes neumes as the melodic inflections of syllables; early neumes encoding melodic contours appeared in the 9th Century [97] with a graphical form over texts as a mnemonic device.

Polyphony refers to multiple musical lines written in a coherent progression to be performed simultaneously by multiple voices or instruments. With the emergence of polyphonic ensemble music, mensural notation became inadequate to support more dynamic musical interaction among players creating complex rhythmic patterns. Out of necessity two concurrent developments emerged in the late 13th and early 14th century: the Ars Nova School of Philippe de Vitry in France, and the Franconian School of Marchettus de Padua in Italy. The two schools disagreed on the fundamental unit value concept for a note, which was called *Brevis*. In French notation, "...(*Brevis*) may be shortened or lengthened ..., but in the Italian system it is an unalterable value" [97] (p. 370). The French system allowed more interpretive and expressive notation techniques preserving the authenticity of the oral tradition of the earlier rhythmic mode where the time values of the short notes (*Brevis*) were often contextually varied based on adjacent notes. With this system, different interpretations were equally logical and plausible, which implies that decoding the French system required more expertise and experiences. On the contrary, the Italian system defined an unalterable unit duration and adopted a strictly additive rhythmic representation built on the smallest unit. The latter afforded a unified approach for encoding different rhythmic patterns, which was easier for performers. Even though the Italian school was heavily criticised by the French school for compromising artful and sophisticated interpretations preserving the oral tradition, the Italian codification of rhythmic division was consistent and efficient and more scalable and led to wider adoption. The latter also might have been more welcoming for new generations of performers and emerging ensembles with new musical genres. In sum, the French system during Ars Nova provided an excellent foundation for a modern notation system, but it was the improvement by the Italian system that afforded the notation system accuracy and scalability that led to the common practice.

We may imagine the increasing specificity in advanced notation leading to an interesting shift in the relationship between notation and oral tradition. For example, the advanced notation system alleviated the dependency of exclusively oral musical practices, which influenced practical musical pedagogy. With the help of notation, musicians could sing unknown melodies without hearing them first, and could reproduce complex passages with greater accuracy. It is possible to argue that in the present era we are encountering an analogous transformation of musical applications through MTI which affords wider adoption of musical interactions, but also demands a more accurate reference system for alignment of multimodal relationships.

Appendix A.2. Definition of Pitch (Frequency) Resolution: Temperament

A musical note has a pitch defined from "low" to "high" by its fundamental frequency. Two notes form an interval, which is the relative frequency separation of two pitches measured by the number of pitches between them. This measure depends on an agreed fixed set of pitches with agreed tunings. The agreement of pitches and tunings is no small matter; your favourite music depends on tuning resulting from 1000 years of debate and prototyping. There are many possibilities for defining frequency intervallic relationship between two notes presented concurrently or in series. The reference rests on how the quality of intervals among those pitches is aimed and tempered. Tuning refers to the adjustment of strings or tubes to control frequency responses in the instruments. Historically, different instruments carried different conventions of tuning and certain music repertoires were written for specific tuning systems.

The theory of harmony we know today was not established until the treatise *Le Istitutioni Harmoniche* (1558) written by the Italian composer and theorist, Geoseffo Zarlino. He introduced the intervals of the 3rd and 6th as harmonic constituents and emphasized the importance of major and minor 3rd, which anticipates the common practice. Zarlino reasoned the 3rd and 6th intervals should be classified as consonances by a "number 6 scenario" for dividing a string into six equal parts, rejecting the four equal parts which create the Pythagorean perfect intervals, the Octave, 5th, and 4th. Zarlino's treatise was a

breakthrough towards establishing the three-note triad as the fundamental harmonic unit which is common in contemporary and popular music we know today.

Tempering describes fine-tuning by introducing small variations in frequency. (The term originates from the temperature conditions effecting micro-frequency responses of physical tubes and strings.) In practice, “tempering” applies a frequency deviation needed to reconcile the Pythagorean perfect intervals such as 4th and 5th, and the consonant intervals such as 3rd and 6th. Temperament is a system of micro-tuning; in history different temperament systems were exercised. Pythagorean is the oldest tuning using four subdivisions of a string to determine all intervals between pitches. *Just intonation* applies whole number ratios to tune all pitch intervals in a scale. The consequence is a pure sonority that works well for monophonic music playing in one key (using the notes of only one scale) but does not support polyphonic music or transposition of one key to another. Just intonation was dominant for nearly 400 years during the medieval era and was favoured by purists. *Mean tone temperament* became dominant circa the late 15th to early 18th centuries, which afforded and promoted the innovation of triadic harmonies by tempering the pure perfect fifth in favour of the major third. This prepared the fully developed major and minor chords as harmonic units in the Common Practice Period, which became the foundation of all chord progressions, including Jazz and popular music today. By the late Baroque, J.S. Bach (1685–1750) worked with an instrument maker to create the well-tempered clavier, a keyboard instrument which applied systematic micro-tuning to achieve uniform pitch intervals. This affordance enabled Bach to compose his Inventions, a series of polyphonic contrapuntal compositions in all keys, 12 major and 12 minor. These 24 keys are the harmonic scaffolding of most contemporary and popular music. Today’s symphony orchestra also leverages the history of compromises among different temperament systems, and is primarily organized around *Equal temperament*, which divides an octave with twelve equally spaced pitches. The modern piano is tuned in equal temperament so that 12 pitches of any octave remain constant across all octaves aiding instruments to play together in a unified tuning reference.

Without the unified system of references from the Common Practice Period, the symphony orchestra and its emblematic statue of a conductor would not exist. Over 80 musicians play in a typical symphony orchestra, who are all trained primarily in the common music repertoire. Today’s symphony orchestra is a product of the Common Practice Period symbolizing the summit of a Western European historical trajectory. This flash survey gives a glimpse of the complexity around musical practice, and how its trajectory arrives at a peak consensus of highly specialized practices such as Equal Temperament, which enabled large ensembles to perform together at the cost of prior arts such as Just intonation, which became archaic when it could not contribute to the wider social norm.

Appendix B. The Relationship between Affordance and Mental Model

The following illustrates the relationship between affordance and mental model. When the mobile phone was invented, a new relationship between people and telephony became possible, from phone at home to phone in a pocket. Portability was an affordance of size and battery operation. However, mobile phones inherited the mental model of a phone at home. We can compare affordance and mental model in use-case syntax, ‘x uses y to do z’: (1) ‘people use mobile phones to make calls on the move’ indicates a mental model, and this makes sense, whereas (2) ‘people use mobile phones to carry around’ indicates the new affordance, but it sounds redundant because, while affordance and mental model are related, the former weighs on relationship and the latter on use cases and workflows.

Referring to the Common Practice Period example (Appendix A), equal temperament enabled an increase in the variety of musical instruments that could play together and remain in tune. Increasing the consistency of tuning introduced an affordance for the formation of an orchestra, the new relationship among musical instrumental groups. Performers had to expand their mental model with the larger ensemble, attending to the

conductor, recognizing different instrumental groups to follow the score, but largely inherited the existing mental model, ‘to play my part correctly and in tune’.

References

- Steinbeis, N.; Koelsch, S. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* **2008**, *18*, 1169–1178.
- Fedorenko, E.; Patel, A.D.; Casasanto, D.; Winawer, J.A.; Gibson, E. Structural integration in language and music: Evidence for a shared system. *Mem. Cogn.* **2009**, *37*, 1–9.
- Koelsch, S. Toward a Neural Basis of Music Perception—A Review and Updated Model. *Front. Psychol.* **2011**, *2*, 110. <https://doi.org/10.3389/fpsyg.2011.00110>.
- Merrett, D.L.; Peretz, I.; Wilson, S.J. Neurobiological, cognitive, and emotional mechanisms in melodic intonation therapy. *Front. Hum. Neurosci.* **2014**, *8*, 401. <https://doi.org/10.3389/fnhum.2014.00401>.
- Nymoën, K.; Caramiaux, B.; Kozak, M.; Torresen, J. Analyzing Sound Tracings—A Multimodal Approach to Music Information Retrieval. In Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies, Scottsdale Arizona, AZ, USA, 30 November 2011; pp. 39–40.
- Xue, H.; Xue, L.; Su, F. Multimodal Music Mood Classification by Fusion of Audio and Lyrics. In *MultiMedia Modeling. MMM 2015. Lecture Notes in Computer Science*; He, X.; Luo, S.; Tao, D.; Xu, C.; Yang, J.; Hasan, M.A. Eds.; Springer: Berlin/Heidelberg, Germany, 2015; Volume 8936. https://doi.org/10.1007/978-3-319-14442-9_3.
- Oramas, S.; Barbieri, F.; Nieto, O.; Serra, X. Multimodal Deep Learning for Music Genre Classification. *Trans. Int. Soc. Music Inf. Retr.* **2018**, *1*, 4–21. <https://doi.org/10.5334/tismir.10>.
- Simonetta, F.; Ntalampiras, S.; Avanzini, F. Multimodal Music Information Processing and Retrieval: Survey and Future Challenges. In Proceedings of the 2019 International Workshop on Multilayer Music Representation and Processing (MMRP), Milano, Italy, 23–24 January 2019. <https://doi.org/10.1109/MMRP.2019.00012>.
- Standley, J.M. A meta-analysis of the efficacy of music therapy for premature infants. *J. Pediatric Nurs.* **2002**, *17*, 107–113.
- Schlaug, G. Listening to and making music facilitates brain recovery processes. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 372–373.
- Schlaug, G.; Altenmüller, E.; Thaut, M. Music listening and music making in the treatment of neurological disorders and impairments. *Music Percept.* **2010**, *27*, 249–250.
- Goldbeck, L.; Ellerkamp, T. A randomized controlled trial of multimodal music therapy for children with anxiety disorders. *J. Music Ther.* **2012**, *49*, 395–413. <https://doi.org/10.1093/jmt/49.4.395>.
- Arieh, Y.; Marks, L.E. Cross-modal interaction between vision and hearing: A speed–accuracy analysis. *Percept. Psychophys.* **2008**, *70*, 412–421. <https://doi.org/10.3758/PP.70.3.412>.
- Algom, D.; Fitousi, D. Half a century of research on Garner interference and the separability–integrality distinction. *Psychol. Bull.* **2016**, *142*, 1352–1383.
- Leman, M. *Embodied Music Cognition and Mediation Technology*; MIT Press: Cambridge, MA, USA, 2008.
- Särkämö, T.; Tervaniemi, M.; Laitinen, S.; Forsblom, A.; Soinila, S.; Mikkonen, M.; Autti, T.; Silvennoinen, H.M.; Erkkilä, J.; Laine, M.; et al. Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain* **2008**, *131 Pt 3*, 866–876. <https://doi.org/10.1093/brain/awn013>.
- Altenmüller, E.; Marco-Pallares, J.; Münte, T.F.; Schneider, S. Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 395–405. <https://doi.org/10.1111/j.1749-6632.2009.04580.x>.
- Wan, C.Y.; Schlaug, G. Music making as a tool for promoting brain plasticity across the life span. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **2010**, *16*, 566–577. <https://doi.org/10.1177/1073858410377805>.
- Ghai, S.; Maso, F.D.; Ogourtsova, T.; Porxas, A.-X.; Villeneuve, M.; Penhune, V.; Boudrias, M.-H.; Baillet, S.; Lamontagne, A. Neurophysiological Changes Induced by Music-Supported Therapy for Recovering Upper Extremity Function after Stroke: A Case Series. *Brain Sci.* **2021**, *11*, 666. <https://doi.org/10.3390/brainsci11050666>.
- Schlaug, G.; Jäncke, L.; Huang, Y.; Staiger, J.F.; Steinmetz, H. Increased corpus callosum size in musicians. *Neuropsychologia* **1995**, *33*, 1047–1055.
- Schlaug, G. The brain of musicians. A model for functional and structural adaptation. *Ann. N. Y. Acad. Sci.* **2001**, *930*, 281–299.
- Gaser, C.; Schlaug, G. Brain structures differ between musicians and non-musicians. *J. Neurosci.* **2003**, *8*, 9240–9245.
- Hyde, K.L.; Lerch, J.; Norton, A.; Forgeard, M.; Winner, E.; Evans, A.C.; Schlaug, G. The effects of musical training on structural brain development: A longitudinal study. *Ann. N. Y. Acad. Sci.* **2009**, *1169*, 182–186.
- Herholz, S.C.; Zatorre, R.J. Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron* **2012**, *76*, 486–502.
- Reybrouck, M.; Brattico, E. Neuroplasticity beyond Sounds: Neural Adaptations Following Long-Term Musical Aesthetic Experiences. *Brain Sci.* **2015**, *5*, 69–91.
- Gibson, J.J. The theory of affordances. In *Perceiving, Acting and Knowing*; Shaw, R.; Bransford, J., Eds.; Erlbaum: Hillsdale, NJ, USA, 1977.
- Gibson, J.J. The Theory of Affordances. In *The Ecological Approach to Visual Perception*; Taylor & Francis: Boulder, CO, USA, 1979; pp. 119–137.

28. von Bertalanffy, L. The history and status of general systems theory. In *Trends in General Systems Theory*; Klir, G., Ed.; Wiley: New York, NY, USA, 1972; pp. 21–41.
29. Umpleby, S.A.; Dent, E.B. The Origins and Purposes of Several Traditions in Systems Theory and Cybernetics. *Cybern. Syst.* **1999**, *30*, 79–104.
30. Von Foerster, H. (Ed.) *Cybernetics of Cybernetics*; BCL Report 73.38; Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois: Urbana, IL, USA, 1974.
31. Nilsen, P. Making sense of implementation theories, models and frameworks. *Implement. Sci.* **2015**, *10*, 53. <https://doi.org/10.1186/s13012-015-0242-0>.
32. Li, L.C.; Grimshaw, J.M.; Nielsen, C. Evolution of Wenger's concept of community of practice. *Implement. Sci.* **2009**, *4*, 11. <https://doi.org/10.1186/1748-5908-4-11>.
33. Kuhn, T.S. *The Structure of Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1970.
34. Horsley, I. Score. In *The New Grove Dictionary of Musical Instruments*; Sadie, S., Ed.; Oxford University Press: Oxford, UK, 1984; Volume 17, pp. 59–67.
35. Tillman, B.; Bharucha, J.J.; Bigand, E. Learning and perceiving musical structures: Further insights from artificial neural networks. In *The Cognitive Neuroscience of Music*; Peretz, I., Zatorre, R., Eds.; Oxford University Press: London, UK, 2003; pp. 109–123.
36. Johnson, W. Musical evenings in the early Empire: New evidence from a Greek papyrus with musical notation. *J. Hell. Stud.* **2000**, *120*, 57–85. <https://doi.org/10.2307/632481>.
37. Weiner, A. The Function of the Tragic Greek Chorus. *Theatre J.* **1980**, *32*, 205–212.
38. Trehub, S.E.; Becker, J.; Morley, I. Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.* **2015**, *370*, 1664. <https://doi.org/10.1098/rstb.2014.0096>.
39. Lim, N. Cultural differences in emotion: Differences in emotional arousal level between the East and the West. *Integr. Med. Res.* **2016**, *5*, 105–109.
40. Wirtz, D.; Chiu, C.Y.; Diener, E.; Oishi, S. What constitutes a good life? Cultural differences in the role of positive and negative affect in subjective well-being. *J. Personal.* **2009**, *77*, 1167–1196.
41. Manen, M.V. *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*; Routledge: London, UK, 1997.
42. Park, D.C.; Huang, C.M. Culture Wires the Brain: A Cognitive Neuroscience Perspective. *Perspect. Psychol. Sci.* **2010**, *5*, 391–400. <https://doi.org/10.1177/1745691610374591>.
43. Paige, L.E.; Ksander, J.C.; Johndro, H.A.; Gutchess, A.H. Cross-cultural differences in the neural correlates of specific and general recognition. *Cortex* **2017**, *91*, 250–261.
44. Gutchess, A.H.; Inneck, A. Cultural influences on memory. *Prog. Brain Res.* **2009**, *178*, 137–150.
45. Hiller, L.; Isaacson, L. Quartet No. 4 for Strings 'Illiac Suite', 1957. Vinyl Disc recording. In *Computer Music from the University of Illinois*; MGM-Heliodore: New York, NY, USA, 1967 (Vinyl H/HS25053); re-released as Quartet No. 4 for Strings 'Illiac Suite'. In *Lejaren Hiller: Computer Music Retrospective*; Wergo: Mainz, Germany, 1987 and 1989 (Vinyl WER 60128, 1987, and CD WER 60128-50, 1989).
46. Mathews, M.; Arthur, R. The Technology Of Computer Music. *Phys. Today* **2008**, *23*, 60. <https://doi.org/10.1063/1.3022299>.
47. Copeland, J.; Long, J. Alan Turing: How his universal machine became musical instrument. *IEEE Spectrum* **2017**, *28*. Available online: <https://spectrum.ieee.org/alan-turing-how-his-universal-machine-became-a-musical-instrument> (accessed on 31 December 2021).
48. Gibson, J.J. *The Senses Considered as Perceptual Systems*. Allen and Unwin, London, England, 1966.
49. Norman, D.A. *The Psychology of Everyday Things*; Basic Books: New York, NY, USA, 1988.
50. Norman, D.A. Affordances, Conventions, and Design. *Interactions* **1999**, *6*, 38–41.
51. Gaver, W. Technology affordances. In Proceedings of the SIGCHI conference on Human factors in computing systems, New Orleans, LA, USA, 27 April–2 May 1991; pp. 79–84.
52. Reybrouck, M. Musical Sense-Making and the Concept of Affordance: An Ecosemiotic and Experiential Approach. *Biosemiotics* **2012**, *5*, 391–409. <https://doi.org/10.1007/s12304-012-9144-6>.
53. Menin, D.; Schiavio, A. Rethinking Musical Affordances. *Avant* **2012**, *3*, 202–215.
54. Krueger, J. Affordances and the musically extended mind. *Front. Psychol.* **2014**, *4*, 1003.
55. Rowe, R. Representations, Affordances, and Interactive Systems. *Multimodal Technol. Interact.* **2021**, *5*, 23. <https://doi.org/10.3390/mti5050023>.
56. Lettvin, J.Y.; Maturana, H.R.; McCulloch, W.S.; Pitts, W. What the Frog's Eye Tells the Frog's Brain. *Proc. Inst. Radio Engr.* **1959**, *47*, 1940–1951.
57. Craik, K. *The Nature of Explanation*; Cambridge University Press: Cambridge, UK, 1943.
58. Johnson-Laird, P.N. Mental models and human reasoning. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18243–18250. <https://doi.org/10.1073/pnas.1012933107>.
59. Ashby, W.R. Requisite Variety and its implications for the control of complex systems. *Cybernetica* **1958**, *1*, 2.
60. von Foerster, H. *Observing Systems*; Intersystems Press: Seaside, CA, USA, 1984.
61. Fuster, J.M. Upper processing stages of the perception-action cycle. *TICS* **2004**, *8*, 143–145.
62. O'Regan, J.; Noë, A. A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* **2001**, *24*, 939–1031.
63. O'Regan, J.; Noë, A. What is it like to see: A sensorimotor theory of perceptual experience. *Synthese* **2002**, *129*, 79–103.

64. Di Paolo, E.A.; Buhrmann, T.; Barandiaran, X.E. *Sensorimotor Life: An Enactive Proposal*; Oxford UP: New York, NY, USA, 2017.
65. Thompson, E. Sensorimotor subjectivity and the enactive approach to experience. *Phenomenol. Cogn. Sci.* **2005**, *4*, 407–427.
66. Gomes, H.; Molholm, S.; Christodoulou, C.; Ritter, W.; Cowan, N. The development of auditory attention in children. *Front. Biosci.* **2000**, *5*, D108–D120. <https://doi.org/10.2741/gomes>.
67. Snyder, J.S.; Gregg, M.K.; Weintraub, D.M.; Alain, C. Attention, awareness, and the perception of auditory scenes. *Front. Psychol.* **2012**, *3*, 15.
68. Oxenham, A.J. How We Hear: The Perception and Neural Coding of Sound. *Annu. Rev. Psychol.* **2018**, *69*, 27–50. <https://doi.org/10.1146/annurev-psych-122216-011635>.
69. Maturana, H.R. Neurophysiology of cognition. In *Cognition: A Multiple View*; Garvin, P., Ed.; Spartan Books: New York, NY, USA, 1970, pp. 3–23.
70. Heylighen, F.; Joslyn, C. Cybernetics and Second-Order Cybernetics. In *Encyclopedia of Physical Science & Technology*, 3rd ed.; Meyers, R.A., Ed.; Academic Press: New York, NY, USA, 2001.
71. Madl, T.; Baars, B.J.; Franklin, S. The Timing of the Cognitive Cycle. *PLoS ONE* **2011**, *6*, e14803.
72. Franklin, S.; Strain, S.; McCall, R.; Baars, B. Conceptual Commitments of the LIDA Model of Cognition. *J. Artif. Gen. Intell.* **2013**, *4*, 1–22. <https://doi.org/10.2478/jagi-2013-0002>.
73. Choi, I. Structured Reciprocity for Musical Performance with Swarm Agents as a Generative Mechanism. In *Advances in Computer Entertainment Technology, Lecture Notes in Computer Science* 10714; Cheok, A.D.; Inami, M.; Romao, T. Eds.; Springer: Berlin/Heidelberg, Germany, 2017, pp. 689–712.
74. Choi, I. Interactive Composition and Performance Framework with Evolutionary Computing. In Proceedings of the 43rd International Computer Music Conference (ICMA), Shanghai, China, 16–20 October 2017, pp.351–356.
75. Craik, K.J. Theory of the human operator in control systems; the operator as an engineering system. *Br. J. Psychol. Gen. Sect.* **1947**, *38 Pt 2*, 56–61. <https://doi.org/10.1111/j.2044-8295.1947.tb01141.x>. PMID: 18917476.
76. Tipei, S.; Craig, A.B.; Rodriguez, P.F. Using High-Performance Computers to Enable Collaborative and Interactive Composition with DISSCO. *Multimodal Technol. Interact.* **2021**, *5*, 24. <https://doi.org/10.3390/mti5050024>.
77. Bruin, H. From Musical Ideas to Computers and Back. In *The Computer and Music*; Lincoln, H. ed.; Cornell University Press: Ithaca, NY, USA, 1970; pp. 23–36.
78. Wu, J. Promoting Contemplative Culture through Media Arts. *Multimodal Technol. Interact.* **2019**, *3*, 35. <https://doi.org/10.3390/mti3020035>.
79. Fraisee, V.; Wanderley, M.M.; Guastavino, C. Comprehensive Framework for Describing Interactive Sound Installations: Highlighting Trends through a Systematic Review. *Multimodal Technol. Interact.* **2021**, *5*, 19. <https://doi.org/10.3390/mti5040019>.
80. Guidotti, R.; Monreale, A.; Turini, F.; Pedreschi, D.; Giannotti, F. A survey of methods for explaining black box models. *ACM Comput. Surv.* **2018**, *51*, 93.
81. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.A.; Kagal, L. Explaining Explanations: An Overview of Interpretability of Machine Learning. In Proceedings of the IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 80–89.
82. Arrieta, A.B.; D’iaz-Rodríguez, N.; Ser, J.D.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115.
83. Chadabe, J. The Limitations of Mapping as a Structural Descriptive in Electronic Instruments. In Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME), Dublin, Ireland, 24–26 May 2002.
84. Zador, A.M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **2019**, *10*, 3770.
85. Bobbe, T.; Oppici, L.; Lüneburg, L.-M.; Münzberg, O.; Li, S.-C.; Narciss, S.; Simon, K.H.; Krzywinski, J.; Muschter, E. What Early User Involvement Could Look Like—Developing Technology Applications for Piano Teaching and Learning. *Multimodal Technol. Interact.* **2021**, *5*, 38. <https://doi.org/10.3390/mti5070038>.
86. Fitzek, F.H.P. *Tactile Internet with Human-in-the-Loop: New Frontiers of Transdisciplinary Research*; Academic Press: Cambridge, MA, USA, 2021.
87. Simsek, M.; Aijaz, A.; Dohler, M.; Sachs, J.; Fettweis, G. 5G-Enabled Tactile Internet. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 460–473. <http://doi.org/10.1109/JSAC.2016.2525398>.
88. Sheridan, T.B. Space teleoperation through time delay: Review and prognosis. *IEEE Trans. Robot. Autom.* **1993**, *9*, 592–606. <https://doi.org/10.1109/70.258052>.
89. Minsky, M. Telepresence. *Omni Magazine*, 1980, pp. 45–51. Available online: <https://web.media.mit.edu/~minsky/papers/Telepresence.html> (accessed on 31 December 2021).
90. Clément, A.; Moreira, L.; Rosa, M.; Bernardes, G. Musical Control Gestures in Mobile Handheld Devices: Design Guidelines Informed by Daily User Experience. *Multimodal Technol. Interact.* **2021**, *5*, 32. <https://doi.org/10.3390/mti5070032>.
91. Poggi, I.; Ranieri, L.; Leone, Y.; Ansani, A. The Power of Gaze in Music. Leonard Bernstein’s Conducting Eyes. *Multimodal Technol. Interact.* **2020**, *4*, 20. <https://doi.org/10.3390/mti4020020>.
92. Haynes, A.; Lawry, J.; Kent, C.; Rossiter, J. FeelMusic: Enriching Our Emotive Experience of Music through Audio-Tactile Mappings. *Multimodal Technol. Interact.* **2021**, *5*, 29. <https://doi.org/10.3390/mti5060029>.

93. Posner, J.; Russell, J.A.; Peterson, B.S. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* **2005**, *17*, 715–734. <https://doi.org/10.1017/S0954579405050340>.
94. Hevner, K. The affective character of major and minor modes in music. *Am. J. Psychol.* **1935**, *47*, 103–118.
95. Senkowski, D.; Schneider, T.R.; Foxe, J.J.; Engel, A.,K. Crossmodal binding through neural coherence: Implications for multi-sensory processing. *Trends Neurosci.* **2008**, *31*, 401–409. <https://doi.org/10.1016/j.tins.2008.05.002>.
96. Treitler, L. The Early History of Music Writing in the West. *J. Am. Musicol. Soc.* **1982**, *35*, 237–279. <https://doi.org/10.2307/831146>.
97. Apel, W. *The Notation of Polyphonic Music, 900–1600*; The Medieval Academy of America: Cambridge, MA, USA, 1953.