Ghent University 2011-2012
Faculty of Arts and Philosophy
IPEM - Department of Musicology

Promotor
Prof. Dr. Marc Leman

Reading committee
Dr. Micheline Lesaffre
Dr. Michiel Demey
Dr. David Worrall
Prof. Thierry Dutoit
Dr. Peter Lambert

Dean
Prof. Dr. Freddy Mortier

Rector
Prof. Paul Van Cauwenberge

Published
Faculty of Arts and Philosophy - Ghent University
ISBN 978 90 7083 087 8

A MULTIMODAL FRAMEWORK FOR INTERACTIVE SONIFICATION AND SOUND-BASED COMMUNICATION

NUNO CORREIA DA SILVA DINIZ

THESIS SUBMITTED TO FULFILL THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR IN ART SCIENCE
PROMOTOR PROF. DR. MARC LEMAN

Acknowledgements

I would like to express my deepest gratitude to all the people without whom the realization of this thesis would have been be much more difficult, if not impossible.

First of all, I would like to express my sincere gratefulness to my promoter, Prof. Dr. Marc Leman, for the trust he deposited in me and my work, the countless inspirational debates that we had, his valuable and experienced guidance, his keen expertise and his endless contribution to this project.

A special word of gratitude goes out to Dr. Michiel Demey, who decisively contributed in the steering and structuring of this project since its early stages, and to Dr. David Worrall, for the very fruitful and pleasurable discussions during our encounter at the ICAD conference in Washington.

I would like to thank all my colleagues and the academic staff at IPEM for creating a stimulating environment over the course of this journey. This thesis would not have been the same without the many energetic and interesting discussions that took place during its development. Also, I would like to express my thanks to the administrative and technical staff at IPEM, who helped me fight the many bureaucratic perils, logistic tasks and technological challenges during this four years.

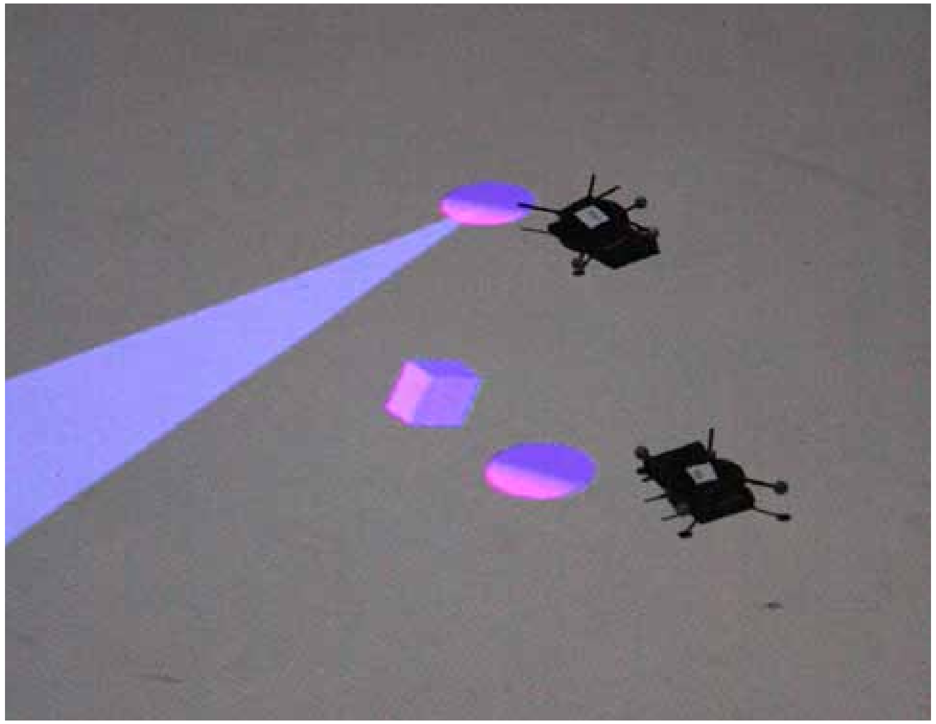
I would like to give a very special cheers to my colleagues and friends Pieter Coussement and Alexander Deweppe, who's true camaraderie was an invaluable support during this quest. Since the very beginning of this adventure, the collaborative work that we have done together as well as our numberless discussions and talks about music, technology and life decisively shaped both the work presented here and my living in this strange, new land. To my "brothers in arms", I say: Santé!

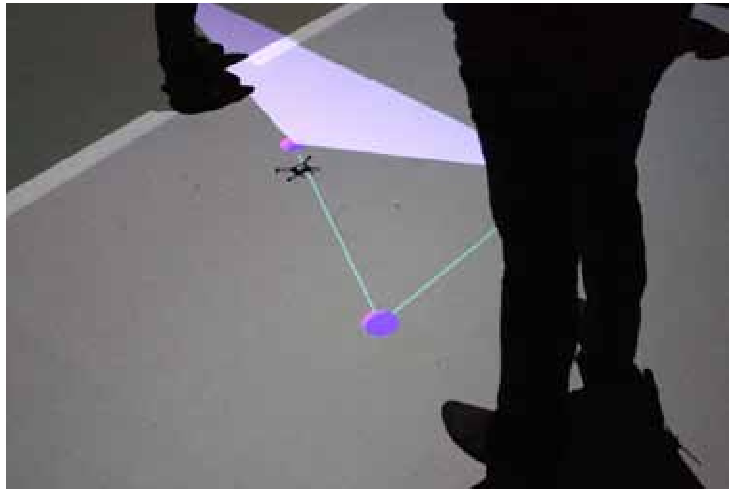
I would like to send my gratitude to all my dearest friends for their encouragement and support during the all good and bad times as well as to the new friendships I developed over this period, who's sympathy and companionship made me feel at home, away from home.

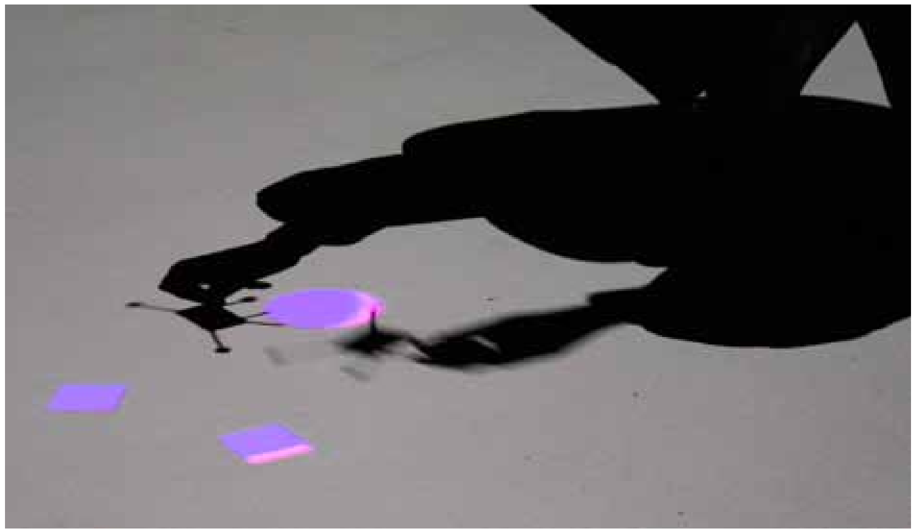
I would like to express my heartfelt love to my parents, Noémia and Augusto Diniz, as well as to my beloved family for their love, comfort and understanding during the many grey days and gratifying victories that I have encountered during this journey. To all, "o meu mais sincero obrigado".

And finally, I would like to express my unlimited love and gratitude for my dearest one, Elvire Delanote. Words are not sufficient to profess my thankfulness for her boundless dedication, strength, help, caring, affection and smile throughout our life together.

Index







chapter 1 / introduction

1.1 Overview

Auditory display uses sound to explore structures in data [94] [72]. However, research on auditory display has yet to provide a comprehensive framework that incorporates the basic principles for this exploration, such as sound design, strategies of interaction and software tools. The goal of this thesis is to provide new building blocks that contribute to the development of such a framework for auditory display. We conceive these building blocks both from the viewpoints of concept and implementation. Our ultimate goal is to establish a working framework that can be used for building different use cases within the auditory display field.

Our inspiration for developing these building blocks is based on two sources, namely (i) the theory and practice of electroacoustic music and (ii) the theory of embodied music cognition and its viewpoint on mediation technologies.

(i) Throughout its evolution, electroacoustic music composition forged several analytical and production directives that aimed at unifying the manipulation of micro and macro levels of sound based communication. Micro levels of sound can be comprised as sonic materials, such as timbre, texture, pitch, while macro levels of sound may be conceived as larger structures that comprise these sonic materials. In addition, the directives are to a large extend based on human factors, as they constrain the materials and structures towards a perceivable and meaningful communication.
The objective is first to identify the relevant guidelines that have been used in electroacoustic music composition, and then to develop these guidelines as functional components of an auditory display framework. The main focus will be on interaction strategies, as this is a domain of study that is less explored than mapping strategies between sounds and meanings. Interaction strategies address principles such as "scope transposition" and "embodied multilevel access". These principles aim to provide the necessary guidelines for the development of tools for a multiple perspective analysis of the sonic context and the individual auditory streams.

(ii) To make the implementation of the above mentioned guidelines successful, we adopt a paradigm of embodied music cognition [99]. The theory of embodied music cognition considers the human body as a mediator between the (subjective) experience and the (objective) environment. Interestingly, this mediator can be extended with technologies, in such a way that mental access to a new and previously unreachable domain becomes accessible. By considering technology as an extension of the human body, a virtual object mediation model is adopted through which the user communicates in a multimodal dialogue, using principles such as "scope transposition" and "multilevel access". In this way, the involvement of the multimodal spectrum (visual, aural, spatial) in both user and system facilitates the perception of its constituent elements and thus relieves the cognitive load imposed to the user when dealing with high dimensional, multivariable scenarios. By acknowledging the interconnection between functional sounds, interactive sonification and music/media arts [72], these guidelines are explored from a utilitarian viewpoint. This viewpoint constitutes the basis for the implementation of a mediation technology for auditory display.

In this thesis, multimodal interaction and multimedia technology are considered to be fundamental for implementing interaction models based on the human body's capabilities to enhance the understanding of sonically illustrated behaviors. Through its active "closed loop" participation in the exploratory, monitoring and production tasks, it then becomes possible to fully integrate and take advantage of the expertise and intuition of the user (either sonifier or composer).

Given the composed and hierarchical nature of sound objects on one hand, and the human auditory and motor system on the other hand, a connection can be achieved through the exploration of multimedia technologies that places them in a shared, three-dimensional multimodal space. Although adaptable to other interaction scenarios, this embodied approach is maximized when operating in a three-dimensional immersive environment where spatial and temporal aspects of human cognition can be fully expressed. This approach constitutes the core of the development presented in this thesis. Given the participation of the human body in the perception and cognitive process, both multimodal elements and the field expert inhabit a common space of multimodal interaction through which their dialogue can be intuitive and streamlined. As such, in order to objectify the participating elements (e.g. data and sound) as well as the tools for accessing it, several electroacoustic inspired metaphors were developed. They are called the "virtual inspection window" and the "virtual inspection tool". Through "scope variation" and "multilevel mapping", the targeted elements are metaphorically transposed and represented in space as virtual volumes onto which the concepts of "virtual inspection window" and "virtual inspection tool" can be applied. This enables its access by user through embodied inspection actions that are of the most natural character.

Drawing on previous work in virtual, augmented and mixed reality domains, we developed the JOINDER software framework as an infrastructure for auditory display research. Based on object oriented modular implementation, this software framework focuses on the support of multiple input sources, multimodal output, distributed execution, multilevel communication and cross platform compatibility. Furthermore, a wrapping strategy is followed in order to establish a central bridge for expert, heterogeneous technologies - such as visual engines, audio synthesis environments and multimodal interface managers. By incorporating mature state of the art technologies and providing a communication platform between them, the JOINDER framework allows the testing of new interfaces through rapid prototyping. In addition, previous developments can be incorporated in a straightforward, modular way.

Consequently, by creating a technological framework for embodied mediation interfaces, this work aims at giving an answer to the research needs in the field of auditory display. More importantly, it aspires to render a scalable information systems's infrastructure over heterogeneous technology and environments used in multimodal and multimedia domains. As such, its implementation rests on a dependable environment - Java - strongly firmed in virtually all IT development areas. This facilitates the expansion of the current version to other technologies such as database connectivity, web access and servicing (e.g. RESTFul, RMI), data mining, CMS integration, and so on.

In order to test and refine the present implementation, a rapid prototyping and user centered design methodology was followed, ranging from laboratory confined preliminary studies to a on-site four months use case development (the Sound- Field project). Focusing on an implementation within interactive arts domain, this last stage offered an expansion of the initial requirements extracted from interaction sonification practices. Furthermore, it provided a demonstration of the functional, neutral character of the developed interfaces and related metaphors concerning its applications in both scientific and artistic domains. In general, this strategy presented the means to refine top down initial design and architecture strategies through bottom up, real world deployments.

Although targeting auditory display, interactive sonification and later music composition, JOINDER's range of application can transcend these domains, given the rapid adoption of multimodal interfacing in many fields and mass consumer products. These might include, among others: interactive advertising, and web content distribution, game development, holographic prototyping, organic interfaces design and mockup implementation.

In summary, this thesis presents:

- An analysis of the intersection between the interaction elements involved in interactive sonification and musical composition (e.g. sound object, sensorimotor constrains of the human body).
- The extraction of interaction metaphors (representation and access) based on electroacoustic theory and practice (e.g. Schaeffer, Stockhausen, Wishart, Smalley) and embodied spatial interfacing.
- The design and implementation of the JOINDER software framework to deploy multimodal applications based on the developed metaphors and its validation (e.g. Interactive sonification use cases, the SoundField project).

By focusing on interaction strategies and the added valued of the relations provided by embodied interfacing, it establishes a conceptual and technological bridge between mapping strategies and multimodal/immersive/multilevel interaction. As such, this approach aims at relieving the burden of the so called "right mapping" quest and at closing the semantic gap between sonic information and high levels semantic representations of the user.

1.2 State of the art I - interactive sonification

The auditory display community has produced several software packages since its establishment as a research field. These results can be considered a consequence of the agenda sketched in the seminal NSF report [94] concerning the development of generic data sonification tools for research and application. On one hand, it acknowledges the existence and variety of software tools for sound analysis, synthesis and manipulation, but one the other hand, it highlights the often encountered limitations concerning computational efficiency, realtime configuration, user accessibility and design flexibility (which is recurrently exposed through integration issues).
Concluding that the current tools are either too complex, specific and/or cumbersome, the report proceeds in outlining several aspects considered to be of imperative relevance for the development of a software infrastructure for the auditory display field. The first directive concerned portability, which means that users should be able to use their tools in various systems and configurations with a high degree of consistency and reliability, independently of their level of complexity. The second directive covered flexibility, as developed tools should be able to embrace new multimodal technology with a minimum configurational effort and use them in realtime scenarios, as interactivity should be a standard feature in accessing a sonification system. And finally, a third directive addressed integrability, which underlines the importance of open design and implementation in facilitating data migration and adoption of new rendering systems within, for instance, virtual reality and assistive technology domains. With these objectives in mind, the new software solutions should offer a "standardized software layer" that is easy to maintain and expand to future needs. This item is finalized with a research proposal that focuses on the implementation of a generalist and accessible software framework through the use of a "platform-independent, object-oriented technology (such as Java)".

As previously stated and illustrated in [7], a number of software systems for sonification have been developed since the NSF report publication. These systems range from applications (ex. Listen [157], MUSE [104]) to software frameworks/toolkits. Among these last, we will focus predominantly on the ones that already aim at addressing the above mentioned portability, flexibility and integrability directives.

- The **Sonification Sandbox** [35] offers a GUI-based Java application and respective class toolkit for sonifying generic data through MIDI. Besides value mapping to common auditory dimensions, it provides interval notification's configuration as well as import and export functions to Excel and QuickTime formats.

- The **SoniPy framework** [163] is a Python based software framework for data sonification. It presents a highly structured implementation and is comprised of modules for data acquisition and mining, mapping and perceptual modeling as well as sound rendering. These modules are monitored by the user through the defined Sonipy Data and Control Network. Besides its comprehensive contribution, it has the advantage of being build in Python, a multi-platform, multi-paradigm development framework that enjoys a vast popularity among the scientific and creative community.

- The **AeSon Toolki**t [7] is a collection of Max/MSP patches that aims to put forward a user-centered customization of the inspected data's aesthetic representations. Its theoretical background lies within musical aesthetics (Xenakis, Varèse) and information visualization theory (Fry, Maeda, Turfe). It focuses on conveying intuitive and simplified tools for sonification of multivariate data by providing scaling and zooming representation of the data. Although being GUI (Graphical User Interface) based, it emphasizes the importance of including gestural controllers for data exploration (ex. touchpad, touch-screen, graphics tablet, WiiMote).

- The **SonEnvir** [36] is a generalized sonification environment built on the SuperCollider3 and Pure Data environments. It enjoys a considerable user community and it targets the analysis of scientific datasets. It offers a modular implementation regarding data handling, data processing, sound synthesis processes, mappings options and real-time interaction possibilities.

- The **Interactive Sonification Toolkit** [121] is a sonification package that focuses on tackling issues that rise both from the interactive navigation and real time data gathering. The base environment of choice is Pure Data and the GUI is divided in two areas, namely data scaling and interactive sonification. The sonification engine offers a considerable range of mappings options (audification, data to note pitch/duration, additive synthesis, noise filtering) and quick dataset comparative tests.

Several observations can be made about these systems. First of all, most of the presented solutions refer to the user as a (near) passive element within the auditory investigation process. The user can turn knobs, cut or enhance sections of the sonic output and even alter, in real time, the mappings. However, none of the presented solutions fully embraces the potential of the embodied listener, that is, a listener who fully uses actions in space as a basis for exploring the auditory environment. As it will be shown further in this thesis, in order to explore the trans-modal affordances given by a fully integrated multimodal system, the user control module is regarded as the main logical kernel in JOINDER. Consequently, its architecture is structured around a human centric modality division.

This, as it will be further developed in Section 2.2, will facilitate both embodied sound exploration processes and multiple user collaboration across different levels of participation.

In addition, the proposed solutions are mostly based on technologies that present limitations regarding the reuse, extension and exchange of the implemented prototypes, such as Max/MSP and Pure Data. Therefore, despite the fact that these environments offer a very efficient prototyping platform for real-time audio (and, in some cases, 3D visuals) through their interactive, visual programming paradigm, they fall short concerning the above stated requirements of portability, flexibility and integrability. In this thesis point of view, the latter are considered fundamental for a system to be adopted and further developed. The same type of criticism can be formulated with respect to the SuperCollider3 environment although, here, the main problem is the specific programming knowledge required to take advantage of his template system.

In summary, although acknowledging the efficiency of these systems in their own domain, most of them should not be considered as the base technology for an immersive, multimodal software infrastructure which aspires to remain open to the rest of the IT world's developments. An exception to this last observation is the David Worrall's soniPy. By choosing Python as a development platform, the extensibility, availability and maintenance of this project are very well safeguarded. In fact, the wrapping as an extension model strategy developed in this thesis had unquestionably an inspiration source here. However, and besides the last point regarding the user's embodied role in the sonification process, the choice of Java for the current implementation is due to the popularity of this development platform (TIOBE index 2 versus Python's index 8), not only in scientific and artistic domains (ex. the Processing environment) but specially also in the business world. This is reflected, for example, in the difference of the volume of projects related to the two platforms in SourceForge (as of July 2012, the ration between Java versus Python projects is approximately 4 to 1). Further advantages concerning the use of Java as main technology will be presented in Section 1.8. However, the existence of several projects regarding the interoperability between Java and Python implementations (e.g. JPython [84], JPype [83], JEPP [81]) as well as their standards' compliance (e.g. SOAP) safeguards the possibility of a future collaboration between these two frameworks. The same applies, to a certain extent, to the above reviewed implementations.

1.3 Functional processes and electroacoustic techniques

As in music, it is relevant to establish a hierarchical context of communication when presenting multivariable data through sonification. For illustration purposes, consider the situation where three data variables are sonified at a given moment with, respectively, the C, E and G pitches. The presence of these pitches implies the existance of higher and intermediate levels of meaning, respectively, the major chord and the ones related to the intervals formed by the combination of the mentioned pitches. These emergent sonic meanings somehow suggests that the data is structured in a similar way. Consequently, the presented sonic structure should take into account all the additional levels of meaning with the same degree of importance as the individual pitches. This can be directly transposed with a concrete sonification example. Consider a pitch sonification of ten stocks' variables representing the assets of a given company. If one wishes to go beyond this immediate representation (e.g. hearing the pitch curves of each stock) and wants to analyze the stock set's trends and correlations (e.g. the company's financial health), the combined behavior of the stocks has to be implemented as well. And perhaps more importantly, the perceptual interpolation between these two levels of sonification should be addressed in order to convey a better comprehension of the dataset (e.g. which stock is determinant for the overall fluctuation of the company's assets).

These examples illustrate that sonification practices bring with them their own emergent meanings that should be taken under consideration as they can be used as an asset to successful data exploration. Although this work does not dispute the fact that music composition and sonification serve mostly different intentions and purposes, this reflection points to a link between these two areas. As such, the central hypothesis of this thesis is that electroacoustic music composition employs systematic data-manipulation processes that can be very useful for developing the framework of interactive sonification. As Delalande pointed out [37], there is a communality of processes in electroacoustic composition practice that concerns the relationship between the singularity and regularity of events used in the musical discourse which underlines their structural dependencies.

Although, as argued by Vickers, one can establish a close relationship between sonification and musical composition through a perspective shift [153], music is different than sonification regarding its origin and purpose. On one hand, music serves as a medium to convey the composer's artistic goal. Sonification, on the other hand, aims at a systematic auralization of the data's properties and behaviors. Furthermore, one of the main differences between musical composition and sonification is that the latter uses exclusively systematic processes in the conversion between data and sound [72]. However, this doesn't imply that musical composition only uses subjective approaches or unsystematic processes. In music composition, systematic technics are often applied in order to create discourse coherence and thus a workable environment for the composer to reach his audience. As expressed by Emmerson, "communication between composer and audience rests to some extent on a common code or at least common expectations and assumptions" [16, pp. 4]. As such, the original idea undergoes transformations according to certain guidelines (either newly created or already established ones) in order for the intention of the composer to be materialized in the musical piece.

Following the seminal inspiration, idea and other subjective motivations leeds a composer to undertake the creation of a musical work, a great part of an artist's endeavor is to apply his craftsmanship in order to convert the original thoughts into an art piece. There is an conceptual stage that is followed by a realization path. This realization path is often, if not always, supported by the application of systematic rules that allows the composer in converting his intention into sound. These systematic processes are typically disrupted by the inclusion of subjective elements and choices which are dictated by the composer's intuition and creativity. If not, the work would easily be interpreted as a sonification. For example, if one disregards the sound mapping choices as authorship (as one does in sonification), one can argue that a musical piece that results from a conversion of the Fibonacci series to pitches is not a musical piece but a sonification, in the sense that the content of the final result is dictated by the series and not by the composer. However, the use of systematic processes in musical composition allows for the composer to establish a common language through an idiomatic soundscape. This regular progression serves then as a platform for the insertion of the composer's own voice and so, the final work is separated from the technique (as a letter document is separated from a generic template). As Emmerson pointed out, "we must establish sufficient agreement to allow communication, but build in the ability to evolve to suit changing situations" [16, pp. 6].

The development and application of processes that allow the transmission of information using sound has always been a main concern of music composition practice. Particularly in the 20th century, several theories have been suggested for establishing a meaningful and coherent binding of individual sound streams or events. In the two main initial historical trends in electroacoustic music's development, namely the french Musique Concrete and the Electronic Music from Cologne, the search for ways of establishing relations between material and form is present in the theoretical and compositional production of their leading advocates, Pierre Schaeffer and Karlheinz Stockhausen. Although based on different backgrounds and artistic ideals, both authors developed highly regarded and widely adopted theories. Respectively, these theories address the dynamic relationship between sonic elements, their confining context and the interdependence of their perceptual relevance. For example, concerning layered sound objects, reference can be made to the "sound object theory" [27]. In regards to scope transposition mechanisms between the micro and the macro level of the musical discourse, the techniques of "integral serialism", "moment form" and "formula composition" [70] can be highlighted.

Besides these examples, one can further refer to Iannis Xenakis' stochastic processes [164] [69] [42] and György Ligeti's micro-polyphony [147] [8] [33] as guiding principles for structuring sonic objects. However diverse these approaches might be, they all address the same problem, namely, how to establish a unified context between hierarchical levels of communication that are exposed simultaneously through time. As such, it was often the case that the technique would then become separated from their composer(s) subjective musical goals and become a compositional devise in its own right.

For example, the so-called french spectralism is a compositional practice that spans over generations of composers, called the french spectralist school [124]. In this practice, "sounds and musical colors (timbres) can be sculpted in time to produce musical effects" [55, pp. 2]. Although being more "an attitude towards music and composition, rather than a set of techniques" [55, pp. 2], spectral music often relies on mathematical spectral analysis and other knowledge provided by the fields of acoustics and psychoacoustics [125] in order to devise interpolation processes that assist the compositional practice [128]. As Gerard Grisey, one of the historical founders of the spectralist movement, stated: "Spectral music offered a formal organization and sonic material that came directly from the physics of sound, as discovered through science and microphonic access" [66, pp. 1]. An illustrative example of this influence can be found on the emblematic electroacoustic work "Mortuos Plango, Vivos Voco" by the english composer Jonathan Harvey, where a spectral analysis of a bell is used for generating both the harmonic content as the sections' duration of the work [3] [71]. Consequently, if systematic processes are extracted from the above compositional practices and the information resulting from their use is not cluttered by the presence of subjective musical elements [26], the conditions presented in [72] concerning the definition of what constitutes a sonification can be respected. For example, the operations contained in Pitch Set Theory [57] or Boulez's chord generation algorithms [14] are systematic and reproducible as it can intentionally be used with different data or in repetition with the same data (yielding the same results).

In summary, the point of departure of this work is to analyze the suitability of electroacoustic compositional concepts within the interactive sonification domain. The goal is to apply these concepts as ways to manipulate sonic materials and structures in relation to multilevel data representations. As such, data-dependent hierarchical levels are generated that preserve their informational identity and significance. As highlighted in the work of Scaletti concerning the specification of the Kyma environment, the adoption of Schaeffer's concept of sound object as a base design directive is fundamental for allowing the manipulation of multiple levels of complexity under one unifying abstract structure [129].

Finally, it should be noted that (i) the techniques that operate on data offered by electroacoustic composition have a large field evaluation throughout music practice and, (ii) they often are founded in close articulation with the human psychoacoustic idiosyncrasies and embodied cognition. This last observation will be fundamental as basis for the extraction of interaction metaphors concerning the representation and access of sound objects through an embodied music cognition approach.

1.4 Electroacoustic techniques as guidelines for interaction metaphors

In this section, we present the main composers and their compositional theories and techniques from which the work in this thesis found inspiration for the creation of interaction metaphors. The application of the compositional techniques to sound material gives rise to intrinsic (relations within the musical work) and extrinsic (relations to non-musical experience) mechanisms of contextualization [135] [22] [161]. In addition, these compositional techniques present a wide flexibility regarding the syntactic representation of sounds [48]. As such, this work takes as premise that it is possible to extract from these compositional techniques general metaphors, which can be used as guidelines for making a given dataset more accessible and easier to mold according to the user's inspection goals. In the following paragraphs, we present the electroacoustic music theories that have been explored in this thesis.

1.4.1 Sound Object Theory

The theoretical work developed by Pierre Schaeffer has been recognized as an important source of concepts for sonification design [153][65]. As Grond and Berger stated, "the first level of the perceptual domain can be considered as the abstract object sonore constituted by idealized reduced listening" [74, pp. 373]. Indeed, the Sound Object Theory pioneered a rethinking of the general composition practice where (recorded) sounds constitute the point of departure for a musical work. This stands in opposition to the abstract foundations (e.g. extra musical concepts) on which the traditional creation of musical works was typically based upon. This process of gathering compositional material relies on acousmatic and reduced listening. By highlighting the pivotal role of the human listening and defining the sound object as "the meeting point of an acoustic action and a listening intention" [27, pp. 27], the separation of perceptually relevant sonic artifacts from both its (physical) origins and its acoustic constitution is emphasized. This provides an important contribution for a new type of understanding of how we listen [159].

Despite Schaeffer's goal of generating purely musical building blocks from captured everyday sounds and the distinctions made between everyday and musical listening (as expressed in [61]), the sound object theory can be useful in providing guidelines for the constitution and classification of metaphors for sonification. Of particular interest are metaphors that apply to the idea of extracting meta-information present in a given sound[65]. By separating the sound from its source, these metaphors can be applied to events that share perceptual equivalences according to the listener's own perspective or viewpoint [89].

Consequently, this led to Chion's formulation of causal (which inquires about the sound source's information) and semantic (which relates the sound to semantical and linguistic codes) listening modes [153], as an expansion of Schaeffer's work. According to Chion, reduced listening "does not consist in denying natural perceptions (...) but in placing them in a new perspective" [27, pp. 29]. This view is also shared by Emmerson who states that "it proves very difficult to hear sound only in terms of an appreciation of its shape and spectral properties as Schaeffer seemed to advocate." [50, pp. 136]. By putting the focus on the listening subject and not on the listened object, it is possible to conceive new sound mappings and sonic interactions that present a higher degree of significance to the user and his expertise. Even if one's intention is to identify the source or cause of a sound object, the preliminary separation proposed by reduced listening enables a deeper evaluation to what is perceptually relevant to the user. As stated by Chion, "attention of reduced listening can use what it knows about the event, or even the meaning, the better to understand how the

object is made" [27, pp. 32].

Furthermore, the identification of sound objects through recursive exploration of its structure is of special interest here. Schaeffer's analysis attributes an identity to a sound object only as a member of a structure and to a structure only as a combination (set) of specific objects (See Fig. 1). As stated in [27, pp. 58], "every object of perception is at the same time an object in so far as it is perceived as a unit locatable in a context, and a structure in so far as it is itself composed of several objects". Concerning multi-level meaning representation and the interpolation between these meanings, this symbiotic relationship establishes a dynamic, perspectival dialogue between levels of sonic presentation. As such, the identification of a sound object in regards to a given structure and the description of a structure based on its (sound objects) composition both become central concepts in top-down decomposition and bottom-up composition processes.. They form the basis for analyzing the morphological structure of sound objects and the informational value that they might carry (concerning

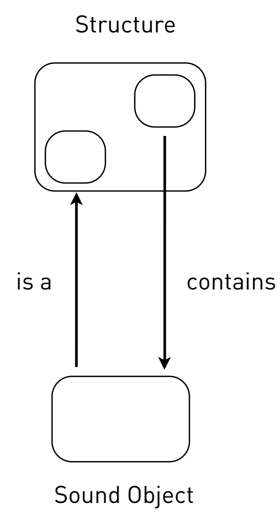


Fig. 1 The sound object and structure relationship in Schaeffer's Sound Object Theory.

source and semantics) in multiple levels. Identification and description unveil the affinities between the micro-structure, sample level and the macro-structure, dataset level.

1.4.2 Material/Form Transposition

In his compositional practice, Karlheinz Stockhausen devised several strategies for a unified control of the musical parameters present in his compositional works. Early in his practice, Stockhausen was an adept of integral serialism, which is a compositional technique where all parameters of a musical work are developed on the basis of a simple series of values. Based on the system invented by Schoenberg and his disciples for dodecaphonic pitch material generation, as well as Messiaen's compositional ideas presented in "Mode de valeurs et d'intensités", integral serialism became widely adopted and expanded by so-called 1950's avant-garde composers such as, besides Stockhausen, Pierre Boulez, Luigi Nono, Bruno Maderna, Luigi Dallapiccola and Luciano Berio. Through matrix-based permutations and transformational processes, compositional material is obtained by regular derivative procedures which would then be applied in melodic, harmonic, duration, dynamic, attack and register composition [18][19]. In this context, Stockhausen used serialism as a way of thinking as well as an algorithm for establishing a bridge between two intuitively defined values (as in, for example, the logarithmic metronomic scales) [29].

However, given the fact that the development process of each musical parameter (pitch, duration, articulation....) was perceptually unrelated as it was merely based on mathematical parameterization procedures, Stockhausen saw in the available electronic studio an opportunity and the technical means for controlling musical structures and their parameters under one single compositional principle [28]. He observed that by accelerating the playing speed

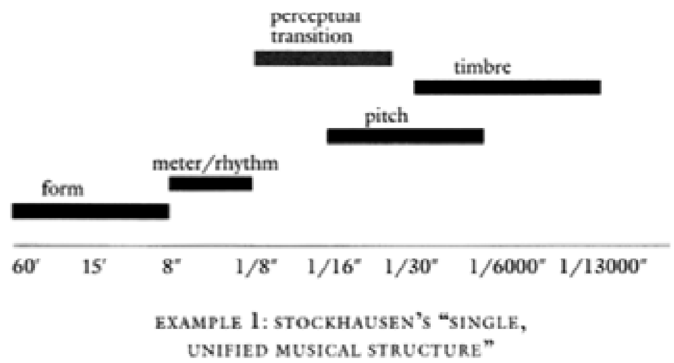


Fig. 2 Stockhausen's unified musical structure (as depicted in [28, pp. 224])

of a recorded series of (rhythmic) pulses, different pitches could be perceived. As such, Stockhausen "was able to make transformations between pulses and frequencies, rhythms and timbres in a single continuous movement, as if part of one spectrum" (which led to the development of a pitch-tempo scale [71]) to even incorporate form, which is "is rhythm expanded to a larger time-scale" [28, pp. 223]. This idea for parameters' unification was applied in the composition "Kontakte" for electronic sounds (see Fig. 2).

Although to some extend he was always relying on serialism as a general construction principle, Stockhausen developed several mechanisms for structuring his works. Of such compositional techniques, one can highlight two interesting techniques called "moment form" and "formula composition".

The technique called "moment form" [141] is based on a non linear distribution of "gestalts" units (known as moments). These units are set apart by their so-called gestalt character (defined in terms of melody, timbre and duration), which results from the interconnection between points (that represent a set of sound parameters) and groups (collections of points brought together by a given quality in terms of statistically measured direction, range and density) [29].

"Moment form" arose from the need to devise a new form that would be coherent with serial material [95] and the pitch/duration unity described above [38], as the above mentioned work Kontakte was one of the earliest pieces where he applied this concept. A detailed exposition of this technique's application in the emblematic work "Momente" can be found in [138][139].

The other technique, called "formula composition", also presents a methodological evolution from the musically sterile material offered by twelve tone rows. Being already music itself, the formula functions as a mechanism for "the pursuit of the highest possible level of coherence between morphological microstructure and musical macrostructure" [103, pp. 64]. Consequently, "formula composition" was used by Stockhausen in almost all of his works since the beginning of the 1970's. As an example, one can refer to his over twenty-nine hours long opera cycle "Licht", which is based on a three-part, eighteen-bar only score. Being comprised of three twelve tone, melodic lines, the formula is subjected to serial transformations (such as inversion and retrogradation) as well as vertical aggregation methods and rhythmic conversions. This provides the base material for projection processes, which are used to generate"all levels of the composition". As such, the formula "At its slowest speed, (...) provides the background structure for all of Licht" [92, pp. 274]. These processes underlying the formula-based composition of "Licht" are described in [91] and [92].

Similar to "moment form" and "formula composition", it is possible to deduce metaphors in relation to Stockhausen's transposition of information theory's concepts into musical composition practice [70]. By basing the musical discourse on the sequence and identity of sound objects, a technique such as moment form is in close affinity with the principles of similarity, opposition and belongingness that are known from Gestalt theory [17]. Furthermore, such techniques establish a close link with the idiosyncrasies of the human auditory system, such as the precedence of relative over absolute relations in parameter discrimination. Such gestalt-based processes constitute the core of, for example, "Mikrophonie I" [140], a work that employs a tam-tam, the active use of microphones for capturing the execution and the processing of the captured sound via bandpass filters and potentiometers. Among others, "Mikrophonie I" is one of the works where the main structural strategy is based on the amount of identity variation of a given gestalt segment or "moment" in relation to the previous ones. Furthermore, by using a single sound source, "each sound and gesture in the work is audibly related to every other, and all contribute to a resonant image of the whole, the tam-tam itself". Consequently, "Mikrophonie I" "is not only theoretically integrated in form, but aurally integrated in sound" [106, pp. 101].

In summary, the presented scope transposition mechanisms illustrate the application of consistent techniques for uniting different levels of sonic perception (see Fig. 3). Furthermore, as these methods are designed to highlight the variational's rate and the quality as well as the character of a given sound material, these procedures can provide powerful guidelines for articulating the presentation and encoding of interrelated levels of sonification [64], from sample (as material) to dataset (as form).

category	*elements*	*parameters*
sound	points	pitch, duration, dynamics, color
statistics	groups	direction, range, density
form	moments (forms)	organization, tendency
transformation	processes	change
composition	formula	elements (nucleus, accessories)

Fig. 3 A summary of Stockhausen's techniques (as depicted in [29, pp. 215])

1.4.3 Dynamic Morphology and Sonic Landscape

According to Trevor Wishart, the concept of dynamic morphology can be applied to a sound object "if all, or most, of its properties are in a state of change" [161, pp. 93]. In this definition, perceptual properties are emphasized over physical parameters as "it is important to view sound objects as totalities, or gestalts"[161, pp. 93] and "the qualities of the processes of change will predominate in our perception" [161, pp. 94]. Such qualities are considered by Wishart to be highly relevant in our perception and apprehension of the relationships between sound objects. This is further linked with the concept of gesture, which Wishart defines as "an articulation of the continuum" [161, pp. 17]. He believes that similarly to the dynamic morphology of objects, gesture cannot be separated into discrete components as "it is essentially a time-varying property of a whole sonic object" [161, pp. 112]. The similarity between dynamic morphologically described sound objects and gesture is considered to be fundamental for why the latter "can be applied to the analysis or control of sound-objects which are varying in a continuous manner in many dimensions of the continuum" [161, pp. 112].

Interestingly, the conceptual affinity between dynamic morphology and the relevance given by Stockhausen to time-based identity variance (see above) leads Wishart to highlight "Mikrophonie I" as "an interesting example" [161, pp. 104]. This remark is related to the rendition and control of dynamic morphologies, namely, in the score's use of morphological and gestural description of the sound objects (e.g. scratching, cracking, whimpering) and the use of relational notation for establishing a coherent discourse throughout the work ("The development and interaction of the individual parts is organized according to a simple set of relational and transformational symbols" [161, pp. 116]).

Closely related with the gesture and dynamic morphology concepts is the redefinition of the notion of sonic landscape [160] (see Fig. 4). Sonic landscape refers to a conceptualized sound source and not the actual, physical one. The advent of sound recording introduced the existence of a virtual acoustic space that offers new creative possibilities for musical creation as "the virtual landscape introduces a new way of sound representation and recognition" [16, pp. 110].

The new creative possibilities can be obtained through variations in aural perspectives during the capturing process, which is implemented through proximity changes of the microphone to the targeted sonic sources. This gestural interplay, which is analogous to the one of a visual magnifying glass, is carried out by the listener and allows the sound object to transcend its physical sources, becoming an entity in itself (as in the sound object as defined by Schaeffer). Once the capturing process is carried out, a vast range of electronic post-production manipulations are available to the composer. For instance, the nature of the acoustic space can be altered by applying a simulation of depth and dynamics to the stored sound object, further detaching it from its physical source. This is achieved, for example, through the manipulation of attributes such as amplitude, reverberation and filtering of higher partials.

Consequently, the created "distance" between the user and sound object can become a tool for foreground/background interpolation within the auditory scene as well as a group/individual comparative analysis of sound objects. Keeping in mind that "the translation of performance-gesture into the gestural-structure of the sound object is most complete and convincing where the technology of the instrument does not present a barrier" [161, pp. 17], the "distance" metaphor entails new possibilities for representing sonic events and contexts within interactive sonification and auditory display when using the appropriate technological means to implement it. If that is the case, this concept can constitute the backbone of user centered inspection of a dataset. The application of the microphone metaphor can make a sound object's content easier to perceive and, consequently, enable a more straightforward comprehension of both its enclosing context and his role in the latter (in relation to Schaeffer's object/structure analysis).The process of continuous feedback, enabled with scope variation (for sound object parsing) and simulation of depth (through amplitude and reverberation), allows the combination and segregation of sound objects (and the data they represent) within a multi-perspectival analysis of environment (dataset). As noted by Paine [117], both the concept of gesture and sound object's dynamic morphology are intimately related and equally relevant if a system aims at fully incorporating the user's gesture within an interactive, real-time operation.

To sum up, this dynamic exploration can contribute to a clearer identification of sound objects through their participation in a sonic context as the user's gestural activity puts forward his interpretation of the information present in the data.

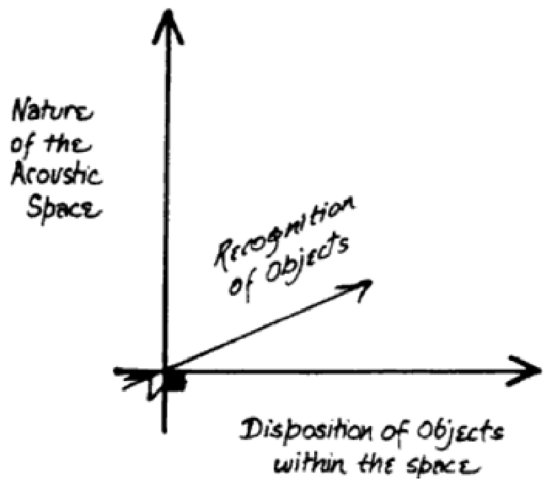


Fig. 4 The sonic landscape components' axis (as depicted in [161, pp. 140])

1.4.4 Spectromorphology

The spectromorphology classification [134] [136], as defined by Dennis Smalley, "is an approach to sound materials and musical structures which concentrates on the spectrum of available pitches and their shaping in time." [134, pp. 61]. Although its unified application in both material and form is an ongoing research topic [49], spectromorphology presents one of the most complete analysis and production frameworks for electroacoustic music's practice as "Smalley's examination becomes a systematic re-formulation and enlargement of Schaeffer's affirmations, preserving some of their original characteristics and conferring generality on many aspects" [22, pp. 6] and the terminology used by Smalley seems to be in tune with "the perceived attributes that people use to classify these kinds of sounds" [21, pp.588].

In order to achieve a fruitful communication through music, Smalley highlights the importance of aural perception as "Listeners can only apprehend music if they discover a perceptual affinity with its material and structure" [134, pp. 62]. This perception is strongly linked with semantic associations with the non-musical realm. As such, the relevance of the intrinsic and extrinsic processes of the electroacoustic listening are emphasized in his theory. As stated by Smalley, "the intrinsic approach emphasizes formal relations within a work ('music is itself'), while the extrinsic approach concentrates on relations with non-musical experience" [135, pp. 105]. Consequently, "If you inquire into a listener's response to a sound or musical work, trying to elucidate what it is that attracts or repels, it is impossible to avoid extrinsic references, such is the nature of verbal communication. However, identifying extrinsic relationships will not of itself uncover the meaning of a sound-event or work. In order to explain extrinsic workings and qualities we shall need to focus our attention on an intrinsic analysis of spectro-morphological features and their structural context. In other words, the extrinsic is determined by the intrinsic and vice versa" [135, pp. 105]. The degrees in which gestural sound objects are related with external sources or causality are defined by Smalley in a "surrogacy" scale, through which the degree of recognition and specialization of the sound body is possible [136]. At the same time, Smalley recognizes the importance of reduced listening (see Sound Object Theory) as a means for analyzing the characteristics of a sound object and assessing its potential for musical communication: "to find out what happens in the life of a sound or sound structure (...) we must temporarily ignore how the sound was made or what caused it, and concentrate on charting its spectro-morphological progress" [134, pp. 64].

Throughout the exposition of his theory, Smalley elaborates a framework for the classification of sound objects by classifying the types of spectral content through a noise-note continuum (constituted by noise, group of notes and the harmonic/ inharmonic note) (see Fig. 5), the types of morphologies (by describing the types of attacks present in the sounds), the models of interconnection (morphological string) and the types of motions (sound's directionality) [134] [22]. Furthermore, by acknowledging the importance of the physical character in the definition of sound objects, concepts such as gesture and texture constitute the foundations of his spectromorphological classification [136]. Using these concepts, he addresses sonic attribution, deployment and articulation within a comprehensive framework. In his classification, gestural energy-motion trajectory (defined by the onset, continuant and termination stages) is viewed as a wrapping mechanism concerning the perceptual grouping, scope transposition and translation of sound objects.

This work considers that the above concepts provide a rich source for interaction metaphors in the context of sonification. Using these concepts, it is possible to think of sonification in terms of an interactive process in which

the user may control the exploration of sounds through spectromorphological feedback.Moreover, the user may control this aspect through gestures. As such, in interactive sonification, upon interactively investigating a dataset, the intentions of the user can be transmitted through gestures. This possibility can further expose or highlight behaviors in the data, as the realtime feedback information may be used to fine tune the way in which the user explores his goals or uncovers unexpected phenomenon in the data. This information may be of use to fine tune the way in which the user explores goals or reveals unexpected phenomenon in the data. As such, the combination of gestural framing (actions upon the data) and sonic texture setting (scope of the actions) can pave the way for a spectromorphological analysis of a dataset through an interactive and embodied sonic exploration.

Finally, it is important to mention the notion of space as a multiple viewpoint provider for the perception of sound objects. In his Spatiomorphology [136] analysis, Smalley highlights the affordance of multiple perspectives as a base component of sonic perception and cognition as "listeners can only become really aware of the variants if they have had an opportunity to compare perceptions of the same work under different listening conditions" [136, pp. 122]. In the space-form analysis [137], the spectral content-driven classification is extended as the concepts of scale, viewpoint and resolution are connected with cross-modal stimuli integration and structure analysis. In this context, Smalley points out the use of immersive spaces as an optimal scenario where "the listener gains from adopting, and is encouraged to adopt, different vantage points (...) freeing aural elements from continuous, concentrated scru- tiny" [137, pp. 52]. Also here, interactive sonification can be informed by electroacoustic practices concerning the use of space. In particular, the use of immersive environments can be promoted as an ideal technological support for this type of sound-based exploration.

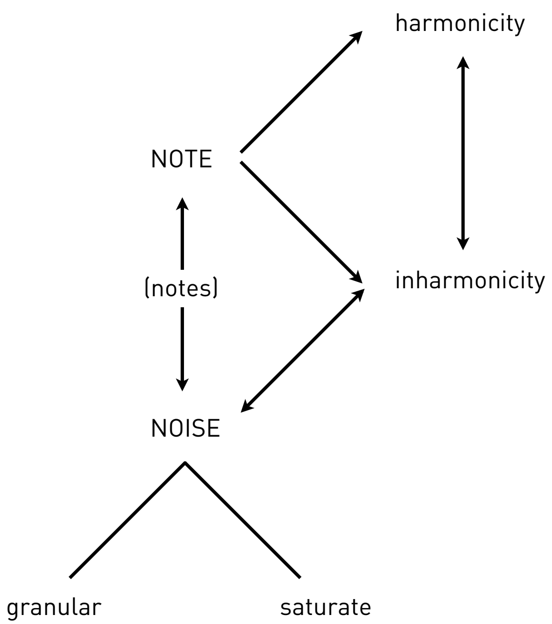


Fig. 5 Smalley's Note-Noise relationships (as depicted in [136, pp. 120])

An in-depth comparative discussion between the spectromorphology and space-form classifications and interactive sonification design is presented in Section 2.3.

In summary, the above described electroacoustic music's theories and practices are adopted throughout this work as guideline providers for the representation and access of sound elements and the data that they illustrate. It is important to mention that these guidelines are all rooted in gestalt theory, as the base applied principle is that sound in itself is multilayered and therefore, can be accessed from different viewpoints. The two main gestalt principles that underlay the above exploration are, first, that representation is conceived from a multi-level perception and meaning viewpoint [17], and second, that access to data relies on a multimodal and dynamic participation of the user in the interaction process.

By acknowledging as well as spectromorphological potential of sound, it is possible to expand sonification design on the relationship between sound objects and their enclosing structure. Furthermore, this analysis underlines the concrete and abstract duality and, consequently, the inherent property of electroacoustic music's compositional practice to symbiotically define and be defined by external elements. Such dependence leeds to the possibility of applying this relationship in interactive sonification research as a dataset can be sonically represented by multilevel sound objects in a structurally coherent, perceivable and thus meaningful way.

1.5 Embodied music cognition, virtual object mediation and interaction metaphors

In order to implement the metaphors described in the previous section, the proposed strategy of interaction follows an embodied music cognition perspective [99]. In this approach, the involvement of the human body in sonification assumes a technology that mediates between body and data. Corporeal mediation technologies are conceived as tools that extend the human body as a kind of prosthesis in order to allow the human explorer to mentally access the digital realm in which the data resources are assessed. Consequently, this methodology can be of service for deploying interfaces that provided an integration of the variable resolution of the human motor system and sound objects [63].

The implementation of this approach is here based on three design directives. They constitute a common ground within human perception mechanisms and software tools through which natural communication can be explored and implemented in an ecological scenario. The design directives are direct manipulation, object-oriented paradigm and virtual object mediation.

1.5.1 Direct manipulation

Directly manipulation [34] is here applied since it focuses on making the target objects that constitute the interface accessible in a way that mimics our natural, physical interaction with real world objects. It brings forward an inspection process that is intimately related with human everyday experience when analyzing their surrounding environment. Actions like grabbing, turning, shaking, hitting or molding are a part of investigating procedures that we have learned to conduct since birth. By using a multimodal approach based on an embodied exploration, one can activate multiple sources in an natural, intuitive way.

In electroacoustic music, the concept of direct manipulation as method for realizing a musical intention is commonly used. For example, one can point out the expressive use of the mixing board's faders in both the mixing and spatialization processes. It is a trivial but nevertheless good example of an embodiment-based discourse that incorporates the physical/gestural factor in the creational process. Even more interesting is the fact that this process can be used across different levels of granularity throughout the work, ranging from the individual amplitude envelope of a sound object to post-production panning of entire sections. Additionally, this procedure can also be found in realtime performance practice. For example, as Stockhausen commented on his work Mikrophonie I, "Someone said, must it be a tam-tam? I said no, I can imagine the score being used to examine an old Volkswagen musically, to go inside the old thing and bang it and scratch it and do all sorts of things to it, and play Mikrophonie I, using the microphone" [141, pp. 89].

1.5.2 Object-oriented paradigm

The adoption of an object oriented paradigm is founded on the composed/hierarchical characteristics that are shared by the entities involved in the interaction process, namely, sound and spatial/visual objects [149]. If interconnected, these structural similarities can positively contribute to further expand the user's cognitive processes by adopting a coherent set of representations and actions within a gestalt theory framework [118]. Also, object oriented interfaces are closely related to direct manipulation approaches since the entities that are the target of direct manipulation procedures are easily defined through an object oriented paradigm. As such, an object oriented interface guides the user's intention by the explicit manipulation that it affords. Its conceptual nature drives on the easy identification of an operation's results via the instant feedback of a given object' state change. Consequently, this work aims to provide an architecture that defines the way components relate and interact. As a result, the system should convey a platform for user-centered, rapid application development by reducing the effort for development of customized applications and by providing a state management. For that purpose, the application of object-oriented design patterns is favored in this environment as it improves quality, reliability, and interoperability of software projects by promoting access and communication standardization between components.

1.5.3 Virtual object mediation

A central concept in the presented interaction system and consequently in JOINDER framework concerns the use of virtual objects to represent both the users and the data within a three dimensional environment. Following Mulder's 3D virtual sculpting as a metaphor for sound synthesis [108] and previous applications of immersive interfaces in sonification [67] [15] , the main concept is that users can access the data in a way that mimics their natural interaction with real world objects. This directly manipulation strategy allows for multiple mapping layers, different viewpoints and degrees of resolution [129].

The superimposing strategy of real and virtual world three dimensional coordinates in the same interaction space, allows to take advantage of the natural constrains of the user's body, according to personal space and extra personal space concepts [101]. Therefore, much of the work presented here is conducted within immersive 3D environments, as it provides an optimal way for the representation of data through real world geometry [148]. Finally, virtual objects mediation and Virtual Reality can be of service in the development of user interfaces by simulating their behavior and functionality without requiring a real life prototype at early stages of development. As a result, the virtual environments can provide a set of functional mockup facilities for prototyping either alternative output/input geometries and activation strategies.

By enabling a configurable location and form representation of the data in space, these methodologies invites the user to adopt a physical approach for the inspection process through a shared space of multilevel interaction. As visual, sound and tactile modalities are integrated within a spatial domain, an embodied cognition approach is expected to further enable a perceptual link between the data under inspection and the semantic high level representations of the user.

1.6 Interaction metaphors

As mentioned previously, the interface paradigm adopted is based on the interaction with virtual objects as a way for accessing the variables belonging to a target dataset. In this section, a summary of the interaction metaphors based on the electroacoustic theory and practice are presented. Our approach is divided into two symbiotic main aspects, called multilevel mapping and scope variation.

- **Multilevel mapping**
  Multilevel mapping aims to aurally illustrate the different degrees of granularity of the sonic landscape generated by a given dataset. It focuses on how to convey information about both individual samples and whole of the collected data (ex. a dataset containing the daily temperatures registered in a given location throughout a month). As stated previously, the activation of these levels should be in conformity with the variable resolution of the human motor system [63].

- **Scope variation**
  Inherently related with the previous metaphor is the scope variation strategy. Scope variation presents a way of grouping a set of sound objects or, in an inverse perspective wise, of analyzing its constituents, by manipulating an activation radius. This zooming facility provides an intuitive support for the our natural methods of processing multi-level information.

Multilevel mapping and scope variation are at the heart of the process for designing and implementing versatile and scalable interfaces that can convey both representation and access using real world metaphors for the active molding of the sonic output. These aspects led to the two main interaction metaphors' categories, called the virtual inspection window and virtual inspection tool (see Fig. 6).

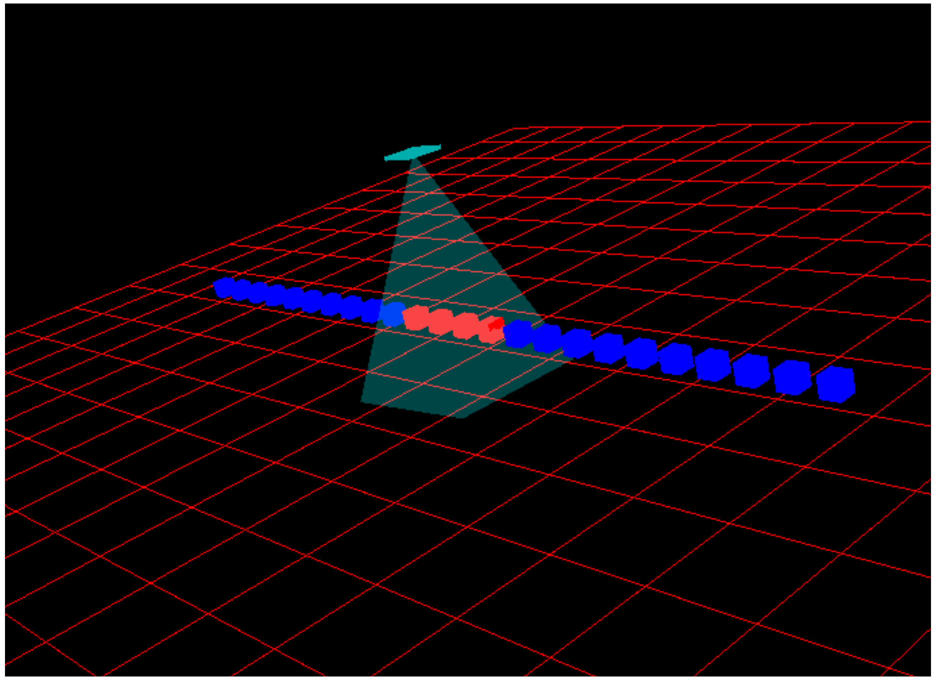


Fig. 6 An early implementation of the virtual inspection window (array of "blue" objects) being activated ("red" objects) by the virtual inspection tool (pyramid-shaped transparent object).

1.6.1 The virtual inspection window

The virtual inspection window offers a generic metaphor for generating access points to sonic entities and the data they portray. This strategy allows a versatile spatial rendition as sonic objects can be represented through any morphology that best fit its origin and the user's preferences. As such, data (through sound) and user are placed in a common, dialogue prone scenario that maximizes embodied cognition through physical interaction, thus alleviating the cognitive load of the specialist. Furthermore, this approach allows the illustration of time dependent variables. For example, a set of virtual objects can be used as a timeframe representation, following a first in, first out methodology: $w(t) = [v(t), \ldots, v(t + \Delta t*size(w))]$. This functionality can be of service in numerous contexts.

For example, consider a dataset containing last year's average day temperatures in 5 cities of a given country is to be sonified. Instead of representing every value for every variable simultaneously, which would easily cause a congestion of the virtual scene and, consequently, difficulties to the inspection process, the representational virtual objects can be configured to allow simultaneous access to a subset of these values. For instance, an array of 30 cubes can be assigned to each variable constituting a temporary access to a period of approximately 1 month. The remaining values of each variable can then be accessed through the "sliding" of the inspection window, controlled by an auxiliary device (Ex. a WiiMote). Finally, the 5 arrays could then be placed in various arrangements in order to allow multiple views of the dataset's content.

Additionally, this approach constitutes a viable option in the analysis and comparison of real-time data. The values can be made available to the user for a certain amount of time (depending on the generation rate) and then "hidden" from him, being available for later inspection. Moreover, the morphology of the virtual object(s) can be data dependent, conveying a more informative visual and spatial representation of the values being analyzed. This dimensionality reduction strategy represents an evolution from the one variable per axis approach found in other sonification explorations within immersive environments [15]. In summary, variables can be represented and accessed through multiple windows (with different time intervals, for example), multiple morphologies (linear, circular, ...) and dynamically assign to different sonic representation through multilevel mapping.

1.6.2 The virtual inspection tool

Virtual inspection tools implement a metaphor for physical and gestural sonic activation of the scene's virtual objects and the data that they represent. It materializes a virtual microphone tool that can be used to inspect the virtual world's objects through a spatial, user centered approach. It allows the user to zoom in and out in order to investigate either one element's output, or its relationship with other members of the set. This scope variation enabler process was strongly inspired by Stockhausen's Mikrophonie I [140] (composition where the active use of microphones is a base concept in the performance of the piece), Wishart's analogy of the visual magnifying glass to the microphone (see Section 1.4) and is related to previous research presented in [158] and [67].

Besides collision detection and distance threshold based activation, it can be used also for controlling depth simulation effects and sonic modulations. Additionally, the activation procedure can be further extended within the physical interaction simulation domain, by taking into account the velocity and direction upon collision and/or gesture's typology (e.g. scratching, shaking, touching, hitting, etc). It can be use with multiple instances (N microphones positioned at different points of the user's body), each paired with one or multiple sonification levels. Furthermore, a global (all microphones can activate all windows) or selective (microphone/window pairing) activation strategy can be explored. Finally, its basic principles can serve for configuring the position, orientation and morphology of the virtual elements. Going back to the example described in the last subsection, the inspection window (the parent object) was composed of 30 spheres (the child objects). Upon activation, the individual object's parameters are fed into the sonification engine. Although the overall sonic output is implemented by the individual elements, it conveys information about the activated set as a whole. Following the previously referenced theoretical guidance, it stimulates a perceptual interpolation between the whole (a month) and the individual nodes (the days).

Moreover, through the use of the inspection tool and the spatial arrangement of the inspection window's elements, the user can group several consecutive "days" and have an "on the fly" composed sound object which conveys the progression of the temperature in one city. On the other hand, several "days" from different cities can be grouped, sonically illustrating the relations between different locations' temperature reading. Being so, the inspection interface paradigm can convey multiple perspective views between different levels in the time domain, by representing the evolution of a variable in time, and in relationship, by establishing and comparing different groups of N variables.

The presented paradigms implement the representation and access of multiple sonic mappings to the virtual element(s). For example, in order to illustrate relationships in the data concerning the previous example, several sonification levels can be superimposed (ex. intervals and chords in addition to individual pitches). Either enabled through distance, collision or realtime configuration, sonification levels can be accessed or assigned to any virtual inspection tool. This strategy expands the multi-perspective analysis possibilities offered by the virtual microphone metaphor and conveys further support for the researcher/composer/performer's expertise and intuition while conducting his task.

These above interaction metaphors were implemented in both prototypes interfaces, namely the "dataset" sonification use case [44] [45] and the "dance" sonification use case [43]. In both cases, the virtual inspection window and tool were used to explore variables' behavior through multilevel mapping and scope variation approaches. Following a demonstration setup where users were invited to interact with a virtual string placed in the middle of a room, the first use case comprised of a more elaborate prototype. A test dataset was represented and explored through a variable number of virtual windows (each window was implemented by an array of fifty virtual objects), inspections tools (using one or both hands of the user) and mapping levels - pitch (one object), interval (two objects) and chord (three or more objects). These mapping levels were assigned to the virtual windows in various combinations (ex. all active in one window, one active per window, etc..). Furthermore, the virtual inspection tool provided both the activation and a distance based modulation of the sonic representation of the dataset (through reverberation and amplitude setting).

The second use case ("the dance use case") applied the virtual window metaphor for illustrating the spatial and kinetic parameters of a dance choreography. Given the scattered nature of the resulting path's object array, the virtual inspection tool was upgraded in order to include radius variation. Being adjustable through the opening degree of the user's arm (measured by a ICubeX bend sensor), this feature allowed the adjustment of the number of activated objects and conveyed a method for interpolating between the sonic representation of the whole path and segments of it. In summary, all interactive sonification iterations provided a study regarding the users' initial reception to immersive virtual object mediation, the coherence within the combination of different metaphors and the system's response to "on the fly" configuration and performance requirements. A detailed description of this research can be found in Section 2.1.

In a later stage, these metaphors were used and further expanded during the SoundField project [31], an interactive art installation developed in collaboration with Pieter Coussement and Alexander Deweppe. During four months, on site development period, the Soundfield project generated multiple interfaces for supporting various artistic disciplines' workshops (such as dance, literature, music, theater, sculpture and audiovisual arts).

Among others, the interfaces developed comprised of:

- a soundscape generation interface based on users' collaboration (where each user controlled a sphere object and through their proximity, other objects - strings and planes - were generated and played),
- a variation of the dance use case path metaphor for two dance performers (where the sonic output was generated by the second performer which accessed in realtime the path laid by the first dancer in a contrapuntal style),
- a grid of virtual objects for melody or sequence of sounds recreation,
- sound sculpting and audio/visual selection and mixing interface based on three dimensional absolute position.

Focusing on collaborative interaction [30], these iterations ranged from sound manipulation to theatrical audiovisual augmentation and multimodal rehabilitation tasks. As such, the SoundField project provided an opportunity for the expansion of the presented methodology and metaphors as well as its validation as a useful mean for the development of embodied, virtual mediated interfaces within an ecological setting. A more detailed description of this project can be found in Chapter 2.

1.7 State of the Art 2 - multimodal technologies

As mentioned previously, the majority of the toolkits/frameworks developed for (interactive) sonification and auditory display relies either on interactive, visual programming environments such as Max/MSP and Pure Data or on specialized platforms such as SuperCollider3. It has been argued that, although these environments offer an excellent base for prototype development in realtime audio (and sometimes, visual rendering), they are often of a self-confined nature, used in a very circumscribe domain and demand a high level of expertise. These properties makes them ineligible for providing a solid and accessible main hub for multimodal management and routing environment that respect the portability and extension requirements stated earlier on.

As such, and given our virtual object mediation interaction strategies, our attention turns to the research and development conducted within the broader fields of multimodal interaction, virtual/augmented reality, distributed computing and environmental/ubiquitous computing domains. The following overview of the state of the art is based on several survey articles [51] [77] [119] [144] [46] as well as on original papers and websites of the reviewed systems (when available). The overview reveals a panorama of the technological efforts in the past decade. However, given the vast bibliography available and the wide range of target scenarios, the software packages that were analyzed are the ones considered closer to the focus of the present thesis. A table containing a reduced version of this discussion is available in Section 2.2.3.

- **Dwarf** [20] is a service oriented, CORBA based middleware. It is conceived as an interactive visualization environment for remote monitoring. It offers a component-based P2P model network for collaborating distributed services, being a runtime reconfigurable system. The services are interdependent and expose their "Needs" and "Abilities" as interfaces respectively for event sinks and sources. The middleware is lightweight enough to run on the Linux-based palmtop. Does not possess a global scene graph for scene management.

- The **AR-Room** [120] presents a set of deployable components for augmented reality technologies, modules for hardware abstraction and an authoring toolkit for the rapid content design. The content scenario is represented by a set of event-action pairs. It provides image analyzer, an interaction handler, a rendering engine and an image synthesizer. The base technology is C/C++ and it runs solely on Windows platforms.

- **Studierstube/Opentracker** [131] [127] offers a collection of C++ classes that extend Open Inventor's scene graph. It provides a distributed system with replicated scene graph and observer pattern implementation for propagating changes in the scene to all hosts (master/slave model). The session manager serves as mediator among hosts and known point of contact for newcomers hosts, providing also load balancing facilities. Further extensions must implement using Inventor objects and it runs solely on Windows, Linux and Android. Its final version is 4.4 and project maintenance stopped in 2008.

- The **ARToolkit** [87] provides a specialized computer vision library for analyzing camera's video stream. More specifically, it allows recognition for markers with their pose, identification and marker position and orientation in a data structure suitable for OpenGL-based rendering. It offers support for a wide range of platforms: Windows, Mac OS, Irix, Linux, Matlab and MS PocketPC. Although supplying Java wrappers, it is based on C API close to OpenGL.

- **Squidy** [93] is a recent framework (2010) with custom configuration visual environment. It implements a data flow (pipes and filters) and semantic zooming paradigm in a multi-threading system for multimodal device integration and versatile usage. It also provides run time configuration. It does not provide a state management infrastructure.

- **Unit** [113] is developed in Java and Java 3D graphical API (Application Programming Interface). It delivers a modular development environment for interactive user interfaces. It provides a semi distributed component through RMI over TCP/IP, a framework, API as well as visual, data flow programming of behaviors. However, it is closely coupled with Java 3D API.

- Based on data flow and pipes/filters paradigms, the **ICARE/Open Interface** [13][132] provides wrapping facilities for integrating components written in several languages (C/C++, Java and Python). Through this integration process, it allows the use of generic devices according to the interaction needs, leveraging expressiveness with reusability. However, being a .NET based system, its usage is restricted to Windows platforms.

- **JReality** [156] is an open sourced thread-safe scene graph API written in the Java language. It provides a separation of the scene graph and rendering backends, flexible attribute-inheritance mechanism in scene graph and device-independent user interaction through the Tool abstraction. It mainly uses JOGL as a visual backend and provides distributed rendering (for CAVE-like environments). For audio rendering, it supplies audio output through Java Sound or JACK and provides spatialization through stereo, VBAP, second-order planar Ambisonics, and first-order 3D Ambisonics. Although a very complete solution, its main architecture design does not provide full modality rendering independence.

- **YVision** [2] is a C#/CLI framework for deploying "natural user interfaces". It is Mono compatible, which allows a wide range of supported platforms. By assuming a component based architecture and implemention on a tree behavior implementation, it allows the replacement of rendering and behavior cores (e.g. Axiom and BulletXNA). Furthermore, it supports concurrent programming (MS Task Parallel library) and provides an input abstraction and management system. Finally, web application deployment is possible through the use of MS Silverlight. Although an interesting approach and Mono framework compatibility, it heavily relies of Microsoft technologies (both specification and implementation), which affects the portability and accessibility, nor does it provide a centralized management kernel.

- The **JMonkey engine** [82] is a open source Java based framework for 3D game development. It provides a very complete solution by including shaders, lighting, effects, multi format import, textures and animation features. It also provides physics integration (JBullet) and networking facilities. However, its main focus is on desktop visually-oriented game development as the other modalities (sound, human interfaces) present a much less level of development.

- **Max/MSP/Jitter** [107] and **PureData/GEM** [126] are popular environments in the creative coding communities, specially in the interactive music performance's systems development. Following a data flow visual programing paradigm, both platforms allow rapid prototyping of realtime multimedia applications and can be extended through externals written in C, C++, Java or JavaScript (Max/MSP). However, it lacks the organizational structure (as in Object Oriented programming) to fully develop extensible code. Moreover, Max/MSP/Jitter is commercially licensed.

- **Corsaire** [52] is a system developed through the combination of heterogeneous technologies in order to achieve high level multimodal solutions. It uses Pymol (python based OpenGL framework) for visual rendering, Open Sound Control for communication, distributed and component based sensor management via ARTrack, central management via VEServer and Max/MSP for sound rendering. It presents a component based architecture for command creation and execution, scoring, rendering and supervision as well as a rule based, context-sensitive platform. Although offering a modality component separation and the use of specialist tools for each purpose, it is unclear how the replacement of these components would affect the whole system.

- **Morgan** [112] is a C++ API for implementing AR/VR prototypes and applications. It provides its own visual rendering engine, synchronization between distributed scene graphs and a publisher/subscriber methodology. However, it is a commercial system and the inclusion of a default proprietary scene graph discourages the re-use of code already developed for another visual engine.

- **VHDPlus/VHD++** [123] is a C++ framework that provides a component oriented simulation engine and software middleware solution for the domain of VR/AR, particularly virtual character simulation. Its design's intentions reflect the need of a unifying "glue environment" for multiple technologies and respective features by establishing a common vocabulary between them. The system provides a generic, high performance runtime engine, set of pluggable service components for encapsulating heterogeneous simulation technologies and hierarchical data components representing both system and simulation states. The latter is used to form the runtime application graph and to allow concurrent mechanisms to operate on them. Finally, Python is use for rapid prototyping scripting. This project addresses one of the main concerns reflected in this article - the unification of heterogeneous technologies. Although having a respectable code base and development team, it is mostly oriented towards virtual human's simulations. Furthermore, it relies on C++/VisualStudio technologies as well on a considerable number of plugins. Lastly, it's distributed capabilities are unclear.

- The **InTML** [54] project focuses on providing a language for representing interaction techniques and applications in the areas of Virtual and Augmented Reality. It describes the high level elements of an application: references to VR objects, to devices, and to interaction techniques and proposes a common way to represent the latter with three main objectives: execution, understanding, and data consolidation. In short, it aims on adding a component-based, reusable architecture to Mixed Reality appications on top of open technologies such as VRJuggler, VRPN, and OpenSG. However, it doesn't describe the visual (i.e. geometry and texture), haptic or sound properties of VR objects, as third-party languages have to be used for such details.

- **Delta3D** [32] is an open-source C++ 3D engine for game and simulations development. It is available as a precompiled SDK for use with Visual Studio 2008 and as source code only (.zip). It aims at connecting multiple open source projects such as OSG, ODE, CAL3D or OpenAL into a custom graphics engine. However, it lacks single point configuration.

- **VR Juggler** [10] provides an abstraction layer between the hardware of a VR system (head-mounted display, projection-based system, etc.) and the virtual world created in software. It has no dependency on a specific rendering engine as it requires applications to bring their own. VRJuggler does support render distribution, but it has no direct high-level support for application distribution. Furthermore, it does not provide a custom scene graph for state management as it relies on the visual modality for scene representation, which undermines modality's independence.

- **Osgart** [105] extends OpenSceneGraph with the ARToolkit. It aims at providing a scene graph to marker-based AR. Furthermore, the design envisions the generalization of available technological concepts, supplying techniques for mixing real and virtual feedback in multiple spaces. However, it does not provide task specific utilities. It is written in portable C++ with Python, Ruby or Lua scripting.

- The **Virtual Choreographer** [76] is a open source 3D graphic engine with XML-based configuration. Its main target is interactive spatialized scenes for artistic design and multimedia scene. Being designed to specify complex geometrical scenes, it provides graphic and sonic components and facilities such as high dimension particle systems through GPGPU, MPEG4 facial and skeletal character animation, vertex and pixel shader support, audio source localization, and collision detection. It runs on Windows, Mac OS/X, and Linux. It uses OpenGL for interactive rendering and POVray or Renderman for offline rendering. Sound computation is achieved through OSC messaging with PureData or Max/MSP. However, it does not present a clear separation between the scene management and the visual rendering.

- **Avango** [97] is a platform developed in C++ for multi-sensory, scalable VR application development and rapid prototyping through Python (previously Scheme) scripting language. Through the use of reflection, all objects can be manipulated in a common way. It provides a concept of active nodes, sink and containers for dynamic event propagation as well as a straightforward design to incorporate new interaction devices. Furthermore, it supports distributed environments by transparently broadcasting scene graph nodes' information.

This is done by sharing the nodes' data between the different client applications viewing the common environment. Previously based on the commercial Performer/SGI library, it now encapsulates and builds on top of OpenSceneGraph. It is mainly developed for Linux (partial multi-platform capability is achieved through ports for MS Windows and Mac OS X).

- The **MRSS** [75] system integrates a collection of concurrent cooperating components. Although MRSS is not strictly tied to Mixed Reality, that's the focus here. The central component of the system is the MR story engine, a container for agents (actors), one for every user. This container stores virtual and real interaction objects and abstract agents that might be useful for the story line. Additionally, the system consists of three rendering engines for multimodal simulation (visual, audio, and special effects) and a fourth engine drives the integration process. The engines are interoperable with other commercial components. The network central node receives and integrates sensor data such as tracking, registration, and orientation. The key technologies used in this MR system are OpenSceneGraph and Cal3D for graphics and Port Audio for sound. Authoring of stories is done in XML, which can include C or Java-style advanced scripting. The MR system can run stand-alone (one user) or in combination with multiple MR systems (each managing one or more users).

- **Processing** [59] is an open source programming language and environment for creating images and animations. Build over Java, it was initially developed to serve as a software sketchbook within a visual context. It provides an authoring environment for creating interactive programs using 2D, 3D or PDF visual output. It uses OpenGL integration (through JOGL) for accelerated 3D and it runs on GNU/Linux, Mac OS X, and Windows. It is extendable, with over 100 libraries for sound, video and computer vision rendering. It is free and open source. It favors the vision modality (original target) and does not provide scene graph management facilities.

- Similar to Processing, the **Open frameworks** [114] project aims at providing an open source toolkit designed for "creative coding". It is written in C++ and it runs on Windows, Mac OS X, and Linux. It offers a simplified programming interface for media, hardware and communication.

- **Open Wonderland** [86] is a 100% Java open source toolkit for creating collaborative 3D virtual worlds. The users can communicate through high-fidelity, immersive video and audio as well as share live desktop applications. It mainly targets educational, business and e-government application development. Additionally, it is completely open for extensions by developers and graphic artists. Its technological foundations are Project Darkstar (platform for scalable communications and persistence in games), Glassfish (highly scalable, open source web server) and JMonkeyEngine (for visuals). Furthermore, it incorporates other libraries such as MTGame Graphics Engine. However, it provides an enterprise scale solution, which can be inadequate to small projects.

- **Syzygy** [130] is a programming toolkit for writing virtual reality or other graphical applications, especially designed for the creation of applications that run on clusters of networked computers. The rendering is synchronized on all hosts, giving different simultaneous views on the virtual world. The integrated utilities handle communication with 6DOF trackers as well as other input devices and spatially-localized sound is supported using the Fmod libraries. It runs on Windows, Linux, MacOS X, and Irix. Furthermore, it supports heterogeneous clusters as different operating systems can be mixed following a master/slave configuration. The toolkit's libraries are written in

C++ and applications can be written in either C++ or Python.

- The **CHAMBRE** [12] system targets the integration of real and virtual world sensors and actuators as well as traditional multimedia environments within a single architecture. It consists of a open framework able to accommodate different protocols, sensors, generation and interaction techniques. It supports multiple processing definition of the same input. This definition can be loaded in real-time and controlled by interaction with users or other software sources. It follows a component-based paradigm, where components are endowed with communication interfaces. These components can act as synchronizers among different inputs, providing a basic form of coordination through an observer/observable pattern implementation. The Chambre network is modified through the addition or removal of plugins. A plugin node defines communication, computational and output features as well as graphical user interfaces (GUI) for user interaction. It uses Java RMI for inter-node information exchange, Java Media Framework for multimedia streaming and Java 3D scene graph. However, it is closely integrated with some technologies, which might compromise modularity.

- **OpenNI** [116] is a API developed in C/C++ for writing natural interaction applications. It consists of a middleware that provides device abstraction via modules and a data processing design through production nodes and chains. It is multi-platform (Windows, Linux, Android and ARC) and it offers Java and C# .Net bindings. However, it provides no global scene management of the target classes involved.

When analyzing the previous descriptions, some of the outlined requirements for our research can be found in the referred systems. Examples are the interoperability oriented design (ex. VHD++), distributed execution (ex. Studierstube;

Corsaire), separation between application and rendering backends (ex. JReality), functionality encapsulation through command classes (ex. Corsaire) and software framework as a design choice (ex. VHD++).

However, the overview of the reviewed systems reveals that:

- Most of them provide development environments that focus on one main modality (ex. visual, in the case of game engines). Additional modalities are implemented as an extension of the original one (ex. visual) which often leads to a dependency on a particular engine. In this way, it is difficult, if not impossible, to use another rendering engine for the original modality. Furthermore, applications that use only the added modality (ex. sound) are difficult to implement without using modules of the latter.
- Most of them don't supply an independent state management solution. The ones that do present a form of centralized access to its elements (ex. scene graph), use the one provided by the main library/framework that supports the rendering of the main modality (ex. Java 3D).

As a result, the focus of this work turned onto the development of the JOINDER, a software framework for multimodal, immersive interfaces implementation. The JOINDER framework will be the subject of the next section.

1.8 The JOINDER software framework

The JOINDER (Java Omni Integrator for NetworkeD Embodied inteRaction) software framework implements our goal to provide a virtual reality based, object-oriented software environment for:

- Interconnecting different hardware interfaces and heterogeneous software environments.
- Developing multimodal, embodied interfaces that use space as their interconnection paradigm.

JOINDER offers a distributed execution design that conveys a multithreaded, modality rendering management, input devices administration and independent state supervision. The application of object-oriented design patterns is favored in this environment, contributing to improve quality, reliability, and interoperability of software projects. As such, the framework offers an architecture that provides element encapsulation through interface definition and reflection class loading in order to better enforce the modularity requirement. Furthermore, it presents a client/server architecture where the central node manages the three dimensional scene content for every element (e.g. visual, audio, tactile) and synchronizes modality and auxiliary (e.g. physics) cores. This implementation follows the state of the art's conclusions for the need of a new infrastructure that is compliant with the NSF report (e.g. portability, flexibility, integrability) along with reliability and realtime performance objectives.

Consider the following example of a JOINDER deployment in Fig.7. The central node (e.g. Virtual Core) provides the interconnection between several technologies. Using different binding levels (see below), both sound and visuals are implemented by distinct rendering technologies, respectively two instances of the

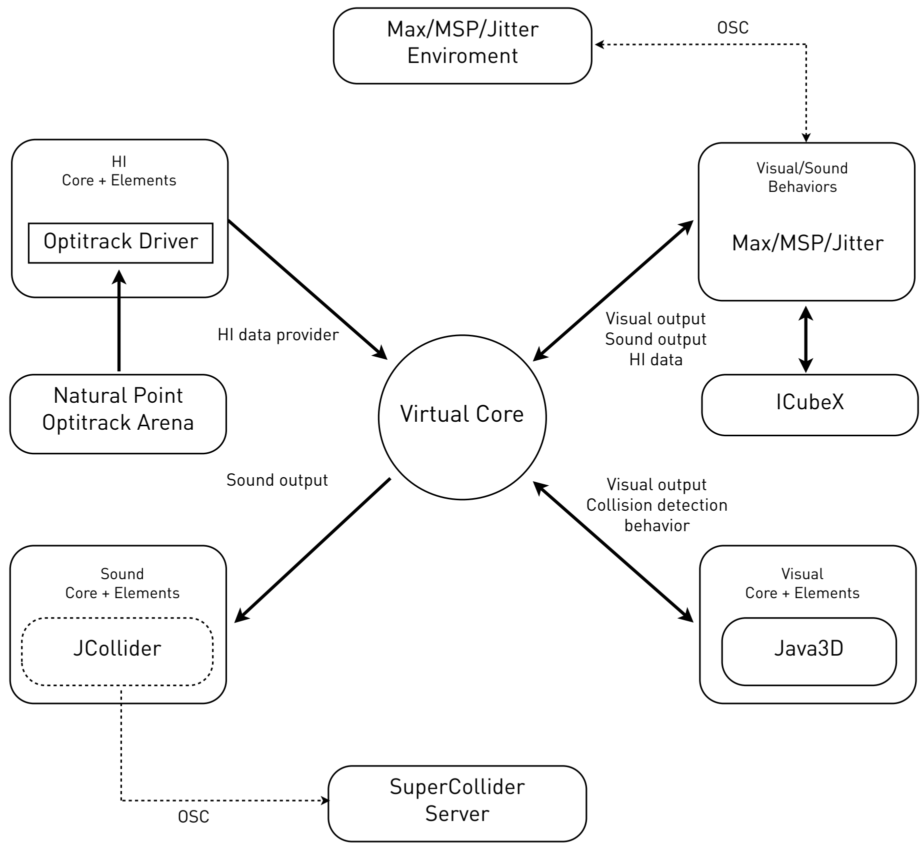


Fig. 7 An example of a JOINDER's deployment. The modality (visual/Java3D, sound/Supercollider) and human interface (HI) cores as well as other technologies (Max/MSP) are interconnected by the central core (virtual).

environment Max/MSP (for scene visuals and sound), Java3D framework (for monitoring visuals) and SupperCollider3 synthesis server (for scene sound), all running in independent machines within a Ethernet network. A visual and audio core were developed using Java3D framework and JCollider library respectively. The sensor interfaces' input is obtained via the Human Interface Core (generic input device manager and virtual element targeting enabler) for ICubeX sensors and Natural Point's Optitrack system (via a custom NatNet to OSC driver).

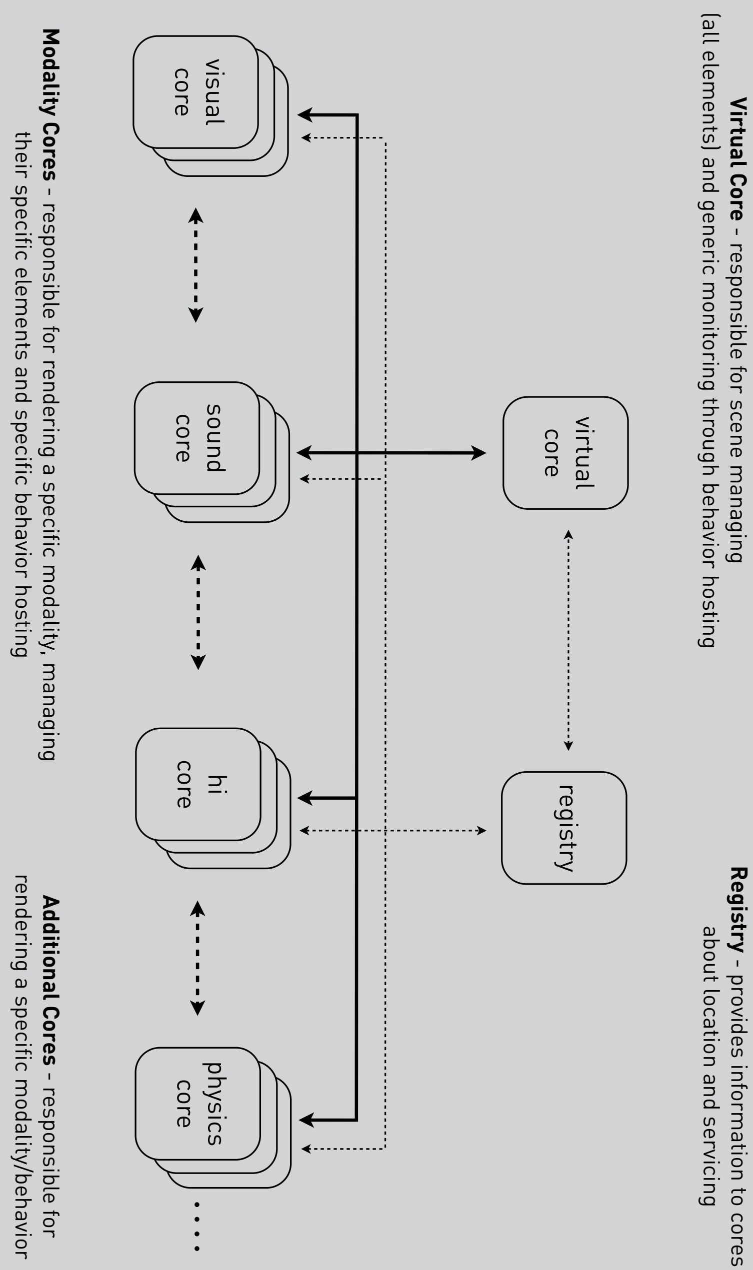


Fig. 8 An overview of the JOINDER's network. Modality and additional cores are interconnected with the virtual core through the registry server.

The advantages that JOINDER offers can be summarized into three main categories.

1.8.1 Wrapping as the central integration strategy

One of the main objectives of JOINDER is to maximize past efforts and specialized tools through integrating them into the framework's work flow (see Fig.8). The objective of JOINDER is not to provide an all round solution to all the areas involved in data sonification and artistic sound production.

Besides the obvious "re-invent the wheel" waste of resources, it would have been unrealistic to try to provide an alternative to specialized tools and their state of expertise and testing. Instead, by developing a system that is capable of interconnecting these tools, it is the purpose of this work to allow developers to take full advantage of the expert level of their functionalities as well as re-use prior developments made in their environments of choice.

For example, all of the previously described toolkits and frameworks using the Max/MSP, SuperCollider3 and Pure Data environments constitute an extremely valuable code base that can be incorporated and used through wrapping within JOINDER. For this purpose, four levels of binding were specified: Strong (for Java libraries), semi-strong (for C/C++ libraries through JNI), semi-loose (element-based network connection) and loose (behavior-based network connection). These levels express the integration degree of a given external technology and standardize the coupling procedure of the latter into the system. As such, the integration levels ranges from full library's embedding with a dedicated management service (provided by the respective core) to stateless, message based communication between JOINDER and the external environment (ex. OSC).

Fig. 9 Users interacting with virtual objects in the SoundField project. Through an Optitrack motion capture system, the users manipulate the virtual objects' parameters (line and plane), by defining the position and orientation of the object's vertices in space. these objects can then be activated through collision detection with a third object.

1.8.2 Space as central mapping paradigm

This strategy aims to represent the multimodal spectrum into one mapping domain (e.g. space). By rendering visual and audio within a shared space of interaction with the user (see Fig. 9), their combined use can occur in a more straightforward fashion. In this way, the user's multimodal interaction possibilities are integrated into the system's design as a fully integrated structural paradigm. This design is derived from both a perception driven musicological framework based on electroacoustic analysis and production as well as immersive environments' interaction possibilities. Given the variety and abundance of specialized technologies in the latter field, these premisses can only be met by establishing a standardized operational and technological bridge between them and by facilitating their interconnection with information technology standards. For this purpose, a simple scene graph is implemented in the central node (e.g. Virtual Core). It provides a synchronization hub for the modality core's instances of a given deployment and a scene element's data center for global behavior tracking and task execution.

1.8.3 Java as the central technology

Oracle's Java SDK 6 was chosen as the main base technology. According to the classification presented in [163], it provides the industry standard mature solutions regarding integrability (Object oriented paradigm; JNI wrapping), flexibility (virtual machine based; real-time specification for Java), extensibility (design pattern prone; various open source libraries), accessibility (standardized language specification; web oriented technology), portability (cross-platform; Web Start), availability (free; open source) and durability (strong rooting in the IT development practice). Furthermore, there are several additional points worth underlining. First of all, Java's server-side infrastructure is very well developed, maintained, documented and tested, in great part by its application in a multifaceted commercial realm. This makes it an excellent bet concerning future web oriented development.

Additionally, the list of Java Virtual Machine's compatible languages (Python, JavaScript, Lua, Ruby, PHP, Scala, Lisp, etc.) further extends the integration possibilities of already existing implementations within the same base environment. Finally, the performance of the chosen platform is of high importance given the interactivity and responsiveness requirements inherent to this project. The standard inclusion of HotSpot/JIT (Just In Time) compilation in Java virtual machines - which replaces interpretation by compiled versions regarding commonly, repetitively executed code blocks - contributes for average performance achievement comparable to commonly used compiled programming languages (ex. C/C++). This is difficult to achieve in other interpreted languages such as Python given its dynamic type nature (in opposition to the statically typed Java) [145].

Areas	Components	Description
Synchronization	Registry - Broadcasting	Cores registry servers Network information management
Representation	Cores - Elements	Technology wrapping classes and processes Specific implementation management
Scheduling	Behaviors - Updaters	Implementation of synchronous and asynchronous command generation and processing Thread for triggering behavioral processing Pool management
Communication	Connectors	Abstraction for intra and inter core communication through multiple protocols support
Execution	Commands	Processing unit executed by the cores Instruction set container Threaded, serializable and log-enabled classes Macro definition, state replication and recollection
Creation	Factories	Reflection based instantiation Runtime component replacement Implementation cache

Fig. 10 A summary of JOINDER's modules and their description.

The table depicted in Fig. 10 summarizes JOINDER's functional areas and their respective components. An in-depth discussion of these components is presented in Section 2.2.4 and a schematic representation of their implementation is available in the Appendix. Up to the current point of development, the JOINDER framework has met the challenges outlined by this research's requirements. In terms of design, the multi-threaded and distributed core environment was able to deliver a solid platform where multiple devices fed data at a high rate (e.g. 100 Hz sampling rate of the Natural Point's Optitrack system) into a concurrent environment, composed by multiple cores. Additionally, JOINDER's modular approach allows independent and switchable assignment of input devices to virtual object's properties. Although the immersion factor is not always possible due to technology access restrictions, one should expect that today's state of the art devices will

Cores/Elements/Components	Technology	
JOINDER	Java SDK 6	
	Kryonet (TCP/UDP simple server library)	
	JavaOSC (OpenSoundControl protocol library)	
	Vecmath (3D vector manipulation library)	
Modality specific implementation	Visual	Java 3D/JOGL
		Max/MSP/Jitter
	Audio	JCollider (SCLang for Java) + SCServer
		JASS
		JavaSound
		Processing/Minim
		Max/MSP
	HI	Processing (WebCamTracker)
		WiiRemoteJ/BlueCove (Wii)
		Max/MSP (ICubeX)
		OSCPack (NatNet to OSC Custom driver)

Fig. 11 Table summarizing the technologies used in JOINDER throughout this thesis.

be tomorrow's standard input paradigms, as tracking technology gets increasingly available (Ex. MS Kinect). As such, the system's architecture takes into account future integration demands in order to reduce later adaptability efforts. Concerning output, the separation of the rendering tasks through independent cores enabled a distribution of the modalities' processing load. For instance, the implementation of two visual cores allowed for differently intended visual feedbacks to be simultaneously rendered, one for monitoring purposes and another for providing aesthetically rich visualizations to the user.

The table presented in Fig. 11 summarizes the technologies used in the JOINDER core implementation and throughout the iterative use case development process. This result originates not only from using a scalable, mature technology (Java) but also from the "on location" refactoring process afforded by use case scenarios. The use cases provided not only the evolutionary scenarios to improve various design and implementation initial decisions but also a phased performance assessment that validated the outlined methodology. The user-centered stages will be focused in the next section.

1.8.4 User centered development

Throughout its development, the design and implementation of JOINDER followed an iterative approach. Given the diversity of the areas from which it originates, a prototyping strategy provided a phased progression for testing the implementation choices devised for the interaction metaphors. It also supplied a way for monitoring the impact on performance as deployments grew wider (in terms of number of technologies involved) and more complex (through the increase of monitoring and action threads). A user centered methodology [110] allowed to collect valuable user feedback at early stages of development regarding the virtual object mediation viability, the realtime performance compliance of the system and the dynamic configuration of the metaphors' implementation. For example, the rendering cores were made independent and completely detached from the central virtual core. As such, the rendering computational weight was spread over multiple machines, in a fully distributed system.

However, given the research laboratory's confinement of the interactive sonification use cases, there was the need for a broader, more ecological approach that would allow the development of features through a participatory design strategy [39]. This strategy was fully deployed in the SoundField project. In contrast with the predefined setup and fixed content of the interactive sonification case studies, the SoundField project used the users' demands as a catalyst for the software's innovation and interaction modes' assessment.

Over a period of four months, the user-centered stages followed the same structure in every use case. After the presentation of selected demonstrators, a pre-interaction meeting would discuss the preliminary evaluation and requirements assessment in relation to the workshop's goals. This would be followed by an implementation of the gathered demands and the respective interaction session. Afterwards, a post-interaction evaluation collected feedback information concerning this session through questionnaires, group interview and debates. In total, 14 distinct deployments and additional variations were developed and used by a total of 77 active participants.

This series of iterations aimed at gathering new usages of the mediation paradigms described earlier and assessing the user's evaluation regarding the implementation of such new approaches. Besides the development of the framework's code base, the main target was to test how well the framework's structure would respond to rapid prototype development challenges in terms of modularity (i.e. addition of device drivers components) and scalability (e.g. the increase in the number of concurrent processes and autonomous machines involved in the deployment). As a result, a wide coverage of user driven requirements was obtained, guidelines for the system's optimization were extracted and steering directives towards future developments were formulated (ex. web based access and multiple virtual scenes' synchronization).

A in-depth discussion of the user centered process is presented in Chapter 2 and will be available in [40].

1.9 Introduction to the Development chapter

In this section, a summary and the main results of the following article based sections is presented. The first section, entitled "An embodied music cognition approach to multilevel interactive sonification", is published in an interactive sonification special issue of the Journal on Multimodal User interfaces [81]. This article has been reformatted in order to fit this thesis' layout. The second section, entitled "JOINDER, a software framework for networked multimodal interface development" is submitted. The third section, entitled "Multilevel immersive interfaces for electroacoustic music composition and performance" is to be submitted.

An embodied music cognition approach to multilevel interactive sonification

JOINDER, a software framework for networked multimodal interface development

Multilevel immersive interfaces for electroacoustic music composition and performance

An embodied music cognition approach to multilevel interactive sonification

1.9.1 An embodied music cognition approach to multilevel interactive sonification

Summary

- In this paper, a new conceptual framework and related implementation for interactive sonification is introduced.
- The conceptual framework consists of a combination of three components, namely, gestalt-based electroacoustic composition techniques (sound), user and body-centered spatial exploration (body), and corporeal mediation technology (tools), which are brought together within an existing paradigm of embodied music cognition.
- The implementation of the conceptual framework is based on an iterative process that involves the development of several use cases.

Results

- The use cases provided a preliminary validation concerning the interaction metaphors, the applicability of immersive interfaces for interactive sonification and the prototype system's performance.
- Through this methodology, it is possible to investigate new approaches for structuring and to interactively explore multivariable data through sound and further develop the related tools for future tasks' compliance.

JOINDER, a software framework for networked multimodal interface development

1.9.2 JOINDER, a software framework for networked multimodal interface development

Summary

- This article introduces JOINDER (Java Omni Integrator for NetworkeD Embodied inteRaction), a software framework for multimodal interface prototyping in Java.
- JOINDER envisions the implementation of a bridge for interconnecting different technologies together within a common, virtual reality based representation.
- This development has been motivated by research in two areas, namely, multilevel sonification of multivariable data, and more recently, the exploration of interactivity within the sound art installations.

Results

- A novel software framework for distributed multimodal interfaces creation was developed through the combination of electroacoustic theory and embodied music cognition models.
- The effectiveness of the applied user-centered methodology for technology development and interaction modes implementation was ascertain within auditory display and interactive art.
- The resulting software solution can provide a scalable, portable and versatile solution for multimodal interfacing, heterogeneous technologies' integration and real-time performance.

Multilevel immersive interfaces for electroacoustic music composition and performance

1.9.3 Multilevel immersive interfaces for electroacoustic music composition and performance

Summary

- This article presents a conceptual framework and related technological environment for electroacoustic music performance and composition. The approach is comprised of multilevel mapping, scope variance and spatial interaction within an embodied music cognition's approach.
- Originally aiming to establish an interface development methodology for the interactive sonification domain, this paper presents a roadmap for the design of user centered interfaces within the electroacoustic music field based on an experimental validation within interactive art's domain.

Results

- A user centered methodology was successfully applied for the interaction strategies' implementation and validation.
- A validation of the conceived interaction metaphors and implemented tools for both interactive sonification domain and electroacoustic practice is presented through Smalley's Spectromorphology and Space-Form classification.
- A framework for the development of embodied, space based, immersive interfaces for music performance and composition is formulated.

chapter 2 /

development

An embodied music cognition approach to multilevel interactive sonification

An embodied music cognition approach to multilevel interactive sonification

Authors

Nuno Diniz - Pieter Coussement - Alexander Deweppe

Michiel Demey - Marc Leman

Abstract

In this paper, a new conceptual framework and related implementation for interactive sonification is introduced. The conceptual framework consists of a combination of three components, namely, gestalt-based electroacoustic composition techniques (sound), user and body-centered spatial exploration (body), and corporeal mediation technology (tools), which are brought together within an existing paradigm of embodied music cognition. The implementation of the conceptual framework is based on an iterative process that involves the development of several use cases. Through this methodology, it is possible to investigate new approaches for structuring and to interactively explore multivariable data through sound.

Keywords

Interaction - Sonification - Embodiment - Electroacoustic - Framework

2.1.1 Introduction

Interactive Sonification has been proposed for displaying multivariable information through non-speech sound-based communication. However, the search for efficient methods to address this remains an open issue [56]. For illustrative purposes, we consider a scenario in which a user has to analyze the simultaneous variation of 10 stocks in real-time and needs to make decisions on the low level (i.e. the individual stock as in buying or selling) and on a higher level (i.e. the combined behavior of sets of stocks as a reflection of the company's assets). When sound is used as a way to investigate each individual stock, then it is clear that the simultaneous sounding of ten sound streams may be rather overwhelming if no internal structure and emerging properties of sound fusion are taken into account. Overwhelming sounds will likely prevent the user from efficiently parsing, analyzing and reacting in real time to the information that is conveyed to him. However, when electroacoustic composition techniques that deal with combined sounds are incorporated, there is a possibility that the data display becomes manageable and more accessible. The main issues can thus be stated as follows:

- How to make multiple levels of sonification both perceivable and meaningful, in such a way so that their interrelated nature can be used to their best advantage?
- How to make sonified information available, in such a way that it can be manipulated and apprehended in a straightforward manner?

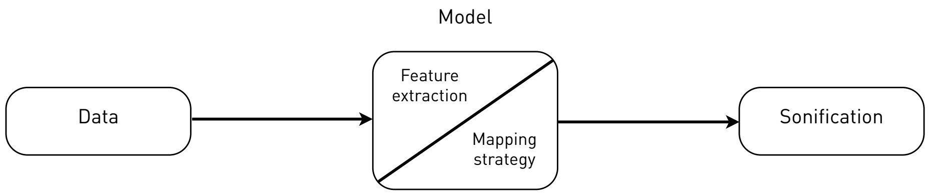


Fig. 12 One-way sonification: the model performs feature extraction on the data, and uses a fixed mapping strategy for sonification.

The traditional approach to sonification is based on a one-way flow of information from system to user. As shown in Fig. 12, this approach consists of a model that first performs feature extraction on the data, and then applies a mapping strategy for sonification. Prototype examples of this approach are the Auditory Icons [60] and Earcons [11].The one-way approach generally entails a high degree of abstraction regarding the static data mining processes intervening between raw data and icons. Other techniques are often closer to the data, such as parameter mapping (e.g. Audification) [94]. However, these approaches seldom allow the user to address the data from multiple perspectives, such as sound streams grouping and global versus local behavior analysis.

Given the limitations of the one-way approach, the question can be raised whether a two-way approach can be developed in which the user would be able to explore the data interactively. This means that the user intervenes between model and sonification. Hence, this requires a more flexible approach to data encoding and exploration techniques in order to cope with emergent possibilities of multivariate data. In fact, the techniques must possess a dynamic nature so that they can mutually influence and redefine each other. This approach to sonification implies a shared responsibility between preconfigured relations on one hand, and the user's expertise and intuition on the other hand.

In this paper, we propose a methodology for the sonification of multivariate data based on the idea of interactive sonification [73]. This involves the display of information on the basis of interaction and scope. Interaction implies that the

perceptual viewpoint of the user can be controlled through his movement, while scope implies the possibility of zooming in and out for a discrimination or fusion of sonified levels. As we will discuss in the following section, these concepts are closely linked with electroacoustic gestalt-based composition and theories of musical embodiment.

2.1.2 Theoretical background

In the two-way approach to sonification (Fig. 13), the user and the sonification system may affect each other on the basis of interaction. To handle the complexity of such an interaction, it is necessary to further decompose the notion of interactivity into different elements. In what follows, we present a perspective on interactive sonification based on three components namely, (1) gestalt-based electroacoustic sound generation (sound), (2) body-centered spatial exploration (body), and (3) corporeal mediation technology (tools). Together they form the pillars of an integrated approach to sonification, which is based on the embodied music cognition paradigm as described in [99].

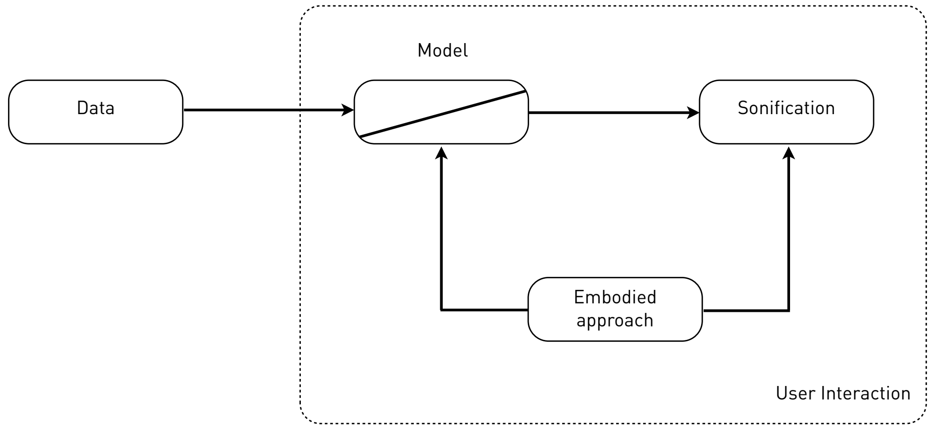


Fig. 13 Two-way sonification, or interactive sonification: the model performs feature extraction on the data, but uses a modifiable mapping strategy for sonification based on embodied cognition.

2.1.2.1 Gestalt-based electroacoustic sound generation

Sound-generation is here conceived from the viewpoint of gestalt theory. As known, gestalt theory is strongly determined by principles that affect the meaning of sound streams in such a way that the whole provides a level of information that is different from the sum of its parts. In our view, we aim at controlling sonification by exploiting principles of gestalt theory as guidelines for multilevel meaning generation. Such an approach considers both analysis and synthesis as well as segregation and integration. Analysis and synthesis account for the decomposition of sounds into frequencies and the subsequent integration into pitches and chords [98]. Segregation and integration account for the grouping of time-varying patterns, depending on intervals of time and pitch [17]. In our view, these concepts need to be implemented using a language that incorporates these elements in a culturally aware context. Our focus then turns to the electroacoustic composition domain, and its internal and external contextualization mechanisms [134]. Given its wide flexibility regarding syntactic representation [49], electroacoustic composition theory and practice may help to make a given dataset more accessible and easier to mold according to the user's inspection goals.

The search for successful scope transposition techniques in sound based communication has always been a central concern in this art form. Several technical processes that are concerned with the relationship between singularity and regularity of events [37] are addressed in Sound Object Theory [27], Concept of Unity [28], Moment Form or Formula based composition [141], as discussed in [45]. Consequently, they might encapsulate guidelines that can be of service in functional sound based communication [72]. For example, the dialogue condition that is imposed to the sound object and the enclosing structure holds a dynamic perspective shift, which reassures the relationship between the two concepts. As a result, this unifying concept connecting sound object to enclosing structure, is taken as a design directive, as in [129], for the manipulation of multiple levels of complexity.

Furthermore, electroacoustic practices also address structure in time. As an example, Stockhausen's take on Information Theory [70] focuses on the behavior of sound objects through time, which is in close affinity with the principles of similarity, opposition and belongingness in Gestalt theory [25]. The sequence in which the auditory stimuli are presented is in Stockhausen's view crucial for perceiving the musical discourse. Among others, Mikrophonie I is one of the works where the main structural strategy is based on the amount of identity variation of a given gestalt segment or "moment" in relation to the previous ones [140]. Such considerations are an imperative requirement, given the human auditory system's idiosyncrasies (i.e. the precedence of relative over absolute relations in parameter discrimination).

These concepts from electroacoustic practice present themselves conform to the flexibility requirements related to the two-way sonification. Furthermore, in our approach, it is assumed that the explorer of the data controls these changes through body movement. As an extension of the referenced guidelines of Schaeffer, the concepts of gesture and texture in spectromorphology theory [134] are included as a methodological base concerning sonic attribution, deployment and articulation. In his classification, Smalley addresses the consequences for the musical discourse emerging from the loss of physical character in sound objects. Such a loss undermines the bases upon which the internal relations of sonic elements are perceived and reasoned upon. As such, gesture can be viewed as a wrapping mechanism for scope transposition. It provides a translation device between the physical user and the sonic texture (or spectral identity). Furthermore, it allows the perceptual grouping of individual sound objects. As a result, multivariate data access should include the representation of the variables' progression as an energy-motion trajectory of gesture.

To summarize, the aim is to transpose the compositional strategies to the interactive sonification domain and to apply the relationships between material and structure to the micro and macro sound levels of data presentation. As a result, functional contexts are generated by data-dependent hierarchical levels that still preserve their informational identity and significance.

2.1.2.2 Body-centered spatial exploration

Interaction with sound assumes a user dealing with information resources from the viewpoint of his/her proper actions [62]. Accordingly, we assume that the user's conception of sonification proceeds from the viewpoint of an attribution of action-relevant values to sonification, in such a way that the information resources may lead to meaningful experiences. Given this action-based viewpoint, it is important to keep in mind that natural ways of processing multi-level information in our environment is often based on the physical distance between the user and the resources. This physical distance defines the scope in which one is dealing with these resources. Therefore, we can say that the action-value, attributed to a resource, is dependent on the distance grounded in actions that imply zooming in or out, and on experience.

For the purpose of sonification, we consider different spaces that surround the user as subjective body-centered spaces. These define the scope for a sonic exploration of the information resources [101]. Three levels may be considered here. A first level is the pericutaneous space, directly surrounding the body, which is defined as the space that provides fine-grained control over interface technologies. A second level is the peri-personal space, which is defined as the area that immediately surrounds a person and that is reachable by the limbs of the subject. A third level is the extra-personal space, which is defined as the space out of the reach of the subject in reference to the absolute position within an environment. This multilevel spatial awareness is attuned with the

hierarchical nature of sound object exploration in such a way that the spectromorphological gesture contained in the sonification can be transposed from audi- tory parameters to an exploratory process. By placing the interaction within a context of spatio-temporal representation, it becomes possible to engage in an embodied dialog with the data. The latter is then converted to a human scale, in reach of the user.

In summary, by enabling a configurable location and form representation of the data in space, this methodology invites the user to physically approach the inspection process through a shared space of multilevel interaction [88]. An embodied cognition approach is thus expected to further enable a perceptual link between the data and the semantic high-level representations of the user.

2.1.2.3 The corporeal mediation technology

The involvement of the human body in sonification assumes a technology that mediates between body and data. Corporeal mediation technologies are conceived as tools that extend the human body as a kind of prosthesis in order to allow the human explorer to mentally access the digital realm in which the data resources are assessed. These tools can provide guidelines for deploying an interface that presents an integration of the variable resolution of the human motor system and sound objects [63]. Our approach is based on expanding the mediating role of the body through interaction with an immersive 3D environment, using virtual entities [108]. The present framework's ongoing implementation [45] is a consequence of the need for generic data sonification tools for both research and applications [94]. It aims at supplying a software infrastructure for integrating the base concepts discussed above while addressing issues concerning portability, flexibility and integrability.

A Java framework has been developed, based on a functional division of the multimodal interaction realm into individual branches around a state representation. In this framework, the concrete implementation of a virtual world, its visual and auditory representations, and the human interfaces, can be defined according to the desired performance, access or functional needs of the intended use cases, using external libraries and platforms (e.g. Java3D, Supercollider 3, Max/MSP, Ableton Live). An overview of the technological setup is shown in Fig. 14.

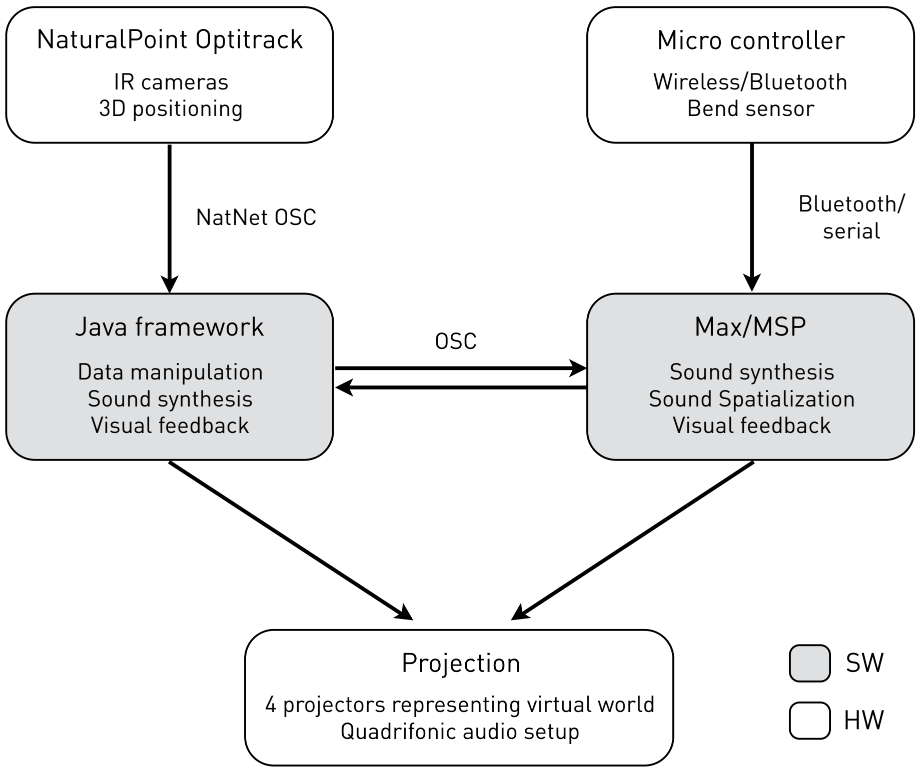


Fig. 14 An overview of the Technological implementation

Since the framework is intended to provide various solutions depending on certain research and/or application needs, the following technical description is restricted to the main use case that is discussed in the present paper namely, the dance use case (see Section 2.1.3.2):

- **Optitrack motion capture system** - This system is used to capture the 3D position and orientation of IR reflective marker sets, that are typically attached to the human body. The data is transmitted through the NatNet protocol via the Arena software and converted to OSC by a custom driver.
- **Micro-controller** - An ICubeX system is used for connecting a bend sensor in order to control the radius of the virtual inspection tool. The data is received in a Max/MSP instance through Bluetooth and sent in the OSC protocol to the Java framework instance.
- **Java Framework** - The Java-based framework constitutes the core of the system concerning virtual scene state representation and monitoring and behavior triggering. All the above OSC formatted information is gathered in the virtual core and used to place and configure the virtual objects in the scene. Based on this information, collision detection and additional information (for example, the distance between virtual inspection tool and activated sound object for sound amplitude and reverb parameterization) is calculated and sent in the OSC format to Max/MSP. Also, through the Java 3D implementation, a visual core is responsible for the rendering of the visual projections and the file logging of all relevant data that can be used for offline analysis (e.g. position of virtual objects; velocity associated with virtual objects at creation time; distance between virtual objects; time-stamped collision detection and activation of sonification elements, etc.).
- **Max/MSP** - The Max/MSP environment is responsible for the sound synthesis and spatialization (together with Ableton Live through MaxforLive) based on the data transmitted by the Java framework.

The above mentioned portability and flexibility requirements are met by means of the use of Java technology in the framework's core implementation for state management and monitoring. Respectively, portability through multiple platform support and flexibility in terms of new functionality incorporation as well as rapid prototyping development through object-oriented component hierarchization are more easily achieved. In the dance use case, the use of an Optitrack system is due to the specific need for capturing 3D position in a large interaction space. It is completely detached from the framework's scene representation, which processes 3D positions, independently of the device through which they are collected. As an example, for a given "laptop centered" use case, a set of three optical cameras might be sufficient for 3D position tracking and this adaptation would not have any repercussions on the core's state management and monitoring implementation.

To sum up, sonification is conceived from the viewpoint of an interaction, which is based on principles that couple perception and action through gestalt-based awareness perception and body-centered spatialization. Mediation technology is needed for the coupling of these elements, in order to make sure that body movements can naturally be extended and deployed in the digital environment. As such, interactive sonification based on scope variation can be formulated as a disambiguation asset towards the variables' behavior and its sonic correspondence. This is achieved through an active molding of the variables' output and the added value concerning the result of this interaction.

2.1.3 Use cases and implementation

In this section, we present two use cases that have been set up in the context of implementing the approach above described. In our approach, the use cases form a core element within an iterative development cycle, in which the interactive system is tested and refined through cycles of user validation. For that aim, interface affordance, hierarchical sound levels generation and the related mapping needs to be developed and tested in an integrated way. By applying this integration in combination with a user-based, phased validation strategy, we obtain an early detection of preliminary issues. On this basis, it is possible to steer the development of a system combining the three main components of our embodied approach to sonification, namely sound, body, and tool. As pointed out in [165] concerning electroacoustic music analysis in music information retrieval research, there is a need for user-oriented approaches. In what follows, two use cases and their evaluations are described.

2.1.3.1 Prototype use case

In the prototype use case, we tested three hierarchical layers of sonic data [45], using the musical concepts of pitch, interval and chord. The pitch mapped data according to a model of variables' value. The interval represented a numerical relation between two variables values. The chord represented the occurrence of a given set of relations among the data. Pitches, intervals and chords were accessed through the use of a virtual inspection window and a virtual inspection tool that allowed an interactive inspection. Sound generation was based on Schaeffer's concept of sound objects and on the structural relationship between these objects. The virtual inspection window consists of a finite set of virtual objects that represent a time frame of the variable's values. When activated through body movement and collision detection, the current value

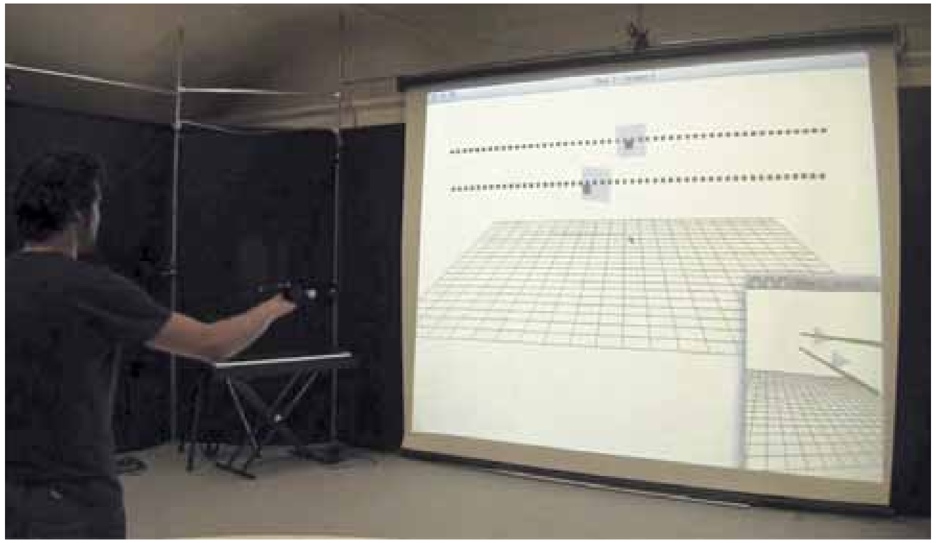


Fig. 15 A user exploring 2 variables' data. The two object arrays represent each the virtual inspection window of each variable. The triangular shaped objects represent two virtual inspection tools (virtual microphones) operated by the user.

attributed to a given virtual object is fed to the sonification engine. In this use case, the virtual inspection window is represented by an array of several cubes. The sonification process is assigned to the latter objects, and conveys information about the activated set as a whole, stimulating a perceptual interpolation between the set and the individual nodes. The virtual inspection tool functions as a virtual microphone that enables the activation of the inspection window.

This approach was inspired by Stockhausen's microphone use in Mikrophonie I, as well as by Schaeffer's view on the recording process in Concrete Music. It provides a perceptual play with the distance between object and microphone in terms of a sonic realization that was related to reverberation and amplitude modulation. The virtual microphone that allows sonic scope variation was visualized as a small pyramid, following the user's hand movements and orientation. An explorative example is illustrated in Fig. 15.

2.1.3.2 The dance use case

The dance use case was based on the previously described prototype, taking into account the users' observations and comments during its evaluation sessions [45]. This time, the array is scattered in space, enabling users to vary their inspection in space and time even more. An extra bend sensor and wireless ADC allowed a basic signaling of an open or closed posture.

This made it possible for the participants to vary the scope of the virtual microphone by changing their posture from an open (global) viewpoint to a closed (detailed) one. Varying the scope of the virtual microphone made it possible to inspect the individual grains of the sound when the user resides in his/her pericutaneous space. A decrease of the index of contraction into the peri-personal space provides access to the macrostructure of the sound-scape. The contraction index, defined in [23], is a time variable value that is defined by the ratio of the area between the bounding box of the human body posture and its silhouette. The sound diffusion of the microstructures was directly linked to the user's scope variation approach (the inspection vector), using sonification ideas based on Smalley's concept of spatiomorphology. That means that on a macro level, this scope-driven sound diffusion defines the spectromorphology of the soundscape. The interactive sonification implies the use of space in relation to a body-centered spatial exploration over time. Since composition is imminent and ongoing, it can only be molded through this kind of interaction.

The dance use case consists of two parts, namely, a setup part and an exploration part. In the setup part, the data objects are set out in space. In this case, the data objects represent dance movements. A sequence of these data objects thus represents a trace of the dance movements. It is worth noticing that the sequence of data objects is not to be regarded as a dance choreography in the

Fig. 16 First phase (Left) where a user sets a trail by performing dance movements. Here the blue objects represent the trail defined by the movement of the head of the user. Second phase (Right) where another user explores the trail. The objects in red represent the objects in collision with the virtual microphone represented by the grey sphere. The radius of the sphere can be adjusted by the opening/closing of the elbow of the user.

strict sense, but as an occupation of space. In the exploration part, the user is given specific tasks in order to explore the trace of the initial movement (Fig. 16). The users had no visual representation of when the virtual objects creation took place within the choreography. The user's task was formulated as follows: "Your predecessor laid out a small choreography, can you recreate this choreography?". The user was informed about the interactive sonic scope. Each user was explained that by reaching out the arms (small contraction index), the sonic scope would increase. Inversely, by closing the arms (large contraction index), the sonic scope would decrease. Furthermore, the user was informed that a low pitched sound signaled points in the beginning of the sequence, while high pitched sounds signaled points towards the ending of the sequence. The pitch would gradually increase from begin to end. When participants were unable to recreate the choreography on their own, the task was divided into smaller tasks, namely, to locate the start position, end position, and global direction of the choreography. Additionally, when activated, the spatially located path elements allowed access to the velocity related information. Besides indicating the direction of the path through an increase in pitch, variation in the velocity of the original movement was conveyed by the variation in the pitch's increase.

For example, if the velocity of the original movement is 3/2 of the maximum velocity (based on physical constrains) at the instant of creation, the pitch would increase a perfect fifth from the previous one. This strategy is inspired on Stockhausen's take on Information Theory, as discussed in Section 2.1.2.

2.1.3.3 User evaluations and feedback

In line with our user-oriented development approach, feedback of the users' exploration was recorded and analyzed in order to better understand the possibilities of interactive sound scope variation. The necessity of this level of user-involvement also stems from the implied wielding strategies [41]. A range of methods, mainly adopted from the field of HCI-usability studies [155] [90], is applied for this purpose [100]. Questionnaires, a focus group discussion and the video footage of the interaction sessions were collected. In addition, all movement data of the users tracked by the motion capture system, and all events occurring with the virtual objects were logged to files for further offline analysis.In the prototype use case, users were able to perceive and to discern the information in multiple levels of sonification [45]. Users found the system to be responsive and its operation to be intuitive. Their performance of appointed tasks was improved by the use of different levels of sonification (i.e. pitch, interval and chord). The outcome of the evaluations made in the prototype stage (concerning performance, maneuverability, precision, distinguishability and completeness of visual and sonification output) showed that some of the most important issues raised, addressed the interaction with the virtual objects.

The need was expressed to dynamically change the morphology of the virtual inspection tool. Other reported problems pertained to the use of a visual aid for a better perception of the user's own movements. All these issues were dealt with in the implementation of the dance use case. The dance use case was difficult for most users. Observations of the recorded footage show that the users experienced difficulties in reconstructing the trajectory, although they were able to rapidly locate the beginning, the end and the direction of the trajectory through the sonification. A particular problem did arise when the movement of the original dancer occurred at a previous spatial location. For users exploring the trajectory, it was hard to disentangle the two events occupying the same place. In future implementations, such difficulty might be avoided by increasing the resolution in the virtual inspection tool and/or by suppressing the generation of virtual path objects when velocity is below a given threshold. The latter would link the creation of virtual objects to the user's energy level. A visualization of an exploration is shown in Fig. 17. In this figure, it is clear that the second user (exploring the choreography of the original dancer) can accurately locate the beginning, turning and end points and correctly reconstruct the direction of the defined path.

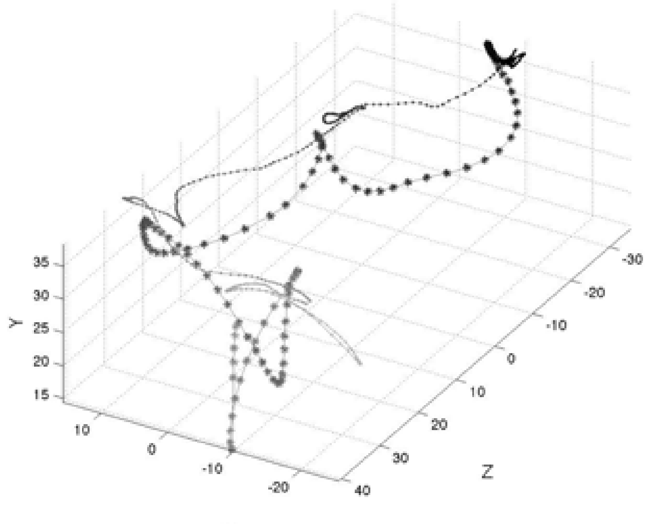


Fig. 17 Visualization of movement data recorded by the motion capture system of two users. The trail set out by the first user is visualized by the starred markers where the starting point is in black and the ending point is in light grey. The trajectory indicated with the dotted markers represent, the movement of a second user exploring and recreating the first user's movement.

Based on Nielsen's guidelines for early stage usability evaluations [109], preliminary tests were conducted with five users. According to Nielsen, this number is advisable to keep evaluations cost-effective. The number is sufficient to extract 85 percent of the usability problems reported in early stage prototype evaluation. Alongside the experiment, the participants were asked to fill out a questionnaire. This generic questionnaire encompassed standard user background information as well as a general appreciation concerning the used interface paradigms and technologies. The questionnaire was divided in two parts. A first group of questions dealt with the evaluation of sonic, visual content and the relation between the two. Additionally, this part questioned social interaction (if applicable) and personal assessment of the tool operation. The majority of these questions used a six point Likert evaluation scale. By using an even number Likert-scale, which eliminates the middle option, the users were forced to make a non neutral rating of their level of satisfaction. The second part of the questionnaire contained questions concerning demographic information, (art) education, cultural profile and new technology use.

Concerning the users' background for the described use case, the users' gender was predominately female (n = 4) and had an even age distribution from 18 to 60 years old. All of the users received higher education and the majority (n = 4) had a background in artistic education for a duration of over five years. The users reported to regularly engage in cultural events (n = 4 for more than monthly activity). Most of the respondents, however, stated to have little or no experience with new media demonstrations or exhibitions (n = 4). Regarding experience and leisure-time involvement with digital communication devices (e.g. computer, smartphone, . . .), most respondents reported manifold and daily usage of the described technologies (n = 2 to be- tween 1 and 3 hours and n = 2 to over 3 hours per day). The users' opinions on the interface were divided. The identification of the sonic output and the precision of sound control were found to be difficult. However, the applicability of sound to actions was classified as rather suitable (n = 3) and the majority positively classified the aesthetic quality of the sonic output (n = 4). Regarding inter-modal relations, the majority classified the relation between visuals and sounds as clear (n = 3), attributing equal importance to both modalities (n = 4). In an overall personal evaluation regarding tool manipulation, all users evaluated the functional usefulness between good and very good and stated that their attention was divided between the use of new technology and the combination of visual and sound.

In a post-interaction evaluation, participants were asked about their experiences by rating different features of the sonification by means of a questionnaire. Concerning the sounds that were used, the most noteworthy findings were that participants reported to be quite able to discern different sounds, although the majority of participants complained about a lack of sonic control. The aesthetic quality of the used sounds and the correlation between sounds and actions were evaluated as good. Participants were able to correlate movements to the visuals more easily than to the sounds. The level of accuracy in their control of the visuals, both functionally and aesthetically, was perceived to be high. Given the combination of sonic and visual feedback in the use-case, most of the participants reported the relation sonic and visual content to be clear and very complementary. Almost all the participants said to have paid equal attention to the sounds although, contrary to the reported poor level of sonic control, a minority of participants said to have paid the most attention to the sonic features of their exploration. The participants' personal appreciation of the aesthetics, the perceived quality of interaction, its functionality and the ergonomics was overall quite positive. In a number of focus groups in between and after the interaction sessions, users were inquired in group about their opinions concerning the operation of the system. Most of the users found that this use case was an apt means to link visual and auditory stimuli together. Most participants agreed that the visuals were primarily used to establish the trajectories whereas the sonification was used to gain a sense of direction (i.e. finding the high and low pitches, the beginning, the end and the course of the trajectory), and this was furthermore reported to be relatively easy. The assessment of single objects was said to be too complex. The confusion stemmed from the fact that most participants were not used to derive 3D virtual positional information from 2D-representations projected on screens. Because of that, some of the participants found that they could better explore the direction of the trail using only the sonification. Although the trajectories were said to be clear enough, isolating discrete pitches through the manipulation of the scope was reported to be problematic. But the flexibility of the scope device was applauded. A suggestion was made to shift to a sonification of discrete pitches when the scope of the inspection tool was at its largest because in the multitude of sonified objects, some resolution was said to be lost. A final positive remark, that was recurrently made during the discussion in all the sessions, was the sense of being completely unhindered by the technology used.

2.1.4 Discussion and future work

This paper exemplifies how an iterative development cycle, rooted in a user centered methodology, can serve as a guide towards the development of an interactive sonification methodology that is fully based on the paradigm of embodied music cognition. All through the development of the interactive sonification tool, the design choices have been based on an integration of the three main components. They support a combination of electroacoustic techniques (sound), embodied space (body), and mediation technologies (tools). By considering the outcome of the user tests and the feedback recorded during the presented development stages, an evolution can be reported over the course of the two use cases. The user evaluation and the incorporation of the feedback provide information about usability issues and their rectification in the subsequent development cycle. Starting with a prototype use case that focused on different levels of investigation of data (using pitch, interval and chord), the dance use case's emphasis was on space occupation and how interactive scope variation allowed spatial access of the data. Though the focus of the use cases varied between iterations, features of the first use case were incorporated in the second use case. This approach led to a more meaningful way of interacting with multiple levels of sound, and ultimately to a better interactive sonification system.

The use cases reflect a work in progress. The feedback and user-tests leave room for improvement and user-requested design decisions made in the previous stages have to be re-evaluated. The concept of scope variance, for example, proved to be too difficult to control, given the operational movements intertwined with natural dance gestures. Given the flexible nature of the two-way sonification, the unpredictability of a user-inspired development process, the arbitrariness common in mapping strategies and reliability issues still confining the technology [111], rapid usability breakthroughs can hardly be expected within a limited number of iterations. Only a limited number of issues can be addressed in one iteration. Nevertheless, as we are reshaping and refining the system, we are hopeful that future developments will occur. For example, in the near future we hope to re-evaluate the problem at hand because we envision a more weighted scope variation through an improved posture sensing. This way, users could listen with their torso to the macrostructure, and reach out with arm (sound chunks) or hand (sound grains), thus giving them control over multiple microphones while all listening to different levels of sonification.

2.1.5 Conclusion

A new conceptual framework has been presented, combining gestalt-based electroacoustic composition techniques (sound), user and body-centered spatial exploration (body), and corporeal mediation technology (tools). The paradigm of embodied music cognition mediated the integration of these components into an implementation of a conceptual framework. By means of an iterative process, involving the development of several use cases, we showed that it is possible to investigate new approaches for structuring and interactively exploring multivariable data through non-speech sound communication and auditory scope variation.

JOINDER, a software framework for networked multimodal interface development

JOINDER, a software framework for networked multimodal interface development

Authors

Nuno Diniz - Alexander Deweppe - Pieter Coussement - Marc Leman

Abstract

This article introduces JOINDER (Java Omni Integrator for NetworkeD Embodied inteRaction), a software framework for networked multimodal interface prototyping and deployment in Java. JOINDER envisions the implementation of a compatibility layer across existing technologies. Therefore, it can be seen as a bridge for interconnecting different technologies together. The development of this framework has been motivated by research in two areas, namely, (1) the generation of, and access to, multilevel sonification of multivariable data, and more recently (2), the exploration of interactivity within the sound art installation domain. Based on these two areas, the software implementation followed a user centered development methodology, addressing the necessary performance requirements while incorporating portability and versatility features that are characteristic for Java-based technologies.

Keywords

Software framework - Interaction - Auditory display - Electroacoustic music Embodiment - Participatory design

2.2.1 Introduction

The motivation for developing a software framework for multimodal interface prototyping is based on the idea that humans interact with their environment in an embodied way, which means that it is action-oriented and strongly dependent on the capabilities of the human body and its use of tools [62]. Digital information can be seen as an aspect of the environment, and therefore, it is natural to assume that the exploration of digital information may profit from interfaces that take into account an embodied viewpoint on human interaction. Such interfaces can be called mediation technologies, as they aim at facilitating the connection between mind, body, and data [99].

The domain of our research is sound driven exploration and monitoring. Therefore, we believe interfaces should be developed in such a way that they become non-obtrusive and have a variable resolution. Ideally, such interfaces extend the human body as a kind of prosthesis and allow the human explorer to access the digital realm in which the data resources are assessed in a most natural way. Furthermore, we believe that interfaces should extend the hierarchical nature of the human motor system and its implied coherent scope variation in performed actions, in such a way that digital information can be accessed on the basis of the human natural bias to consider multiple viewpoints and variable resolutions of the objects in our environment [63]. The concepts of embodiment, non-obtrusiveness, multiple viewpoints and variable resolution, put human behavior, and therefore the user, at the center. Consequently, a software framework for multimodal interface prototyping has been implemented in close connection with users. In particular, we adopted a spiral development cycle based on case studies in which users had the opportunity (given the different constraints of available media technologies) to decide upon the technology and the content [39]. Our task was to provide a software framework that could integrate different hardware devices and software programs together in such a way that users could deal with their content through embodiment, non-obtrusiveness, multiple viewpoints and variable resolution. in other words, our objective is to interconnect digital resources so that the interfaces can be adapted to the natural human way of data exploration.

This article presents the design strategies and implementation stages of a software framework that aims at applying the above premises while respecting the performance, portability and versatility requirements necessary for its successful usage. The next section introduces the conceptual background of JOINDER. This is followed by a state-of-the-art review of multimodal interfacing technologies and the framework's general properties. The next section goes into more specific detail concerning the concrete functional components of the system, which is followed by a description of the development stages regarding the user centered development approach. Finally, a discussion and future work for this project is provided.

2.2.2 Conceptual background

The present software framework for multimodal interface prototyping is the result of research that was initially conducted in the field of interactive sonification and auditory display. As defined by Hermann and Hunt, interactive sonification is "the use of sound within a tightly closed human computer interface where the auditory signal provides information about data under analysis" [73]. A major challenge of interactive sonification concerns the need for a more flexible approach to data encoding and exploration techniques in order to cope with emergent possibilities of multivariate data. Furthermore, these techniques and related implementation must possess a dynamic nature so that they can mutually influence and redefine each other.

In [43] [45], an approach for the sonification of multivariate data based on the concepts of interaction and scope was described. Interaction implies that the perceptual viewpoint of the user can be controlled through body movement, while scope implies the possibility of zooming in and out for a discrimination or fusion of sonified levels. This enables a shared responsibility between preconfigured relations of the data on the one hand, and the user's expertise and intuition on the other hand. Furthermore, it was shown that these concepts could be closely linked with electroacoustic gestalt-based composition and theories of musical embodiment.
The work on interactive sonification generated a number of questions, which provided a background for the development of a software environment. These can be summarized in the following points:

- How to encode both the global and the detailed aspects of a dataset into sound objects?
- How to define relations between the resulting multilevel sound objects and human exploratory actions?
- How to design and implement proper mediation tools that will allow the user to explore a dataset through sound in a interactive and embodied way?

In order to research an answer to these questions, a central concept of JOINDER concerns the use of virtual objects to represent both the users and the data within a 3D environment (see Fig. 18). Following Mulder's 3D virtual sculpting as a metaphor for sound synthesis [108] the users can access the data in a way that mimics his natural and physical interaction with real world objects. This direct manipulation strategy [34] allows for multiple mapping layers, different viewpoints and degrees of resolution [129]. Such embodied approach aims at further enabling links between the explored data, its sonic representation and the semantic high-level representations of the user.

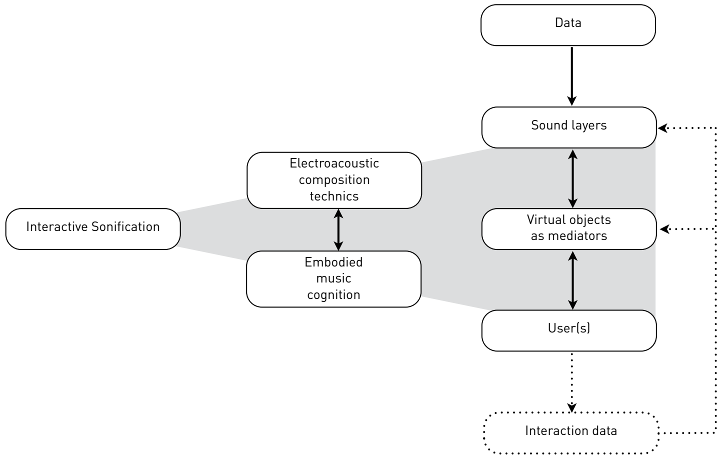


Fig. 18 Relational representation of the intervening domains and concepts of this project.

It should be added that JOINDER can also be used to develop prototypes and applications that don't use the above mentioned virtual object mediation paradigm. In fact, the software facilities offered by the framework may be used for management, representation and synchronization purposes alone between multiple input and output devices. However, much of our work is conducted with the use of immersive 3D environment, as it provide an optimal way the representation of data through real world geometry. For instance, by employing dimensionality reduction (e.g. N variables represented as a 1D time based arrays in a 3D space), the data can be positioned around and between the users. This superimposing of real and virtual world coordinates in the same interaction space allows to take advantage of the natural constrains of the user's body, according to personal space and extra personal space concepts. In summary, in our approach to interactive sonification, the goal is to optimize the mediation technology [99] as creative tool for data exploration, thus maximizing the human exploration's natural capabilities.

Given our approach (embodiment) to the application domain (interactive sonification), we believe that a technological support for the described research goals should provide a solution for the following conceptual requirements:

- The transposition of certain compositional concepts present in electroacoustic music theory, namely Schaeffer's sound object theory [27], Stockhausen's concept of unity [28], Wishart's dynamic morphology [160] and Smalley's spectromorphology [136], to the interactive sonification domain.
- The possibility of a dynamic exploration of data, so that virtual object mediation between data and sound can be handled from different perspectives.
- The possibility of using scope variation through virtual objects in order to allow the manipulation of multiple levels of detail present both in the auditory representation of the analyzed data.
- The use of an embodied interface in order to connect the bi-directional top-down/bottom-up processes of human cognition to scope variation.

From the discussion presented above, it is clear that the research context and goals entails specific technological needs. It is our proposal that such needs to be met through the design and implementation of a software framework for Virtual Reality based multimodal prototyping. Following a user oriented investigation methodology, top down considerations such as cross-modality representation and interaction through virtual object mediation are accessed and finetuned through use case driven bottom up exploration. Both the design and the technology used to achieve this goal would have to be open to the inclusion of heterogeneous systems, in order to take advantage of such systems' expertise and state of development in their specific field of action. Furthermore, the resulting software should not be closed to the original field of interactive sonification, as its application in other related areas will certainly contribute to its further functionality development and fine-tuning. Finally, in order to comply with the above requirements and accommodate future extensions, the design and the technology used would have to provide a compatible solution in terms of real-time performance, scalability and portability. For example, the use of web based access and mobile compatible technologies can provide a suitable base for these two later items.

2.2.3 Multimodal interfacing technologies

Within the auditory display realm, the need expressed in [94] for a generic approach to data sonification tools still maintains its validity today. In fact, even within the sound and music computing domain, there are only a few environments that partially address this issue. An example is the Max/MSP/Jitter and its open source fork, Pure Data. Both these environments offer a very good platform for real-time audio and 3D rendering through their interactive, visual programming paradigm. They also provide an extensive code base for multimodal devices given its popularity in the design and art communities. However, the often complex and scattered resulting code and the expansion through C/C++ external programming constrains the reuse, extension and exchange of the implemented prototypes. Being so, the need for a software development that is based on an extensibility premise is still a pertinent one. This extensibility awareness is fundamental for maximizing code reuse. Additionally, this gap is even more visible when one considers the existence of software environments that enable multimodal interaction together with embodied mediation concepts. Again, it is our opinion that, in order for an environment to support these concepts, it should adopt from the start a modality integration strategy in its core design and implementation process. In this way, the design weight of the original modality (e.g. audio synthesis and manipulation) on the system's future development stages is minimized.

More broadly, the search for effective tools for development and prototyping of applications which apply virtual, mixed and augmented reality user interface paradigms has been an active topic of research in their related fields. Several reports can be found in survey articles [46] [51] [77] [119] [144], which give a panorama of the technological efforts in the past decade. Given the vast bibliography available and the wide range of target scenarios, we will confine our discussion of the state-of-the-art to software packages that are closer to the focus of this article. The review (see Fig. 19) concerning the presented technology is both based on the survey articles as well as the system's original papers and websites, when available.

System	Main application context	Technology	Management	Architecture	Comments
Dwarf	Interactive visualization environment for remote monitoring	CORBA	-	P2P Component-based Event sink/sources	Lightweight middleware Runtime configuration
Studiers tube Open tracker	Augmented reality	C++ Windows, linux and android	Open inventor scene graph Session manager as mediator among hosts	Distributed Master/Slave Observer	Limited portability project ended 2008
Squidy	Multimodal device integration	Java	-	Data flow Multithread	Runtime configuration Semantic zooming interface
Open Interface	Multimodal device integration	.NET Windows Components in C/C++/Java/C#/Matlab/Python	-	Data flow Component	SKEMMI graphical front end Linux & Windows
JReality	Mathematical visualization and animation interactive art and installation	Java JOGL JACK JavaSound Ambisonics ...	Scene graph	Thread safe Rendering A/V backends CAVE distributed rendering Spatialization	HI management integrated with main core Closed inter-core update modules
JMonkey Engine	Framework for 3D game development	Java LWJGL JBullet Spider monkey	Scene graph	Modular Shader based architecture Game logic utilities	Desktop and monolithic oriented design
Max/MSP/Jitter Pure Data/GEM	Audio processing Interactive music performance Creative coding	C++ , C, Java, JScript	-	Graphic data flow interface Expansion through modules (externals) development	Large extension codebase No strong organizational structure (i.e. OO) Commercial (Max/MSP)
Corsaire	High level multimodal applications Bioinformatics solutions	Pymol OpenGL OSC VEServer Max/MSP ARTrack	VEServer's scene graph	Component based (Command; Scoring; Rendering; Supervision) Distributed execution	Dependencies between components is unclear
VHD++	VR/AR domain Virtual character simulation emphasis 3D animation and interaction engine	C++ framework Python/Lua CORBA Windows Linux	Runtime application graph External scene graphs (ex. OpenSG) Share data via properties	Pluggable services oriented components for heterogeneous technologies encapsulation Swappable rendering engines XML config.	"Glue environment" "Common vocabulary" Unclear distributed capabilities Windows and Linux only
VRJuggler	VR applications	C++ framework OSG, OpenAL, ...	Relies on visual engine's scene graph	Rendering engine independence Distributed rendering support Decoupled modules	High portability and scalability "Glue environment" no state management
Processing	Image and animation authoring environment	Java JOGL Other through libraries	Scene graph through external library (?)	JOGL based rendering engine External libraries	Large extension codebase Scripting to Java prototyping No internal (scene) manager
CHAMBRE	Integration of real and virtual world sensors/effectors and multimedia environment	Java Java RMI JMF Java3D	Java3D scene graph	Components with communication interfaces Observable/Observer	Close integration with Java3D
Open Framework	Toolkit designed for creative coding	C++ OpenGL, rtAudio, among others	-	Similar to Processing (see above) Extension through wrapping add-ons	High portability (Win, OSX, Linux, Android, iOS) No internal (scene) manager
OpenNI	API for writing Natural Interaction applications	C/C++ API Win32, Linux, Android, ARC Java/C# .Net bindings	-	Middleware and device abstraction via modules Production nodes and chains for data processing	API wrapping strategy No internal (scene) manager

Fig. 19 A comparative analysis of multimodal interfacing technologies

An overview of the different systems shows two main tendencies. First of all, most of the solutions in the sound and music computing realm focus on the development of specialized tools within one modality scope (e.g. audio synthesis). Furthermore, even if they incorporate more than one modality, this results from an expansion effort (e.g. (visual) game engines) rather than from a top down methodology that aims at embracing the whole modality range and the research of commonalities and specificities between their nature and/or applications. This mono-modality often leads to a dependency between the system's original modality and the rendering implementation used (e.g. CHAMBRE). As a result, custom solutions targeting other modalities different from the system's primary one will more likely reveal component encapsulation issues. Additionally, when a multimodal input routing service is the main functional target, most inspected systems don't provide an independent state management solution. Again, when such functionality is present, it results from extending an already existing managing library (e.g. scene graph) of a rendering engine (e.g. Java3D).

Based on this critique of the existing systems, we believe that a software environment for interconnecting different hardware interfaces and software environments should possess an underlying design that favors a multithreaded distributed rendering per modality, element encapsulation through interface definition and a modality independent state management. As such, our efforts focus on the design and implementation of an object-oriented software framework. On one hand, the choice for object-oriented paradigm is based on the intrinsic composed/hierarchic characteristics of all entities involved in the interaction process (i.e. sound objects, virtual objects and human motor system). On the other hand, the framework strategy aims to provide an implicit architecture that defines the way components relate and interact. As a result, the system conveys a way to reduce the effort for development of customized applications, to provide state management and ultimately, to provide a platform for user-centered, rapid application development. For that purpose, the application of object-oriented design patterns is favored in this environment, contributing to improve quality, reliability, and interoperability of software projects. For instance, all operations that act upon JOINDER's elements should be encapsulated into specific classes (commands). This will establish a standardization concerning both access to the core's elements and inter-core communication. These guidelines are implemented respectively through the command and behavior classes (see Sections 2.2.5.2 and 2.2.5.4).

Further analyzing the previous table, some of the outlined requirements for our research purposes can be found in the referred systems. Examples are the interoperability-oriented design (ex. VHD++), distributed execution (ex. Studierstube; Corsaire), separation between application and rendering backends (ex. JReality), functionality encapsulation through command classes (ex. Corsaire) and software framework as a design choice (ex. VHD++). However, the presence of all the targeted requirements (including portability) cannot be found in any of the freely available systems. Furthermore, although there are quite a number of frameworks that use different systems for each modality, few had their main focus on providing a meta layer for enabling the intercommunication between random heterogeneous frameworks. In fact, in order to comprise the whole multimodal spectrum, the re-implementation effort of the desired functionalities already existing in toolkits/frameworks/environments/languages can be very high. If such external tools are well tested, to reach a comparable level of quality represents an enormous task in time and expertise. Therefore, the key issue here is interoperability. In fact, this design directive was pointed out as a matter of primary importance given the increasing number and specialization of frameworks developed over the last years. As stated in [6]: "To our knowledge this was the first time two AR frameworks were rendered interoperable. We believe that this approach could start a positive trend in the Augmented Reality community. An ongoing discussion of the advantages of interoperability as other frameworks develop and mature can only benefit the field as a whole". However, despite the fact that several powerful frameworks have been proposed in the past, they mainly focused on specific areas within the augmented reality, distributed computing and environmental/ubiquitous computing domains. Although claiming to posses easily interfacing modules, few examples of versatile interoperability achievements have been reported. In summary, instead of trying to provide a comprehensive coverage and related implementation of one particular issue (e.g. multimodal input management), the JOINDER framework aims at developing a system that is capable of uniting expert tools. This target is pursued through both theoretical top down guidance as well as bottom up driven needs. As a consequence, it assumes only at present data representation and functional strategies that are imperative for this integration effort. Throughout its development, the interface specification for the cores, elements and other components followed a need driven refinement while complying with the imposed constrains by global conceptual considerations previously addressed.

2.2.4 The JOINDER framework: general properties

The foundations of our solution is based on the following premises (see Fig. 20):

- The design of a supra framework for providing a common infrastructure and state management between heterogeneous technologies in order to better take advantage of their expertise and advance state of development.
- The application of object oriented modular development and design patterns for the encapsulation of state of the art modality technologies.

In the following subsections, the JOINDER framework will be described on the basis of a number of properties, which include (i) integration, (ii) modular architecture, (iii) Java as enabling technology, (iv) levels of binding and (v) client-server model.

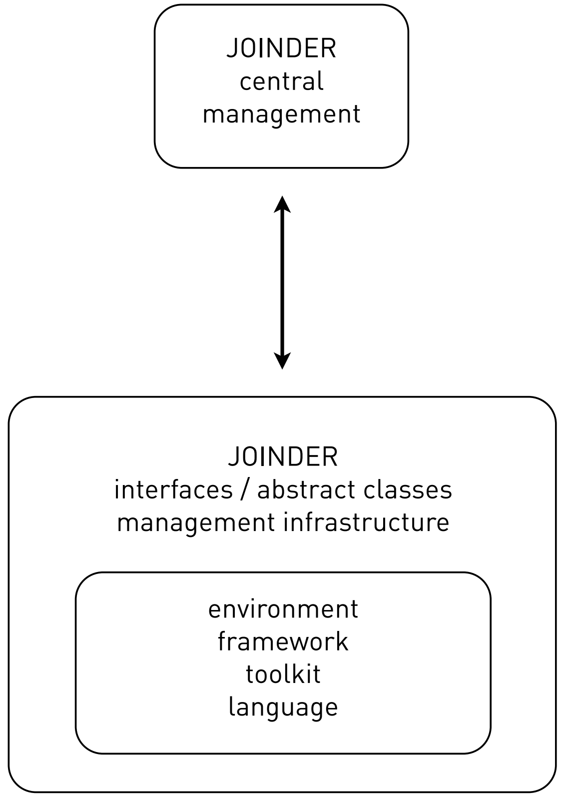


Fig. 20 JOINDER framework's integration strategy for external technology

2.2.4.1. Integration

An important characteristic of our approach concerns the integration of software technologies. In other words, the software technologies offered by the systems described above can be, when possible, integrated as a member of the framework and provide their specialized services (see Fig. 21). This integration strategy aims at not only to take advantage of the high degree of reliability, performance and functionality offered by the state of the art specialized tools but also to allow and promote the recycling of previous developments made with these tools, reducing the implementation effort of future applications.

Also, this approach promotes the usage of technologies adherent to a wide range of hardware devices. For example, a scene can be shared in real time by participants using CAVE environments with optical tracking systems, and by participants using desktop computers or smart phones, all running rendering engines according to their technical possibilities, degree and type of desired participation. The JOINDER framework provides the "translation" service between these heterogeneous nodes (see Fig. 22).

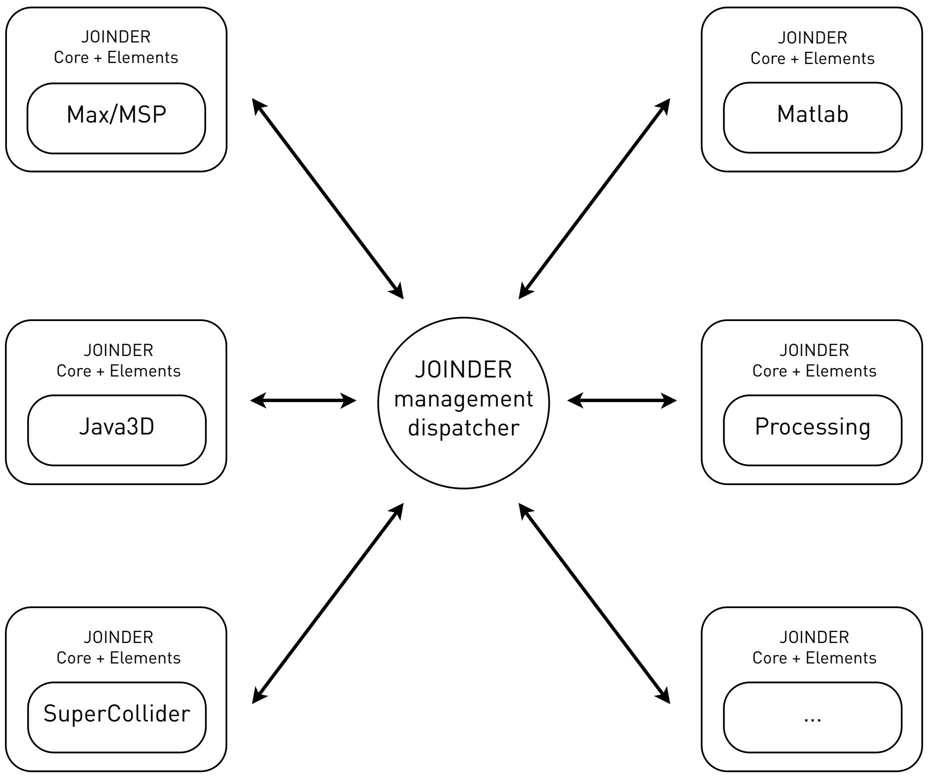


Fig. 21 JOINDER framework provides intercommunication and state management to heterogeneous environments and frameworks.

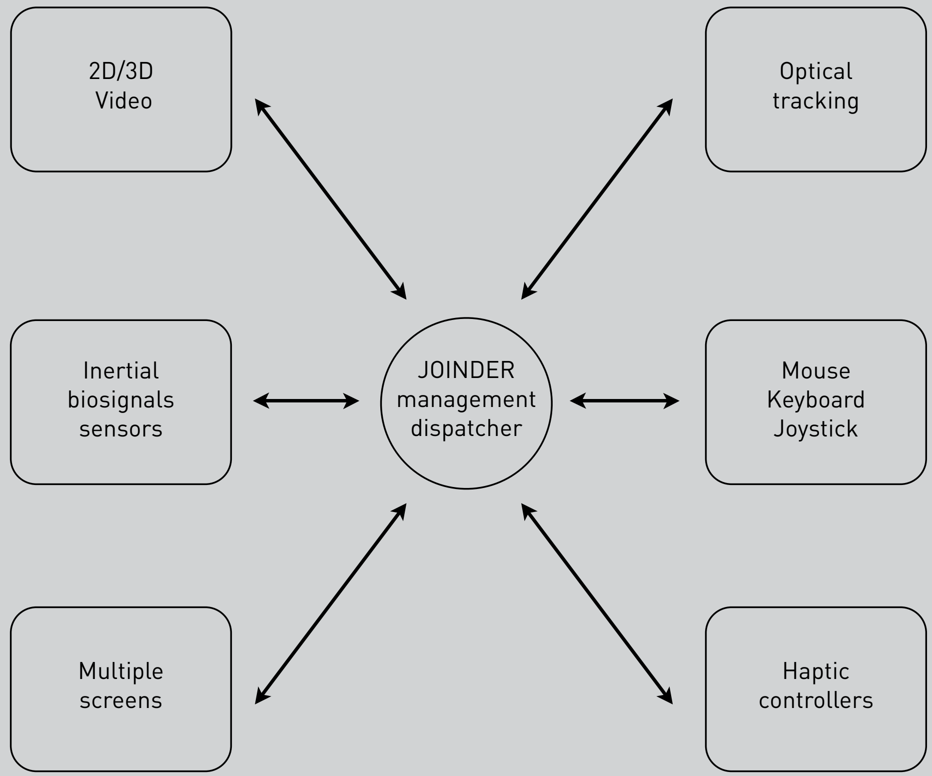


Fig. 22 Through centralized management and common data conversion, multimodal scenarios are obtained through multiple I/O devices integration.

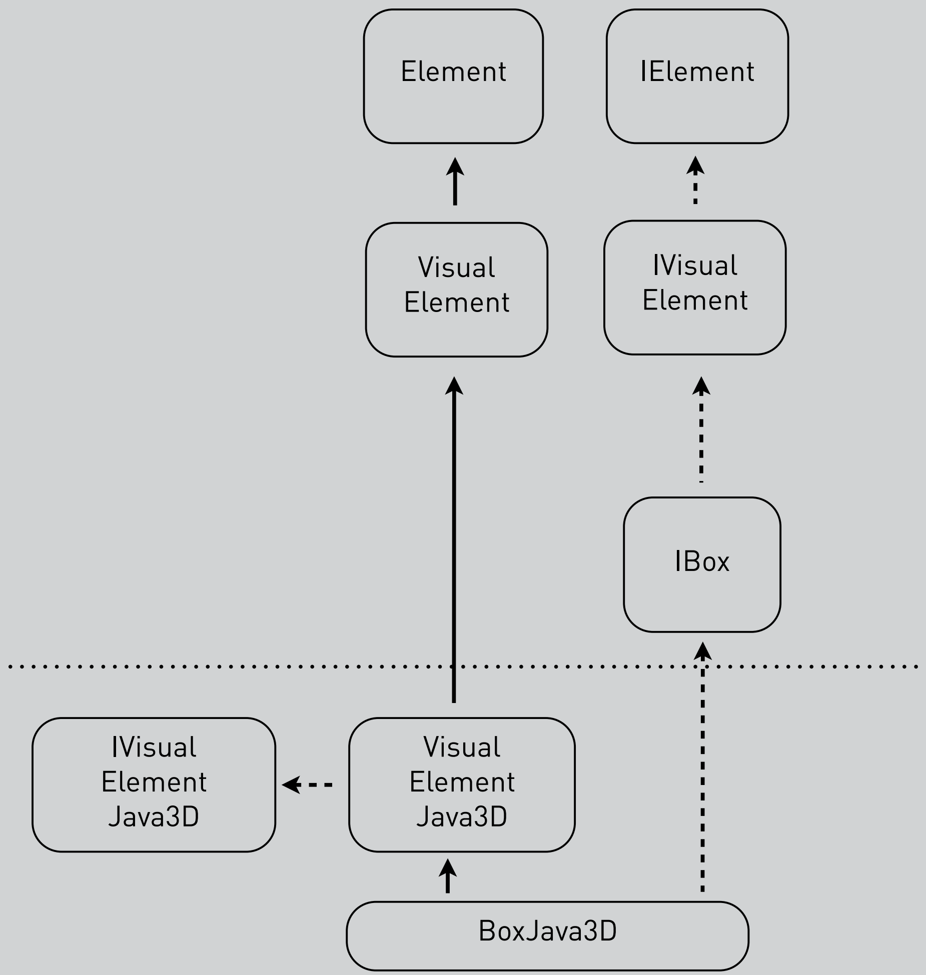


Fig. 23 An example of the framework's class structure. The Java3D's implementation of the visual element BOX implementation based on fixed (above dotted line) and specific (below dotted line) parent classes (full arrow) and interfaces (dotted arrow).

2.2.4.2 Modular architecture

Our system follows a plug-in based structure, which implies that the access of different technologies is conveyed through the use of generic interfaces and factory based class instantiation. Such encapsulation process leads to a loose coupling between the object access and his implementation. In other words, the application development is abstracted from the external software's specificities when using a given element. However, as all the classes are a product of multilevel class extension, one can always access a specific functionality only provided by a certain software package. For instance, if a specific command or action targets a specific implementation (e.g. Java 3D), the class inherence allows the element/core to be type casted to the wanted implementation (e.g. IElementJava3D), which provides access the specific functionality, through a more specific interface (see Fig. 23).

2.2.4.3 Java as enabling technology

Oracle's Java SDK 6 was chosen as the main base technology as it provides the industry standard solutions regarding integrability (Object oriented paradigm; JNI wrapping), flexibility (virtual machine's multiple language support like Python; Real-time specification for Java), extensibility (design pattern prone; various open source libraries), accessibility (standardized language specification; web oriented technology), portability (cross-platform; Web Start), availability (free; open source) and durability (strong rooting in the IT development practice). The application of this base technology allows the framework to operate at different levels of binding with the external software (see Fig. 24).

The four levels of binding are defined as:

- Strong - when the core and its respective elements are implemented using Java based libraries;
- Semi-strong - when the core and its respective elements constitute JNI/JNA wrappers to C/C++ based classes;
- Semi-loose - when the core and its respective elements constitute a network communication interface to the external environments;
- Loose - when specific core and elements do not possess a state and are not implemented. In this case, commands are directly generated and provided to the external environments through behaviors classes (see Section 2.2.5.2).

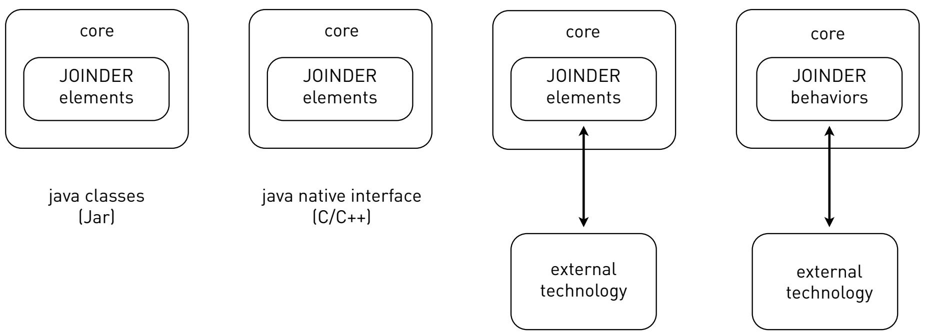


Fig. 24 The different levels of binding for external technology encapsulation.

Through these adaptable levels of connection, it is easy to connect JOINDER with other heterogeneous frameworks and toolkits. An example of loose binding is exemplified in the NatNet protocol integration through the use of NatNet2OSC driver and Natural Points' Arena software while a strong binding example can be found in the use of Java3D (although using JOGL as a low level resource). These different levels of binding contribute to a versatile inclusion of external frameworks and environments. On one hand, it contributes to application level code reuse. For the application code, the command's instantiation and execution process is independent of the element upon which it operates. The latter's implementation can be based on any of the defined binding modes without repercussion to the application's code while providing a best fit for the integration of the external system. For example, a sound element class for rendering a string-like physical model implemented using SuperCollider 3 can be swapped for an implementation using JavaSound for web deployment reasons. Since both implement a generic interface IPluckedDataString that specifies the functionality for both cases (Ex. play()), this change has no effect in the remaining elements (visual and HI elements, triggering observers) of the system. So, while the JavaSound and the SuperCollider implementations would follow different levels of binding (respectively, strong and loose), the access and manipulation of the sound element base functionalities are independent of the chosen binding level.

2.2.4.5 Client-server model

The JOINDER framework follows a client/server model when the virtual core serves as an explicit mediator and certifies that all the information that circulates in the network is valid according to the user's rules specified in the central virtual core. As such, in the case of novel information by the modality cores, the latter transmits the respective command to the virtual core, which can perform validation tasks and in turn broadcasts it to the relevant cores.

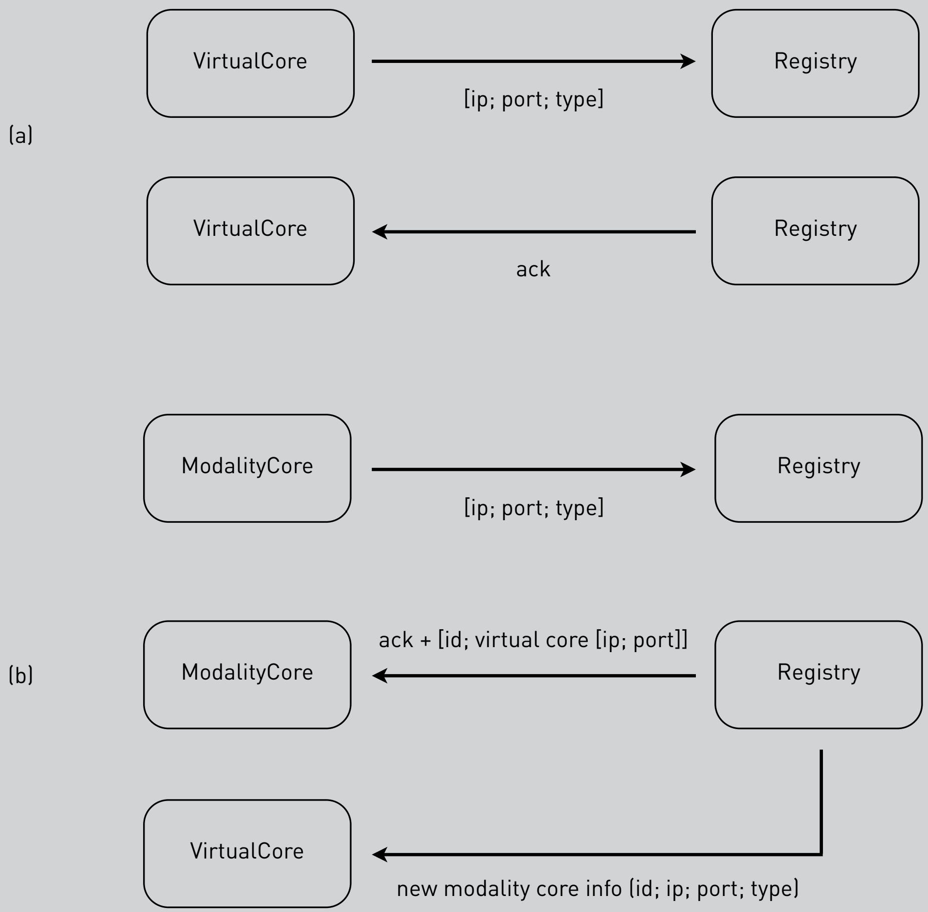


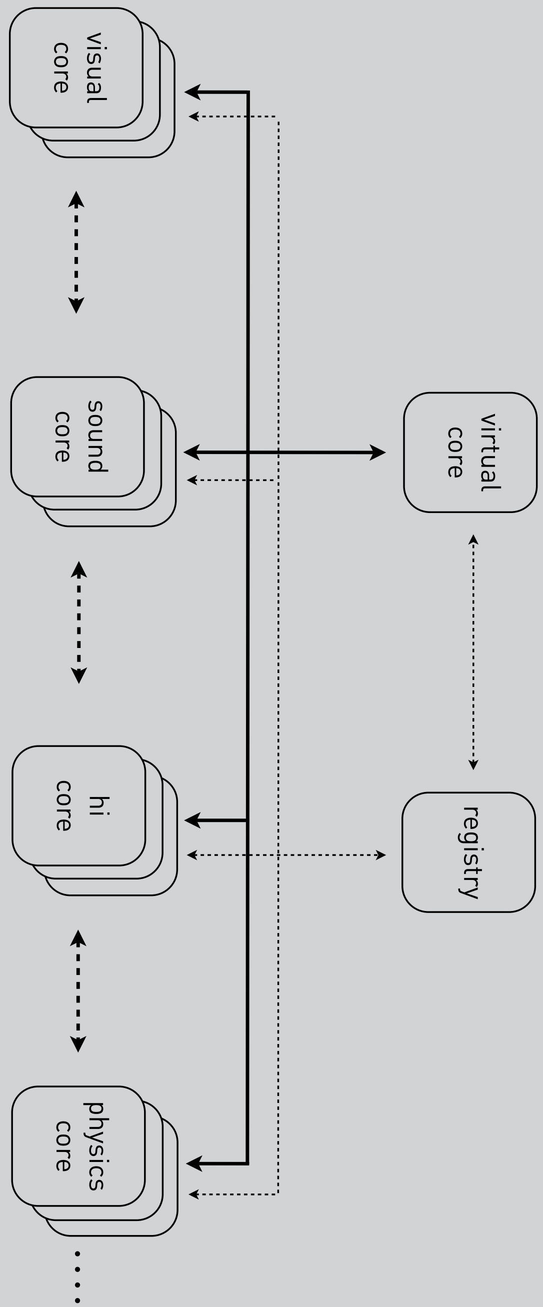
Fig. 25 The Virtual (a) and Modality core's (b) registration procedure in the Registry Server.

2.2.5 The JOINDER framework: specific properties

In the following subsections, the working principles and modules of the framework are addressed. They are (i) synchronization, (ii) representation, (iii) scheduling, (iv) communication, (v) execution and (vi) creation.

2.2.5.1 Synchronization

In order to manage and update the core's routing information in the system, a registry server is provided (see Fig. 25). The registry server is responsible for keeping track of the communication information (IP and connector's ports) of every live core in the system. When a modality core (e.g. visual) is deployed, the registry stores its related information and sends it to the virtual core, which creates a dedicated connector for the new core concerning the updating process related to changes in the virtual elements. This way, the targets of the broadcasting procedure are automatically managed in a transparent way to the end application developer. The communication between cores and registry is realized through the use of Kryonet, an open source library that provides simple TCP/UDP server functionalities [96]. Furthermore, this component can constitute in the future a gateway for generic web based access, further enabling software as a service capabilities of JOINDER through automatic core lookup and access, and for inter-deployment synchronization with other running instances of the framework. The state logger class implements a mechanism for automatically storing executed commands. Although applicable to all cores, it is a default component in the virtual core. This functionality allows for a newly registered core to be set up-to- date in an ongoing interaction session, contributing to the "resources on demand" flexibility through dynamic setup functionality. Furthermore, it constitutes a storage base for session recording and playback as well as session mirroring between distinct virtual cores and respective modality cores.



Virtual Core - responsible for scene managing (all elements) and generic monitoring through behavior hosting

Registry - provides information to cores about location and servicing

Modality Cores - responsible for rendering a specific modality, managing their specific elements and specific behavior hosting

Additional Cores - responsible for rendering a specific modality/behavior

Fig. 26 An overview of JOINDER's network

2.2.5.2 Representation

The JOINDER framework is based on a functional division of modalities into individual branches around a virtual scene representation (see Fig. 26). Following a top-down approach, a first level is composed of abstract managing cores and their respective elements per modality - visual, auditory and human interface. A second level is then obtained by concrete implementations of these cores in correspondence to the external toolkits/frameworks/libraries/environments chosen by their particular capabilities.

A similar decomposition process is also applied to the elements that map the targeted functional implementation. Both the cores and the corresponding elements that they manage implement generic interfaces according to their role in the framework. The resulting abstraction layer, combined with a command-based access, enables the simultaneous use and undifferentiated access between elements through their specific cores independently from the specific library that implements them. So, as a result of this encapsulation, the concrete implementations of the virtual worlds, their visual and auditory representations and the human interfaces that enable the manipulation of the virtual objects can be either refined or substituted according to the desired performance, access or functional needs of the intended use cases.

Further cores may be implemented as provider of extra functionalities and configuration of the interaction context. As an example, collision detection between virtual objects can be provided through a dedicated physics core. Each core in the JOINDER framework manages a collection of related elements through which the user(s) interact. Since all cores inherit from the same common structure, undifferentiated access to each core's managed data is provided. This includes both in and out connectors, executed command cache for remote updating (see

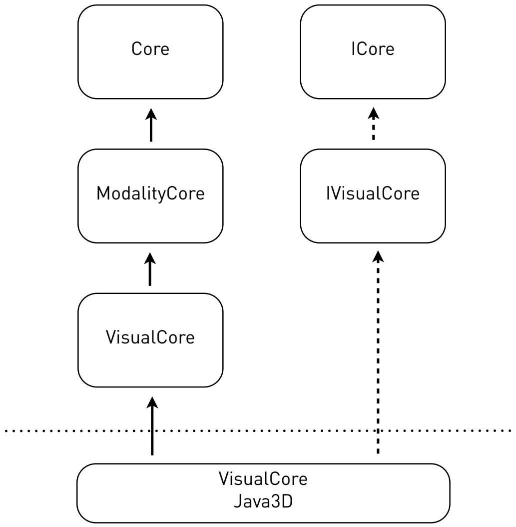


Fig. 27 The class structure of JOINDER's cores, here exemplified through an Java3D based visual core's implementation.

broadcast behavior in Section 2.2.5.3), the behavior and updater managers, the core's element factory as well as registry related information (see Fig. 27).

The following description of the cores follows the package and classes structure imposed by the modality segregation described above. The virtual core holds a data representation of all the elements that interact in the virtual scene. Acting as a deployment global manager for a specific framework configuration, the virtual core enables the data access to the behaviors classes by mapping the state of all elements provided by the modality cores through the implementation of their respective interface. As such, the tasks defined by the behaviors classes are executed upon local and remote modality cores and elements locally in a transparent way, leaving the redundancy synchronization tasks to the internal mechanisms offered by the framework. The default virtual core implementation provides a simple scene graph structure for storing and accessing the state data of the modality core's elements. The Vecmath library [152] is used for 3D attribute manipulation. Their common attributes are provided by a virtual element class, which is extended in order to supply element specific data in conformity with the related modality defined interface specification. For example, a virtual.Box object extends VirtualElement class and visual.IBox interface. However, within the virtual elements' specification, a special type can be highlighted: the Bind element. The Bind element is defined as an element which some attributes are dictated by their children. For example, the position of an array of objects can be defined as the axis average of the positions of its child objects. The modality cores make available, along with their related elements, the concrete implementation of the virtual entities. As of now, visual, sound and human interfaces (HI) cores definitions are defined. Their purpose is to provide a common structure among all implementations regarding a given modality. As such, a specific deployment of the framework may contain, for example, two visual cores implemented through the use of different technologies (Java3D and Jitter), being their access undifferentiated to the remaining cores, elements and behaviors.

Although the human interface (HI) elements retain the basic generic element interface, they differ from a modality element as they perform a different role in the JOINDER's architecture and consequently do not possess scene graph related properties. As expected, the instances of HI elements provide data input for the active participants in a scene, i.e. the modality elements. For that purpose, each HI core holds ID information of the existing virtual elements for input target assignment. This list is updated each time the AddElement command that results of an updating process originated in the virtual core. The addition of an HI element though is related to the same respective command only when executed in an HI core. For executing the binding process, a set of virtual elements can be declared as targets of an HI element. This is done through the AddHITarget command. After certifying that the target candidates are valid, through the HI core, the HI element enters the elements as subscribers for his data source

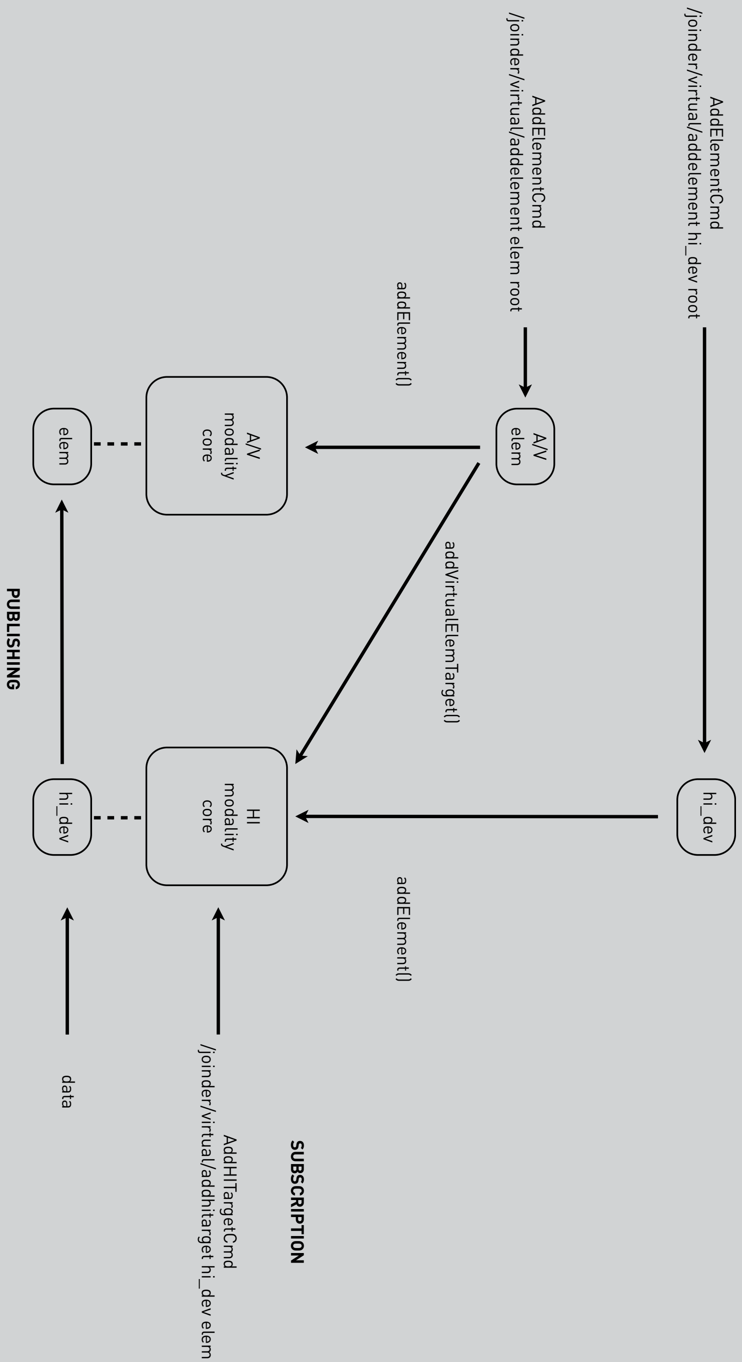


Fig. 28 The instantiation, subscription and publishing processes related to a human interface (HI) element.

and eventual manipulation and/or filtering operations. Finally, upon new data readiness, the HI element executes the publishing process through the creation and broadcasting of commands for the targeted element's update (See Fig. 28).

Even though the HI core does not always constitute a managing point for a modality in the strict sense (like visual or audio), it follows to a great extent the general core structure specified by the framework. By placing the HI core conceptually alongside the modality cores, modularity and versatility features are gained concerning the connection and input management of hardware devices. For example, the adoption of the common distribution denominator makes it easy to use locally available devices in a remotely located scene (and vice versa) by detaching input management from global deployment management.

2.2.5.3 Scheduling

Synchronous and asynchronous tasks are encoded through Behavior and Updaters instances. All of the classes described below can be extended or re-implemented transparently since they all implement related interfaces and are instantiated in appropriate factories.
A Behavior is defined as the task execution component of the framework, being implemented according to the observer/observable pattern. For instance, whether awaken at a given user defined rate or asynchronously due to an exterior event, a behavior instance can monitor the state of a given set of elements and, depending or not on a certain condition, generate a set of commands. As an example, consider a behavior who's task is to activate a sound element if two virtual objects are below a given distance threshold. An Updater instance triggers the latter condition check in the Behavior class every n milliseconds. If a condition is met, the behavior is responsible for generating and scheduling the proper command for the activation of the sound element. The behavior

subclasses extend the Observer interface. The behavior classes can function in close articulation with the updater classes but are not compelled to have a one to one relationship with the latter, as a single Updater class can serve as a timer for several Behavior instances' execution. Management wise, a behavior manager is a container class for behaviors. It's responsible for the creation, removal and access to all active behaviors in a given core. The broadcast behavior constitutes the heart of the updating mechanism, which insures the representational coherence of the virtual scene throughout the modality cores rendering. This behavior is instantiated on the launch of every core. At a user defined refresh rate, all the commands that have been executed in the respective core and stored for broadcast are parsed by modality and send to all the applicable cores with which the core has active connections. The information flow between the virtual and modality is bidirectional as every core can generate instruction that affect the network, either from behavioral execution or user/external input. An example of this can be found in the human interface core, which manages inputs from external hardware sources into the framework. Another example can be found in a collision detection behavior, which can be hosted by a visual core. In this case, in order to prevent the loops occurring when a modality core sends a locally generated command to the virtual core (which then will forward it to the registered modality cores), a sender id was added in the command class, identifying from which core originated the instruction or if it's source was external (e.g. user). The sender id is then equal to zero.

Finally, as stated before, in order to prevent bottleneck effects when multiple cores send information to the virtual core at high rates, the virtual core possesses a dedicated threaded connector instance (along with a dedicated port) for each modality core belonging to the network. Consequently, the incoming actions perform updates to the information managed in the virtual core in a concurrent, synchronized way. An updater is a class of objects that extend the observable class, being implemented according to the observer and singleton patterns. Its envisioned main function is to trigger or activate the execution of one or more behaviors. A simple example of this class can be a single thread that triggers the execution of a set of behaviors each n milliseconds (performing a similar task as the metro object in the Max/MSP environment). The updater manager supplies and manages the updater's instances to the system. It functions according to the Pool pattern by which an updater is created if the required rate is not already supplied by an already existing one. If such is the case, a new updater is created. If that is not the case, the behavior is added to the observer list of an existing updater instance. Additionally, if at a certain point an updater is no longer need (i.e. its observer list is empty), the instance is terminated and removed.

Alternatively, the monitoring of elements and their properties can be processed through callback hooks. In this situation, all elements possess an observable field. The observer behavior registers in the respective elements he wants to monitor. Upon change in field, the observable notifies all observers and supplies the updated value for the monitored property. This option can improve performance in the cases where the change in the element's property is not regular as it avoids the unnecessary lookups from the method above.

2.2.5.4 Communication

The Connector package implements the high level syntax with classes that allow intra and inter core communication. It is accomplished through two main interfaces, the IConnectorSender and IConnectorReceiver. Classes that implement these interfaces have to provide translation and execution methods for their specific communication protocol. The provided implementations are examples of this. This allows each core to receive and send commands to other deployed cores or external user apps in a transparent way, independently of the specific protocol or technology that the connector's implementation uses. At present, two implementations of connectors for command transmission are provided, java object serialization and OSC (Open Sound Control protocol) through Java-OSC [78]. The latter supplies an interface to a commonly used protocol in the sound and music computing research and development field.

In addition to the class infrastructure made available by the JOINDER framework, further types of implementations can be added to give response to additional protocol or design needs for a specific deployment configuration. For example, the inclusion of web servicing through Remote Method Invocation (ex. Java RMI; CORBA/IIOP) for intra and inter deployment communication can be implemented without any repercussions to the already existing connectors. Likewise, other web servicing paradigms as REST architectures can be incorporated in parallel with the implementations above mentioned. This expansion path will be further addressed in Section 2.2.8.

2.2.5.5 Execution

The command specification and its class hierarchy is a reflection of the modularity requirements of the described design. Therefore, all actions are defined as commands classes as these constitute a processing unit to be executed by the cores. This allows the encapsulation of generic and implementation specific core functionalities in a self contained executable, serializable and log-enabled item. Additionally, it provides the composition of actions through macros and class inheritance and an easy conversion base for multiple protocol communication. The command classes encapsulate all the accessible functionality applicable to cores and elements, as well as core subsidiary components such as behaviors, updaters and connectors. A command is therefore an instruction and related data container that is executed by the framework's cores upon reception. This design choice makes it easier to handle and implement a common task execution procedure. Additionally, it simplifies and structures the generation of instructions that induce state changes. Consequently, operations as instructions logging for state replication, recollection or analysis or "jump to state point" scripting are standard features. As such, the command design pattern is closely related with the updating mechanism present in the core classes. Besides holding all the information necessary for its execution (ex. core reference, execution parameters), all commands implement a common interface, ICommand and extends Command abstract class, which implements both Serializable and Runnable interfaces. The Command class specifies the execute() method, which serves as an execution macro. It instantiates the thread for running the command class that extends it, which on its turn implements the run() method of the Runnable interface. The usage of a command driven architecture furthermore facilitates thread based execution, networking (and related code mobility) and parallel processing, three base features targeted by the JOINDER framework.

2.2.5.6 Creation

The abstract factory pattern is applied in the creation of objects factories. This includes not only the creation of cores and their respective elements but also commands, behaviors, updaters, connectors and auxiliary packages (ex. OSC library wrappers).

We have defined two types of factories: instance and static. The instance factories are specific to the core type and implementation as in, for example, the core's element factory. For example, a "Box" entity is implemented by virtual.impl.Box in the virtual core and visual.java3d.BoxJ3D in the Java 3D

implementation of the visual core. The static factories are transversal to all cores, elements and further components in a given deployment of the framework. Nevertheless, the status of a given static factory can be changed to instance on need to basis through the core class inherence by adding the relevant field and access method and refactoring of the framework's Factory class. For instance, a custom configuration might need to define the Command and/or the Behavior factories as core specific. In order to be able to easily exchange implementation classes of components, the factories in the framework execute their class instantiation through reflection. Reflection is a process available in several programming languages, including Java, by which software can alter its behavior at runtime. This is often achieved through instantiating and using a class without explicitly coding its definition type.
So, each JOINDER's component class is loaded on demand according to its specification on the configuration XML file and its usage is conducted through interfaces which the class implements. However, in order maintain the performance requirements, a cache is provided for reducing the reflection class lookup's processing costs. The factories perform only a class lookup, storing the respective class object in a cache. After this first instantiation, the reloading of a class can occur only by the explicit execution of two commands, LoadSystemFactoriesCmd that targets the static factories and LoadInstanceFactories Cmd, which targets the instance factories of a specific core.

All component class's names are specified in a XML file and accessed through JDOM. This file constitutes JOINDER's main configuration tool. Through the loadXMLFile method in the Factory class, it is possible at any point to reload the XML file. This allows the class redefinition of a given element, for example. Further configuration parameters are supplied through properties files.

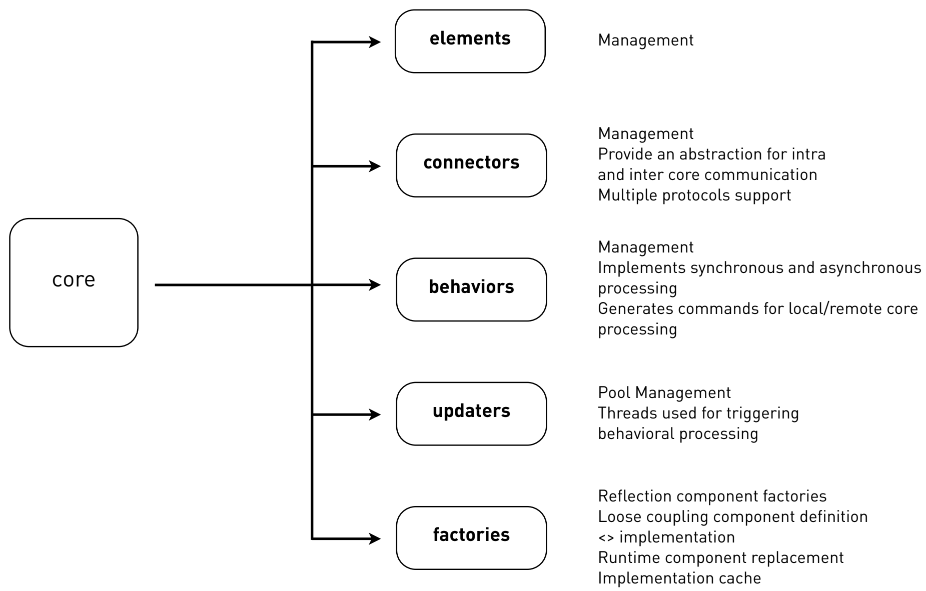


Fig. 29 A summary diagram of all components of the JOINDER framework

These normally are specified according to user needs. The previous section described the framework's modules. A summary of the scheduling, communication, execution and creation components is presented in Fig. 29.

2.2.6 User-centered development

JOINDER has been developed on the basis of a user-centered development cycle that was based on several case studies. A first series of case studies had a focus on interactive sonification and includes work in virtual string interaction, data exploration/monitoring [45] and dance [43]. A second series of case studies had a focus on the SoundField art installation [39].

2.2.6.1 Virtual string interaction

This case study targeted an assessment of the mediation principle described above, in which the user's body parts were extended with markers that allow the manipulation of objects in a virtual environment. In addition, this case study tested both the design and the performance of the framework's early implementation. The case study implemented "playing" a virtual string object, using two virtual objects that are controlled by the user's hands (see Fig. 30). Thus, a physical space was mapped onto the virtual world in which a cylindrical object representing the sound producing string was positioned in the middle. Through 3D positioning and orientation, by means of optical tracking, the users could "pluck" the string when it collided with the users' tracked "hands" (also represented through virtual objects). A scheme of the software and hardware systems and components used in the virtual string use case is depicted in Fig. 31 and an illustration of the data flow regarding this use case is shown in Fig. 32.

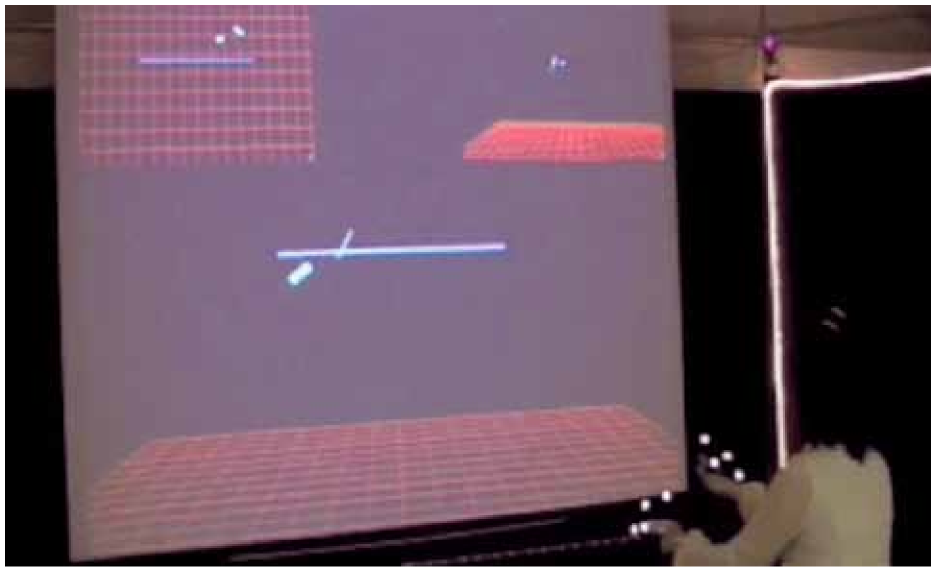


Fig. 30 User interacting with the virtual string prototype

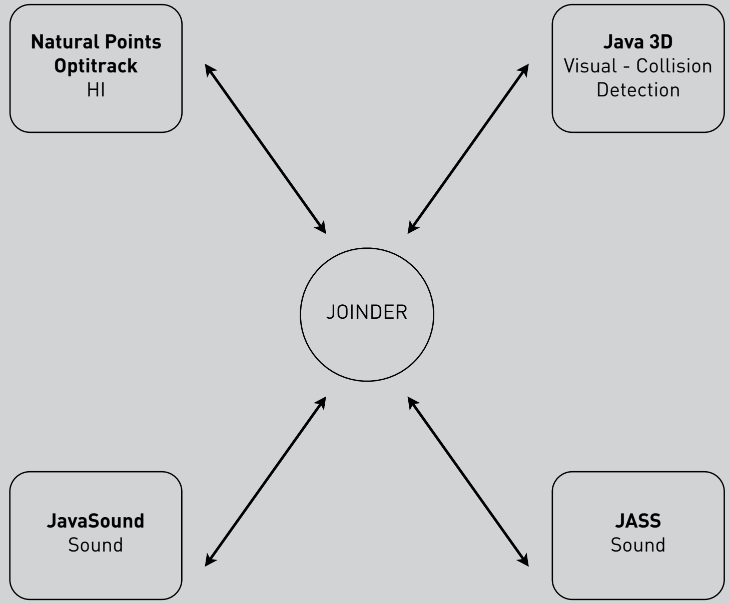


Fig. 31 Technologies used in virtual string use case

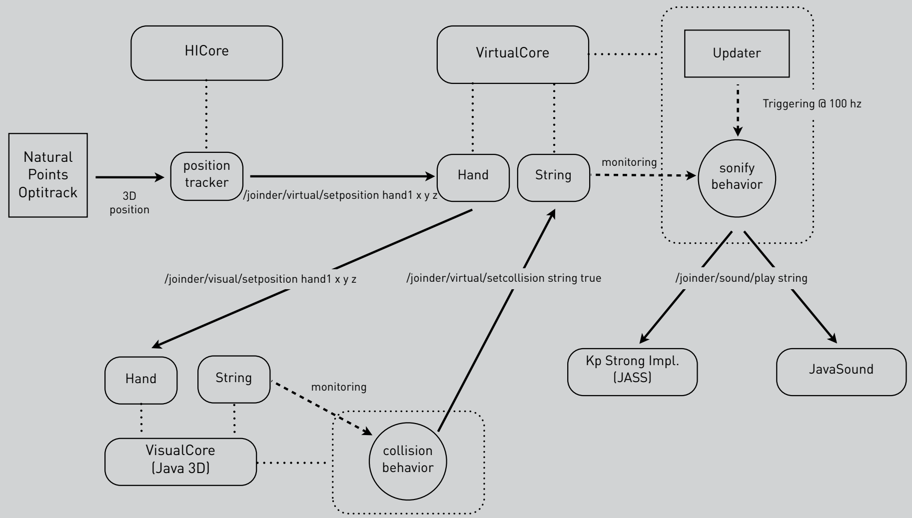


Fig. 32 Data flow representation in the virtual string use case

2.2.6.2 Data exploration and monitoring

In a second case study, a sonification system was developed that targeted the interactive exploration of multivariable data through non-speech audio communication [45]. The goal of this implementation was to investigate the possibility of establishing a unified context between individual sound streams that are exposed simultaneously through time. The sonification paradigm is based on concepts from electroacoustic musical composition. The inspection process was conducted through the interaction with virtual objects in an immersive 3D environment (see Fig. 33). This case extended the study of the relationship between the users' embodied behavior and the virtual entities on one hand, and the different levels of sonic objects on the other hand, through the exploration of a test dataset. The conceptual design was founded on the combination of two main metaphors, namely, the virtual inspection window, providing access points to the variables and their values belonging to a given dataset, and the virtual microphone, which implemented a sonic inspection tool controlled by the user's hand. This approach is inspired on Stockhausen's work Mikrophonie I [141]. Each independent virtual object belonging to the virtual inspection window functions as a sound source. The sonic representation of its value was activated through collision detection when the inspection volume of the virtual microphone intersected the virtual objects. Moreover, two additional sonic representations of the relationship between the data's values were made available, interval and chord, in order to provide a sonic feedback of numeric proportions in the data. In addition, the position of the activated objects in the inspection volume of the virtual microphone influenced the respective output in terms of auditory relevance. This was implemented through distance based amplitude and reverberation real time adjustment, as suggested in [141] [160] concerning distance simulation. In summary, the adopted interface paradigms conveyed

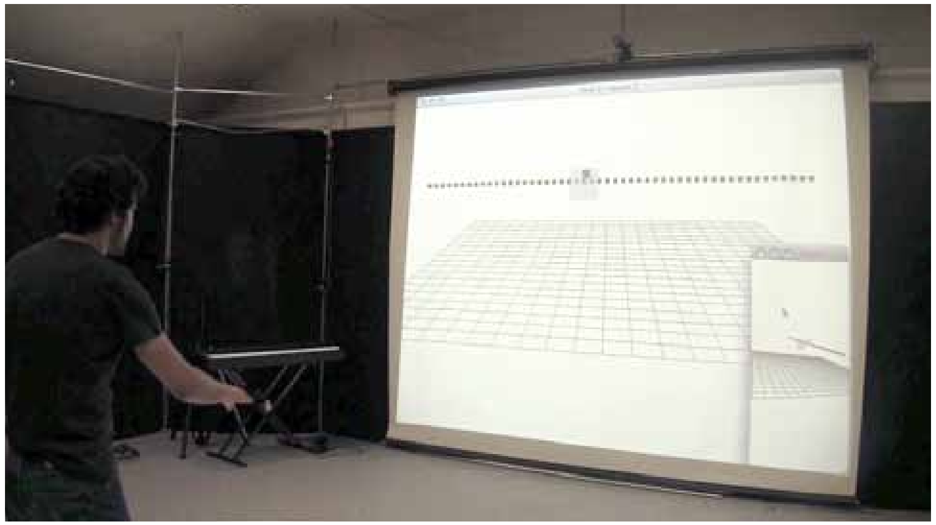


Fig. 33 A user sonically exploring a dataset in an immersive environment. The room in which the user stands is depicted through the projection screen. The virtual scene contains the virtual inspection window (array of square shaped elements), which is activated (red elements) through collision detection with the virtual inspection tool (pyramid shaped element).

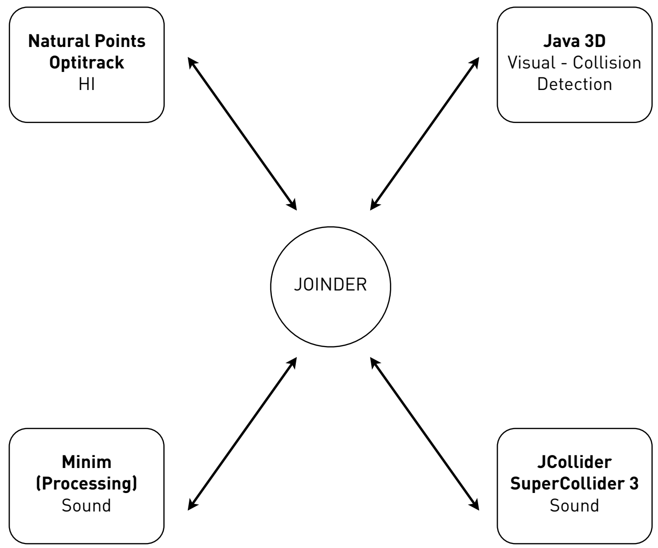


Fig. 34 Technologies used in the data exploration and monitoring use case

multiple perspective views between different levels by representing the evolution of the sound object/data in time and in structure, and by establishing and comparing different groupings of variables. A scheme of the software and hardware systems and components used in data exploration and monitoring use case is depicted in Fig. 34.

2.2.6.3 The dance use case

The dance use case [43] was derived from the previously described prototype. This time, the array is scattered in space, enabling users to vary their inspection in space and time even more. An extra bend sensor and wireless ADC allowed for the participants to vary the scope of the virtual microphone by changing their posture from an open (global) viewpoint to a closed (detailed) one. The sound diffusion of the microstructures was directly linked to the user's scope variation approach (the inspection vector), using sonification ideas based on Smalley's concept of spectromorphology and spatiomorphology [160]. Here, the interactive sonification implies the use of space in relation to a body-centered spatial exploration over time. (see Fig. 35)

Fig. 35 Users executing the dance use case. The first user (on the left picture) performs his dance choreography, which was later analyzed and mimicked by the second user (on the right picture).

The dance use case consisted of two parts, namely, a setup part and an exploration part. In the setup part, the data objects are set out in space. Here, the data objects represent dance movements. A sequence of these data objects thus represents a trace of the dance movements, as an occupation of space. In the exploration part, the user is given specific tasks in order to explore the trace of the initial movement. The users had no visual representation of when the virtual objects creation took place within the choreography. Besides indicating the direction of the path through an increase in pitch, variation in the velocity of the original movement was conveyed by the variation of the pitch's increase. This strategy is inspired on Stockhausen's take on Information Theory [70].

A brief note concerning the need for multiple views visualization system. Within the visual modality, multiple views are regarded as an integrated feature of the visual modality within the JOINDER (IView). As a result, this feature can be applied for multiple display scenarios, either one view per visual core/machine (in case the complexity of scene and/or the rendering performance so requires) or simply multiple views in one single visual core instance. Furthermore, several instances of the same conceptual view can be grouped, providing a distributed, easy to manage functionality for group collaboration.

2.2.6.4 Preliminary evaluation

The user centered development strategy was implemented throughout the use cases above described. Participants were asked to perform exploratory tasks in order to evaluate the human interface used in terms of performance, maneuverability and precision. Additionally, they were asked to comment on the visual output completeness (used to find and interact with the virtual objects) and on the sonification output in terms of distinguishability, information carrying potential and aesthetic design. In all use cases, the users reported having no problems interacting through the virtual object's metaphor after a brief introduction and exploration of the interface. This opinion was maintained through all use cases, despite their increasing level of complexity. However, some recommendations were made in the two initial use cases for future development stages. For instance, some participants indicated the need for additional screens that would convey a better depth perception. This would facilitate the adaptation to the spatial 3D nature of the interaction (e.g. finding the virtual objects). Also, it was mentioned that the virtual objects' morphology should be dynamically configurable during the interaction sessions. These observations led to a further refinement of the framework element's class structure, fields and access in order to comply with these requirements. Consequently, all these issues were addressed in the dance use case, in which more manipulation possibilities were given to the user resulting in a more controllable and fulfilling experience. As mentioned, real time configuration requirements were recurrent in both initial use cases (ex. dynamic morphology configuration). Additionally, the inclusion of multiple objects was desired, given the inherent collaboration aspect prone to the immersive environment used, and requests for more aesthetically attractive visuals were made. During the related implementation process, a performance decrease was observed on the initial single laptop running prototype. In order to prevent the system from reaching levels of latency that would be incompatible with the interface's real time compliance, an architectural improvement concerning the cores' modularity was undertaken. In accordance with the required performance and platform flexibility goals, the rendering cores were made independent and completely detached from the central virtual core. As such, the rendering computational weight would be spread over multiple hosts, in a fully distributed system. Finally, concerning the sonification levels in use cases described in Sections 2.2.6.2 and 2.2.6.3, it was proposed that the relative distance of the virtual microphone and the object(s) under inspection could also be used for the activation and mixing of the sonification planes. This feature would be made available as an option to the standard regulation of the amplitude and reverberation parameters. These requirements also led to a refinement of the modular design to improve the communication between sound and human interface elements through the virtual core. Also, it provided structural directives for the definition and access of task specific packages (e.g. sonification mapping classes). Furthermore, additional changes in the factory classes were adapted to use reflection for enabling dynamic switching between core/element's implementations.

2.2.6.5 SoundField Art Installation

In order to develop JOINDER as a rapid prototyping tool that would be useful by different users, there was a need for a broader, more ecological approach that would allow the development of features through a participatory design strategy [39]. In contrast with the predefined media setup and fixed content of the previous case studies, a participatory design strategy provides a greater freedom to what users want to do with different media and different contents. In this approach, software development in collaboration with users typically led to a focused and comprehensive bottom up development of software given the top down premises concerning background, mediation paradigms, and other preliminary decisions (such as choice of different media technologies) (see Fig. 36).

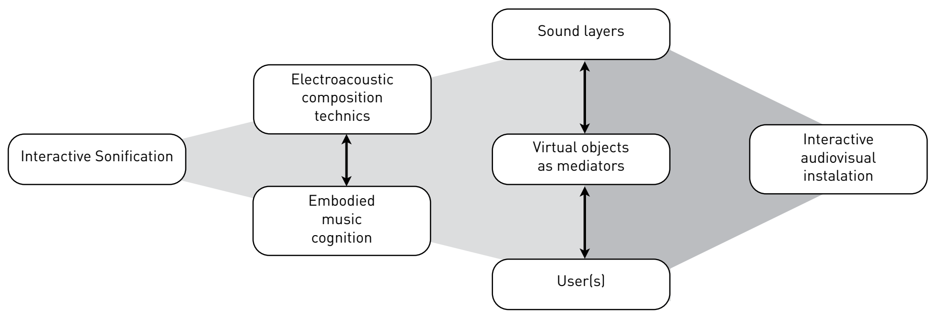


Fig. 36 The initial domains of the JOINDER framework were expanded through the realization of SoundField, an interactive audiovisual installation.

This participatory design strategy was fully deployed in a case study called SoundField. The first iteration of the user-centered development within the SoundField project context took place at the Destelheide Educational Center in Dworp, Belgium. This project's focus was mainly on the exploration of interactivity in new media arts through social interaction principles [30], extending the virtual objects mediation approach discussed earlier in Section 2.2.2. The Destelheide Educational Center yearly provides workshops and courses in various disciplines such as dance, literature, music, theater, sculpture and audiovisual arts. Upon DHArts (a section of the educational center focusing on arts promotion) invitation, SoundField was presented as a work in progress artistic installation to a broad audience from October 2010 to May 2011, in close collaboration with the Flanders "Youth and Music" program.

During this period, several fully functional implementations of the technology were developed according to the prospective users' requirements, gathered during evaluation sessions (see Fig. 37). These requirements were incorporated (through rapid prototyping) and presented to the same user groups. After another set of tests and interaction sessions, the specifications of the given use case, and the experience of the users with the technology was recorded, in order to incorporate this knowledge in the following user-cases. In summary, the iterative development process, which included 1) demonstration, 2) interrogation, 3) implementation, 4) interaction and 5) evaluation was carried out throughout the entire duration of the project. The use cases developed during the SoundField sessions encompassed various interaction strategies through multiple implementation scenarios. A prototype demonstration deployment, which expanded the interaction possibilities presented in the previous section, constituted the starting point for the development of the subsequent use cases. Here, the envisioned social dimensionality, as a means of establishing cooperative interaction, was implemented

Fig. 37 A group of users testing a prototype implementation during the SoundField project sessions. After a short presentation and discussion, the recommended features made by the focus group were incorporated and evaluated after a second interaction round.

through simple metaphors - point, line and plane (see Fig. 37). From that point of departure, 14 distinct deployments and additional variations were developed in close collaboration with their respective user group, ranging from 3D sound manipulation to theatrical audiovisual augmentation and multimodal rehabilitation tasks.

The user-centered process applied throughout SoundField's iterations provided a methodology for functionality expansion, interface integration and design re-factoring and consolidation. During the numerous sessions at Destelheide, the use cases requirement inspection and rapid prototyping provided cyclic in-house process through which a number of guidelines were formulated. These guidelines mainly concerned the generalization and accessibility of functionalities. They can be partially summarized in additional re-factoring of class structure and fine-tuning some implementation aspects of inter-core communication (ex. dedicated port/thread per core pairing) and task execution performance (ex. command thread scheduling). Still on this latter item, other optimizations included providing a cache for the core/elements factory classes and distance calculation behaviors. The first stem from the necessity to contra-balance the performance toll resulting from reflection based instantiation in use cases with a high creation of objects rate (ex. dance use case). A class cache based instantiation was then implemented to reduce the reflection class lookup delay. This resulted in a considerable performance improvement upon the creation of cores and objects. The second aimed to improve the collision detection performance and limit the use of the Java 3D internal collision detection mechanism to complex forms. This optimization was achieved through the usage of a distance-based behavior when the used virtual objects' morphology would allow it (ex. sphere shaped collider). Variations of the distance calculator behavior were also applied to proximity based activation and VBAP (Vector Based Amplitude Panning) mixing. Generally, the sessions supplied a fertile context for behavior and element gathering and increment, contributing for architecture validity assessment and code base expansion. As in the previous use cases, interaction was said to be fairly immediate and self-explanatory, as latent affordances were easily understood through interface consistence. Such verification stems from the fact that the use cases maintaining a common base of operation while increasing in complexity. This made it possible to make iterative transitions between setups, which involved the incremental inclusion of several software technologies and hardware resources as well as an increase of number of simultaneous participants, and tasks (e.g. from a laptop-based demonstrator to immersive, multiple participant environment). Here, the JOINDER's modular design and scalable technology provided a adequate platform for the expansibility-wise requirements of the project. Both new software elements and environments (ex. Max/MSP) as hardware devices (ex. ICubeX bend sensor, WiiMote) were easily incorporated without any repercussion to the already present technologies, functions and data management facilities.

The Soundfield sessions also allowed consolidating design choices. The updaters and behaviors classes for event monitoring proved to be adequate design choice for most use cases by its recurrent usage straightforwardness. Nevertheless, alternative programming strategies were developed envisioning the optimization of the system's prototyping cycle (ex. the monitoring of virtual object's through callback hooks). The integration of new technology and their respective articulation by elements and behavior specification was achieved in conformity with the levels of binding defined in Section 2.2.4.4. This process helped to circumscribe the applicability scope of these levels through implementing the use cases' interaction needs. For instance, new elements (and respective cores) definition would be reserved to the cases where state maintenance is necessary as behaviors, that communicate directly with external environments such as Max/MSP, would be chosen for stateless communication needs. Later that year, the Artbots festival 2011 provided an additional opportunity for further develop SoundField in terms of new use cases, which led to new technological additions and further implementation refinement, and collect participant's appreciation and suggestions. The Artbots festival 2011 was held on October 6th to 9th 2011 at UFO Ghent University (see Fig. 38).

Fig. 38 Visitors of the Artbots festival 2011 interacting with the presented showcase. As in the previous iterations, new devices and functionalities were incorporated in order expand the framework's code base and test its performance with new interaction techniques.

The participation in this event aimed again at collecting new interaction requirements inherent to a new interaction environment such as an artistic showcase. Consequently, additional interaction techniques were incorporated during this iteration of SoundField. These techniques aimed at further extend the manipulation of the virtual objects possibilities beyond those facilitated in the previous use cases. These new interaction techniques included the "grabbing" of virtual objects, through position proximity to the user.

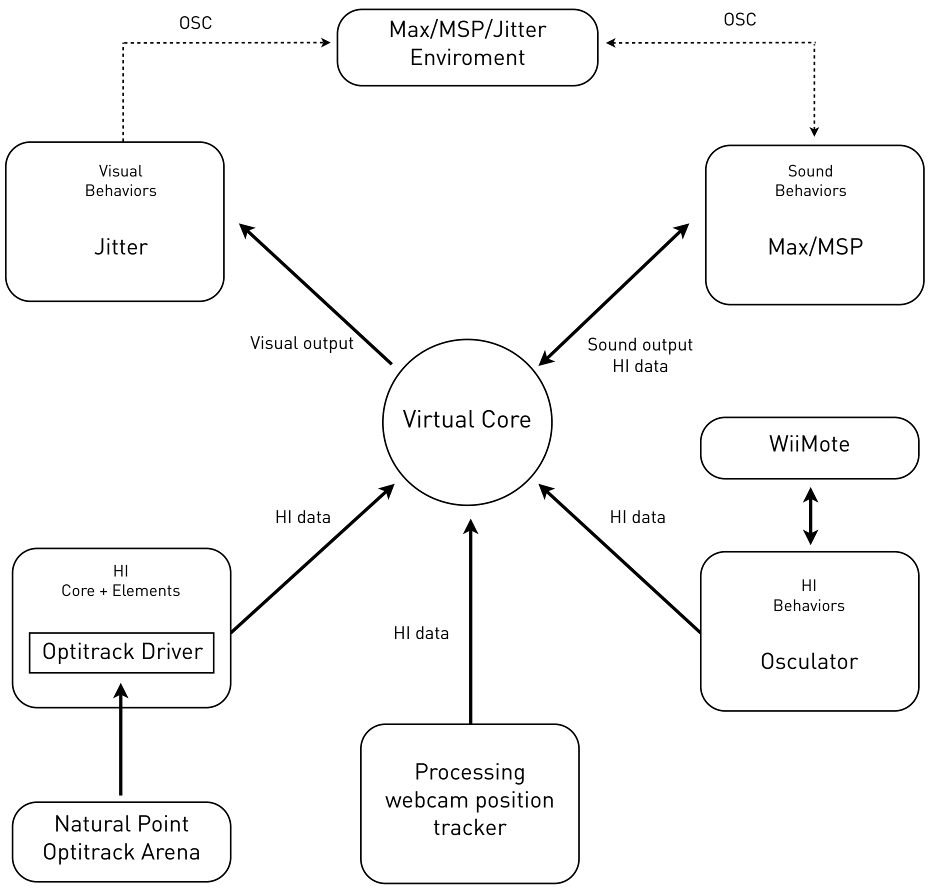


Fig. 39 Technologies used during the Artbots festival 2011 showcase

As in the previous use cases, the addition of this functionality implied the addition of new fields in the Virtual Element class (dataMaster and dataSlave), the development of new behavior classes (to map the grabbing feature controlled by a Nintendo Wiimote) as well as a thorough verification of any incompatibilities to the already available modes of interaction. The different hardware and software setups as well as the data flow concerning SoundField and the Artbots showcase are shown in Fig. 39 and Fig. 40.

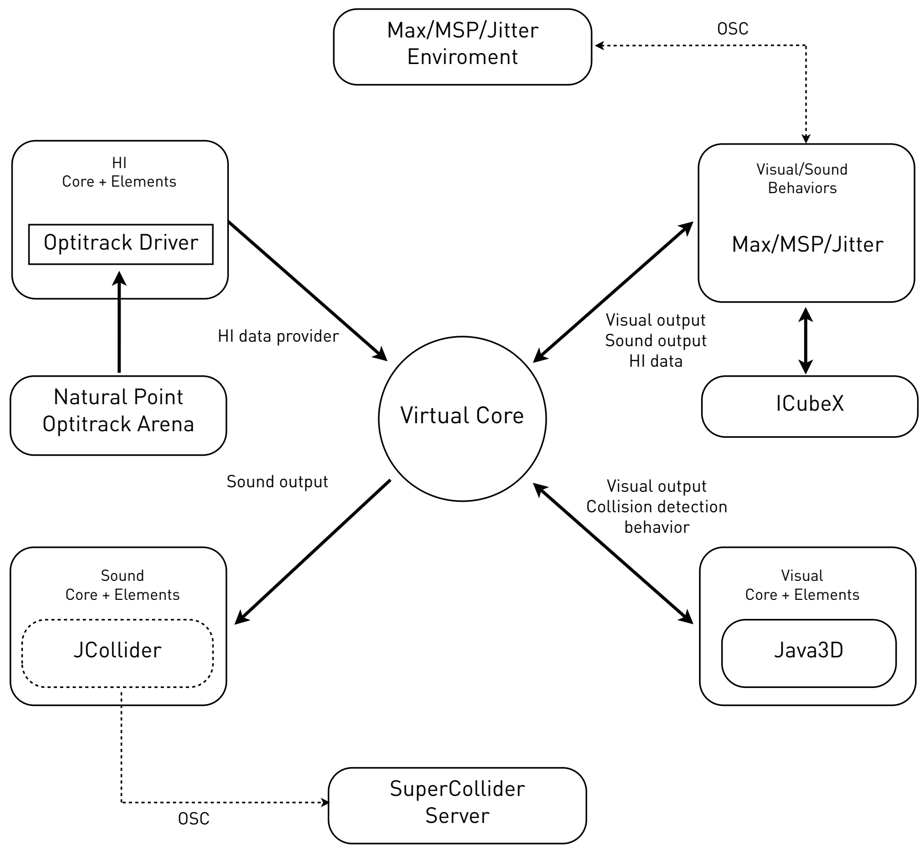


Fig. 40 Technologies used during the SoundField project sessions

2.2.7 Discussion

Up to the current point of development, the JOINDER framework has met the challenges outlined by the research context's requirements. These comprised the development of a software solution that would provide a platform for VR based interactive sonification application development through a user centered oriented methodology. This software solution would have to supply data management facilities and comply with the research field's requirements of multimodal interfacing, heterogeneous technologies integration and real-time performance. Furthermore, in order to accommodate future extensions, the design and founding technology would have to be flexible in terms of scalability (e.g. inclusion of standard web technologies) and portability (e. g. multi platform compatibility). This result originates not only from using a scalable and mature technology but also from the refactoring process of the top-down driven "global" strategies with bottom-up, "on location" use case scenarios. As described above, the use cases provided not only evolutionary means to re-factor the various design and implementation decisions that were followed but also a phased performance assessment throughout the development stages that validate the outlined methodology.

In terms of design, the multi-threaded and distributed core environment was able to deliver solid platform where multiple devices fed data at a high rate (e.g. 100 Hz sampling rate of the Natural Point's Optitrack system) into a concurrent behavior environment, rendered simultaneously in heterogeneous modality cores. Additionally, the JOINDER's modular approach allows independent and switchable assignment of input devices to virtual object's properties. For example, as depicted in figure 39, a webcam based position tracker was implemented through the Processing environment and imported as a pure java element to the framework's interface code base. This device was used during the development

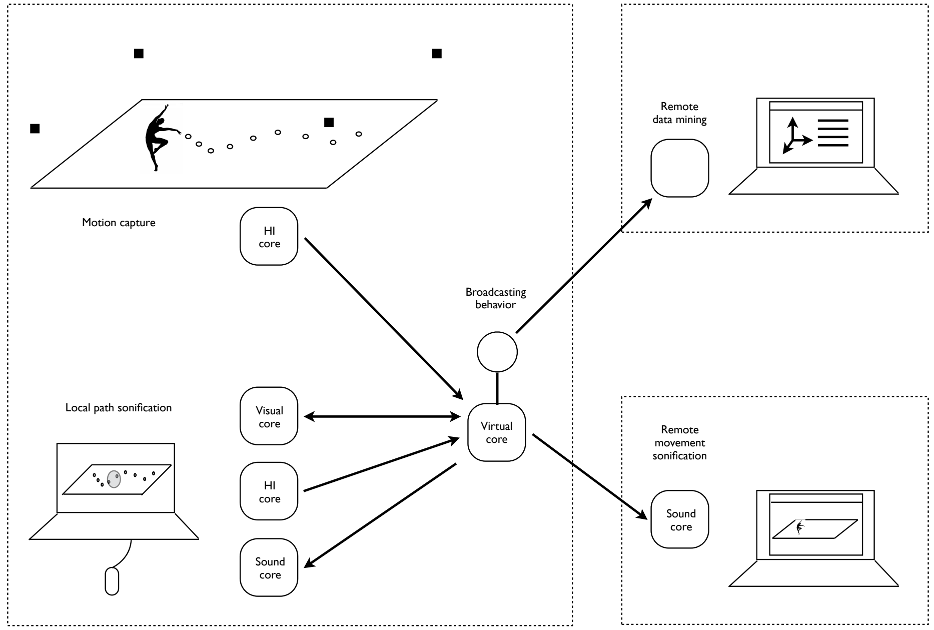


Fig. 41 Local and remote analysis scenarios via JOINDER

stages as an approximated position input, which allowed a faster prototype development cycle by not having to always recur to the motion capture system for simple testing procedures. This example also illustrates the versatility conveyed by the distributed execution strategy concerning local and remote user participation in a given deployment. One can easily imagine a scenario where users simultaneously interact at a given location using a motion capture system and remotely via laptops and more accessible input devices such as a webcam or a mouse (see Fig. 41).

Although the immersion factor is considerably diminished in the latter case, the participation of the remote users is still possible which is, in our view, an overall positive balance. Furthermore, as tracking technology gets increasingly available, one can expect that today's state of the art devices will be tomorrow's standard input paradigms. As such, the system's architecture takes into account future integration demands in order to reduce consequent adaptability efforts. Concerning output, the separation of the rendering tasks through independent cores enabled a distribution of the processing load within a given modality's output. For instance, an implementation of two visual cores allowed for differently purposed feedbacks to be simultaneously presented to the user, one rendering the latter's image in the virtual world and another providing aesthetically driven depictions. This multiple representations make it possible to furnish an interaction scenario with both interactive, latency sensitive contexts as with non-interactive, latency tolerable context.

In terms of technology, the bet in Java paid off, successfully delivering a consistence response to the performance and on-demand development needs, along with its widely proven stability, expandability and integration characteristics. Another factor for this positive outcome was an incremental approach to component integration into the core processes of the framework. The default adoption of enterprise ready subsystems could easily put an initial performance and resource strain that would be difficult to circumvent later. For instance, the inclusion of a multi-user enterprise scale system for the command broadcasting update behavior alone would represent an unnecessary computational weight in a deployment running on laptop based setup of a conventional research laboratory. Consequently, by valorizing architectural design over exhaustive features codebase, the result is an inviting, easy to understand architecture that allows a straightforward addition of functionalities.

In conclusion, both the development methodology (definition of context requirements, global requirements driven technology choice, top down context driven approach) and the implementation (user centered prototyping, bottom up component integration) have proven to be successful choices.

2.2.8 Conclusion and future work

This article presented JOINDER, a software framework for multimodal interface prototyping. Although the diversity of use cases as well as the design and methodology that were followed covered a relatively wide range of application scenarios, further expansions are ongoing. In fact, the requirements enumerated in Section 2.2.1 envisioned accommodating not only the immediate support for the research goals in the field of interactive sonification. They aim at illustrating a way in which the inclusion of new participation and interaction mediums is naturally achieved, without refactoring the current architecture and implementation. That was, after all, the reason for applying several well-known design patterns in the first place and choosing Java as the base technology. Consequently, after a first phase consolidation of the base requirements through a user centered approach and an expansion of the initial application context - from lab based interactive sonification research to public space interactive art installations -, the inclusion of new technological features that will convey a significant expansion of the use case and application scenarios is ongoing. In the near future, these might include projects in the numerous areas such as interactive advertising, game implementation and web content distribution.

The ongoing and short term feature goals for this project include the development of a GUI control panel in Processing [59], the addition of fully featured physics core with JBullet [80] as a reference implementation, a P2P network configuration as a configurable startup option (with the virtual core as an implicit instead of an explicit mediator role), runtime scripting with Python (for increasing the environment's accessibility regarding interactive control and debugging) and synchronization between multiple deployments (through the mirroring of each virtual core's nodes and a master/slave conflict resolution policy) (see Fig. 42).

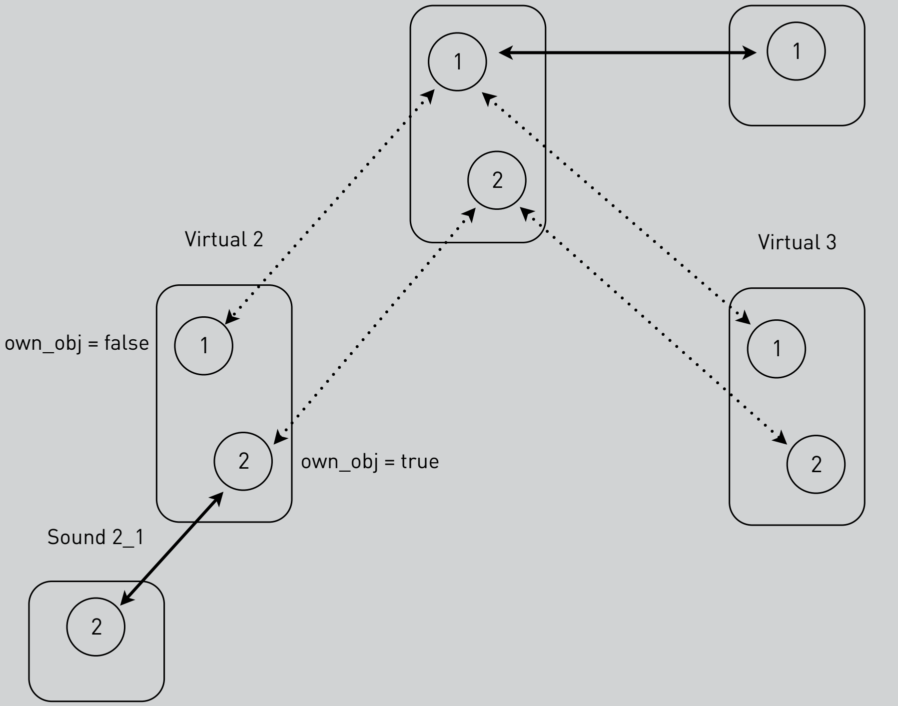


Fig. 42 A schematic representation of the synchronization algorithm for multiple JOINDER deployments.

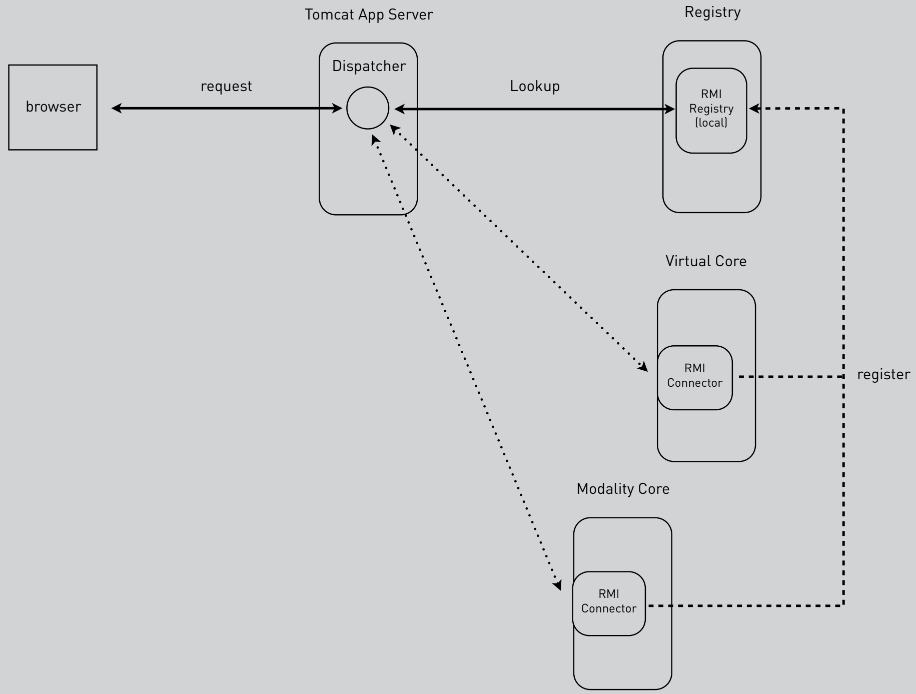


Fig. 43 A schematic representation of the RMI based communication with JOINDER's cores and elements.

As medium term goals, the design and implementation efforts will target real time capabilities improvement (through the total or partial - ex. virtual core - inclusion of RTJS implementations [143] and high performance libraries such as Javolution [79] versus a portability impact assessment), persistence storage for data mining support (through Java Persistence API and Java DataBase Connectivity) and web servicing (through RMI [162] connector's incorporation (see Fig. 43), HTTP based RESTful [53] front-end's design and implementation as an alternative to SOAP/WSDL based tightly coupled services and additional messaging facilities through OpenJMS [115]. Regarding this last item, the inclusion for web servicing technologies has the main objective of expanding the connectivity possibilities of the JOINDER framework, which rely until now on Open Sound Control and socket based transmission of standard java serialization of the command classes, towards a service oriented architecture (SOA).

Although OSC can address most of requirements postulated for this project, there are various standards lacking proper implementation in this protocol, as expressed in [58]. As such, the framework interfacing possibilities would be broadened beyond the confined space of sound and music computing field and embrace, when applicable, web communication standards. For instance, an immediate application would be the support of the mirroring updating process concerning the synchronization between multiple deployments.

Additionally, the adoption of such technologies can increase cross platform compatibility. For example, given the restrictions related to Java execution on Apple's iOS, application development based on web standards like HTML5, CSS3 and JavaScript offer a viable solution for multi-platform compatibility. Lastly, the expansion of the human interface's codebase (e.g. multi-touch and haptic devices support) will be a parallel task throughout these growth stages.

In conclusion, we are currently analyzing the technologies available in order to define the best extension strategy and related implementation roadmap for the above mentioned functionality targets. This roadmap will define how to incorporate the above-mentioned functionalities so that the JOINDER's original context of application and related performance requirements are not compromised.

Multilevel immersive interfaces for electroacoustic music composition and performance

Multilevel immersive interfaces for electroacoustic music composition and performance

Authors

Nuno Diniz - Pieter Coussement - Alexander Deweppe - Marc Leman

Abstract

This paper presents a conceptual foundation and a technological implementation of interfaces for electroacoustic music composition and performance. This concept has been developed in the context of interactive sonification. In this approach, theoretical conceptions of music composition are combined with theoretical insights of embodied music cognition, leading to interfaces that support operations such as multilevel mapping, scope variance and spatial interaction. A software framework was built for supporting a flexible wrapping and integration of different hardware and software solutions, providing a platform for the development of embodied, immersive interfaces for music performance and composition. The concepts and implementations of this framework have been tested and expanded using a participatory design within the domain of interactive arts.

Keywords

Framework - Interaction -Sonification - Electroacoustic - Sound art

2.3.1 Introduction

Non-speech sounds (such as warnings, commands, emotional utterances, musical sounds, etc.) form an important part of the way humans interact with their environment. However, not many tools are available that support non-speech sound-based communication in a convincing way. The problem mainly concerns the link between what is called "mapping" and "interaction". Mapping is concerned with how sounds are linked to meanings, whereas interaction deals with how human actions control the sounds and their meaning [72]. An integrated approach for the simultaneous development of these two poles is needed in order to provide more efficient tools that convey meaning based on intentional control of sound objects.

We believe that data sonification and music composition form two interesting domains of application in which tools, their related concepts and interactive devices can be studied. Data sonification stresses a more objective approach to listening as it aims at illustrating intrinsic structures and properties of data using sound as mediator. In contrast, music composition stresses on a more subjective approach to listening because here the goal is to reveal intrinsic properties of sounds and structures linked with particular intentions and emotions. However, both sonification and music composition have many common features. Sonification specialists may use certain sonic mappings to appeal to the users' cultural background, in an effort to stimulate the users' engagement and optimize their expertise. Likewise, composers may use all kinds of extra musical data (e.g. sounds that imitate particular devices) for their creative purposes. Consequently, by exploring alternatively both an objective side and a subjective side, non-speech sounds can be linked to specific intentions and therefore serve in a communicative context. The connection between the sounds and their intentions is called a mapping.

We believe that a technological platform can be developed for implementing these mappings, both for functional and aesthetic purposes, as applied in sonification and music composition, respectively. Such mappings can be accessed through interaction metaphors, namely through specific ways of control that cope with the cognitive and corporeal capacities of human beings. Applications in sonification and music composition provide an interesting context for the development and testing of tools for non-speech sound based communication. So far, the attention in the field has been focused mainly on mappings (e.g. auditory icons [60] and earcons [11]) but less on interaction metaphors. Yet, we think that interesting interaction metaphors have been explored in the context of electroacoustic music and that these interaction metaphors can be further developed in the context of non-verbal sonic communication.

In the presented study, we started from a musicological analysis of the compositional theories of Schaeffer, Stockhausen, Wishart and Smalley in order to identify interaction metaphors that could serve as guidelines for implementing an auditory exploration of data. As the resulting interaction metaphors constitute mediation strategies between listener and auditory artifacts, their semantic content is not limited to auditory display. Indeed, the same interaction metaphors can be tested and expanded within an artistic context, more specifically in an interactive art context.

As shown in this paper, we found that the same mediation strategies can be also used in the contemporary music production realm, as they are syntax driven and implement user centered strategies for gesture-based access to sonic entities. This application of the mediation strategies suggests the possibility for an overall validation of the developed tools, which will be based on an investigation of smalley's spectromorphological and space-form classification.

Following a state of the art of immersive interfaces for musical expression, we first present the electroacoustic interaction metaphors that guided the interface development for interactive sonification and related application results. In the subsequent part, we provide a description of an installation that aimed at exploring these metaphors within an artistic context. The installation is called SoundField, a long term artistic project undertaken to achieve an expansion and consolidation of the implemented techniques and technology. Afterwards, we present an analysis of these techniques through an in-depth spectromorphological based analysis. Finally, we address a preliminary space oriented classification and example applications of the discussed strategies to electroacoustic music composition.

2.3.2 State of the art

The use of virtual reality interfaces in the music domain has been fertile over the recent years in both performance and compositional scenarios. Ranging from software systems to components, recent developments illustrate their suitability for implementing sonic and visual interface elements through a unified conceptual strategy. In the following section, a selection of the related systems and their relevance for our proposed approach will be discussed.

Following the footsteps of pioneer works such as Enkel's SoundSpheres installation [47] and Kaper's M4CAVE [85] (system for music composition, score editing, digital sound synthesis, and sound visualization in CAVE settings), many software platforms were created for providing immersive, multimodal environments targeting musical creation.

Among others, we can highlight:

- The WAVE system [150], a multicomponent low cost environment which enables sound synthesis and sample playing through a virtual desk like interface;
- SoundStudio4D [133], a system using three dimensional drawing as a metaphor for composing spatial soundscapes in VEs;
- HarmonyGrid [1], a floor projected grid for creating and controlling a musical score;
- The GAVIP [24] /CoRSAIRe [52], project, an auditory and visual interactive platform which explores space through gestures allowing 3D navigation audio visual partition;
- DRILE [9], a virtual environment for hierarchical, interactive live looping;
- COSM [154], Networked, Max/MSP extensions for spatial electronic music within immersive audiovisual environment;
- The ToneWall/HarmonicField [142], an immersive 3D compositional environment based on volume data sonification);
- MvM [102], a computer vision oriented framework for interactive installations and dance performance.

However, although the possibilities offered by the above referenced systems are quite extensive, none of these proposals present a comprehensive response to problems of portability and scalability (e.g. COSM and GAVIP), and distributed execution (e.g. DRILE). Instead, what we are aiming at, is a more solid and open technology that incorporates the means for interconnecting heterogeneous and specialized tools. In addition, such a technology should provide a more comprehensive infrastructure for creation plurality, ensuring further development beyond the demonstration stages and elevation of the referred platform to real world application development level. A similar concern has been expressed in relation to the software components employed in the Allobrain project (which COSM is a part of), where the authors [146] state that Max/MSP is unsuitable as a foundation environment for such a long-term strategy.

Finally, some inspiring, non-immersive oriented implementations are worth mentioning here. Frameworks3D [122] introduces a three dimensional workspace for the traditional two dimensional DAW (Digital Audio Workstation) interface paradigm. While maintaining the standard track and time axis, the added depth dimension permits the hierarchical composition of sound objects. Additionally, it uses Max/MSP as a sound rendering engine to overcome the audio shortcomings of the central Java implementation. This strategy of using different environments to unite expert tools can also be found in the Versum system [5], an environment for creating 3D spatial compositions and performance. While the main coordinating process is Java based, the visual and sound rendering are delegated to specialized environments, respectively Max/MSP/Jitter and SuperCollider3.

2.3.3 Interactive sonification development

2.3.3.1 Theoretical background

We aim at generating and controlling meaningful sonic interaction by exploiting principles of gestalt theory in the context of electroacoustic composition. This context is of particular interest because of the sound manipulation principles that have been developed in order to deal with the enormous variety of sonic material. As such, one could conceive electroacoustic composition theory and practice as an attempt to make, interface wise, a given dataset easier to mold and access according to the user's inspection or monitoring goals. As previously addressed in [43] and [45], our musicological investigation was centered around the theoretical work and practice of Pierre Schaeffer and Karlheinz Stockhausen as well as on Trevor Wishart and Dennis Smalley. We believe that their compositional practices provide a structural framework for relating the components of sound based communication and embodied perception into two interrelated approaches, which are called: "multilevel mapping" and "scope variation".

Multilevel mapping concerns the consistent use of layered sonic mappings to individual or groups of samples within a given dataset in a consistent way. The Sound Object Theory [27] provided the initial basis for the implementation of this strategy. According to Schaeffer, a sound object is perceived as an object only in an enclosing context as "every object of perception is at the same time an object in so far as it is perceived as a unit locatable in a context, and a structure in so far as it is itself composed of several objects" [27, pp. 58]. This imposed dialogue condition to sound objects and structure, constitutes a fundamental design directive for allowing the manipulation of multiples levels of complexity as in, for example, [129].

Additionally, Stockhausen addressed the unified control of material (low level) and form (high level) through transpositional techniques such as integral serialism - where musical parameters (pitch, duration, dynamics, register) are derived from a series of values and are used in melodic, harmonic, and structural progression construction - and formula composition - where all aspects of a given work derive from an initial short composition [141]. As an example, his over twenty-nine hours long opera cycle "Licht" is based on a three-part, eighteen-bar only score formula. Also, Stockhausen's take on Information Theory provided a valuable input as it focuses on the behavior of sound objects through time. In Stockhausen's view, the sequence in which the auditory stimuli are presented is crucial for perceiving the musical discourse as different sequences can afford multiple views on the musical material. This interpretation provides the foundations for his "moment form", a structuring paradigm based on a non linear distribution of "gestalts" known as moments [70].

Scope variation is concerned with inspection mechanisms that allow both the triggering of generated sonification levels and an aural interpolation between them. For this matter, Wishart's considerations and practice regarding the impact of sound recording technologies on our recognition of a sound's source (e.g. the landscape of a sound) were taken into account. By changing the microphone's proximity, variations in aural perspective are created during the capturing process [160]. This technique is also used by Stockhausen, namely in Mikrophonie I, where microphones are employed for providing sonic close-ups of tam-tam sounds [141]. This manipulation process leads to new sound objects that transcend its physical sources, becoming an entity in itself. Furthermore, additional transformations can be applied concerning simulation of depth and dynamics of the sound object. Finally, Smalley's Spectromorphology classification was incorporated as it addresses the concept of gesture as wrapping mechanism for scope transposition, providing a translation device between the listener and the sonic texture. The Spectromorphology theory will be the target of further analysis in Section 2.3.5.

Multilevel mapping and scope variation are metaphors for sound manipulation. In transposing these metaphors to a technological implementation, we relied on the theory of embodied music cognition [99]. This theory dictates that the human body plays a prominent role as mediator between the (subjective) musical experience and the (objective) sonic environment. Moreover, it states that the human body can be connected with technologies in such a way that the human mind considers these technologies as part of the body. As such, the human mind can operate in a virtual sonic environment, using extended body technologies. We believe that, through the expansion of the mediating role of the body with an immersive 3D environment, it is possible to turn the interaction metaphors of electroacoustic music into an interesting technological implementation for exploring sonification. By using physical space as a common representation of the real and virtual world, and as an access strategy for sonic objects, it becomes possible to explore direct manipulation [34] strategies that bring forward the user's intention through an explicit manipulation of the sound objects as they are afforded by the space and spatial setting. Additionally, an object oriented paradigm was chosen in order to better take advantage of structural similarities between the variable resolution of the human motor system [63], the sound objects and data. Based on the above considerations, two main categories of interfacing metaphors were designed, namely, the virtual inspection window and virtual inspection tool.

- The **virtual inspection window** presents a strategy for representing a variable's values (or a subset of them) in a three dimensional environment. If the variable is time dependent, it can be used as a timeframe representation, in which an array of virtual objects follows a First In First Out management (w(t) = [v(t), ... , v(t + !t*size(w))]]). This strategy significantly increases the representational versatility of a given variable. For instance, the same variable can be represented by multiple windows (with the same or distinct time intervals), multiple window's morphology (linear, circular, ...) and multiple window's sizes to better fit the user's inspection procedures.
- The **virtual inspection tool** allows the activation of the constituents of the virtual inspection window. Following a "virtual microphone" strategy, as suggested by the work of Stockhausen and Wishart (see above), such an activation can occur through collision detection or a spatial distance threshold approach. The spatial distance between the user and sound object becomes a tool for foreground/background interpolation of the data. as well as for individual/group comparative analysis of data samples. The virtual inspection tool can be used with multiple instances of microphones. For example, it is possible to apply different microphones on multiple points of the user's body, and each microphone can be paired with one or multiple mapping levels. Additionally, a global (all microphones can activate all windows) or selective (microphone/window pairing) activation strategy can be explored.

A summary of the relationship between the presented electroacoustic techniques, communication approaches and interfacing metaphors is illustrated in Fig. 44. Through these metaphors, multiple possibilities for exploration and monitoring perspectives are conveyed to the user. The application of the described multilevel mapping and scope variation strategies will be the focus of the following section.

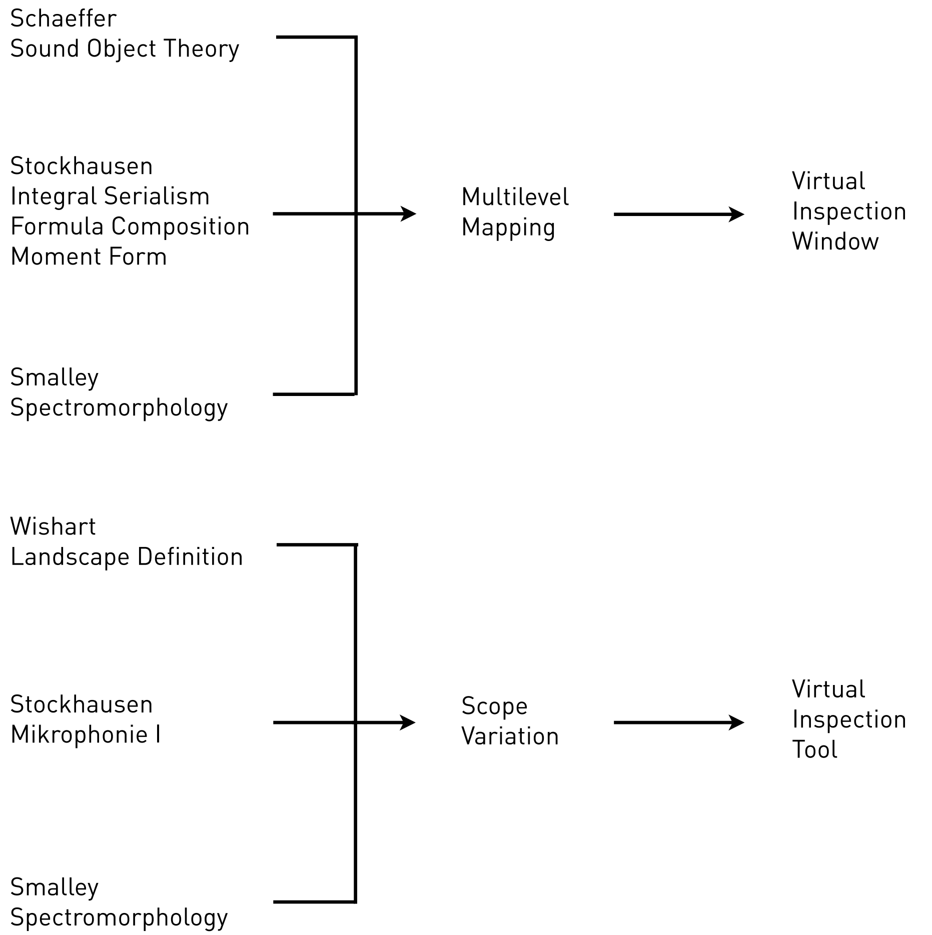


Fig. 44 Relationship between the presented electroacoustic techniques, communication approaches and interfacing metaphors.

2.3.3.2 Sonification use cases

A sonification system was developed, targeting the interactive exploration of multivariable data through non-speech audio communication [45]. The goal of this implementation was to investigate the possibility of establishing a unified context between individual sound streams, simultaneously exposed through time. The sonification paradigm was based on concepts from electroacoustic musical composition. The inspection process was conducted through the user's interaction with virtual objects in an immersive 3D environment (see Fig. 45). Through several use cases, we investigated the user's relationship with spatial sound object representation and the use of different sonification levels upon his exploration of a test dataset.
The conceptual design was based on the combination of two main metaphors, namely, the virtual inspection window, providing access points to the variables and their values belonging to a given dataset, and the virtual microphone, which implemented a sonic inspection tool controlled by the user's hand. This approach is inspired on Stockhausen's work Mikrophonie I (see previous section). In the first use case, each independent virtual object belonging to the virtual inspection window functions as a sound source. The pitch representation of its correspondent data value was activated through collision detection when the inspection volume of the virtual microphone intersected the virtual objects.

Moreover, two additional sonic representations of the relationship between the values of the data were made available, namely, interval and chord, in order to provide a sonic feedback of numeric proportions in the data. In addition, the position of the activated objects in the inspection volume of the virtual microphone influenced the respective output in terms of auditory relevance. This was implemented through distance based real-time adjustment of amplitude and reverberation, as suggested in [160] concerning distance simulation. In summary, the

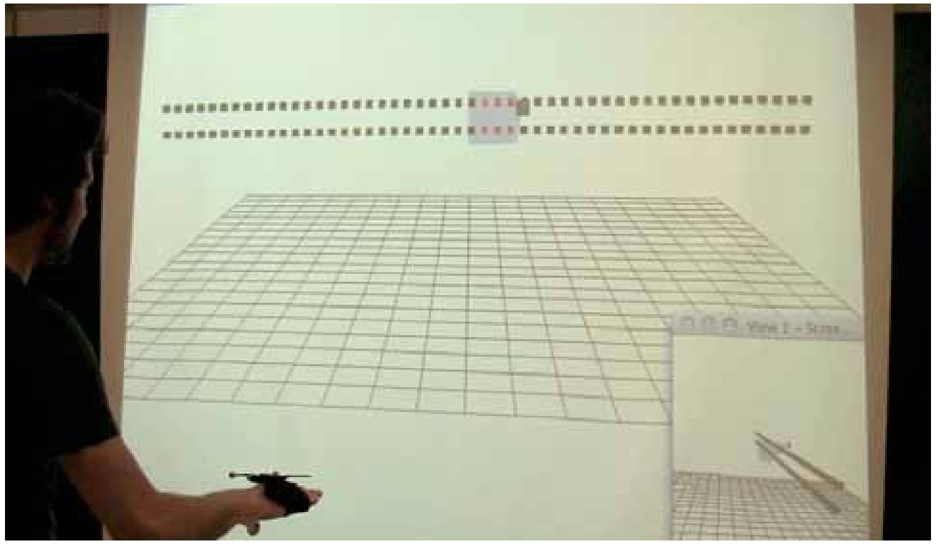


Fig. 45 User interactively exploring a dataset, using two virtual inspection windows (e.g. array of cubes) and one virtual inspection tool (e.g. pyramid shaped object).

adopted interface paradigms conveyed multiple perspective views between different levels by representing the evolution of the sound object/data in time and structure, and by establishing and comparing different groupings of variables.

The second use case, called the Dance use case [43], was based on the previously described prototype. This time, the array is scattered in space, enabling users to vary their inspection in space and time even more (see Fig. 46). An extra bend sensor and wireless ADC allowed for the participants to vary the scope of the virtual microphone by changing their posture from an open, global to a closed, detailed viewpoint, which tied the use of space in relation to a body-centered spatial exploration over time.

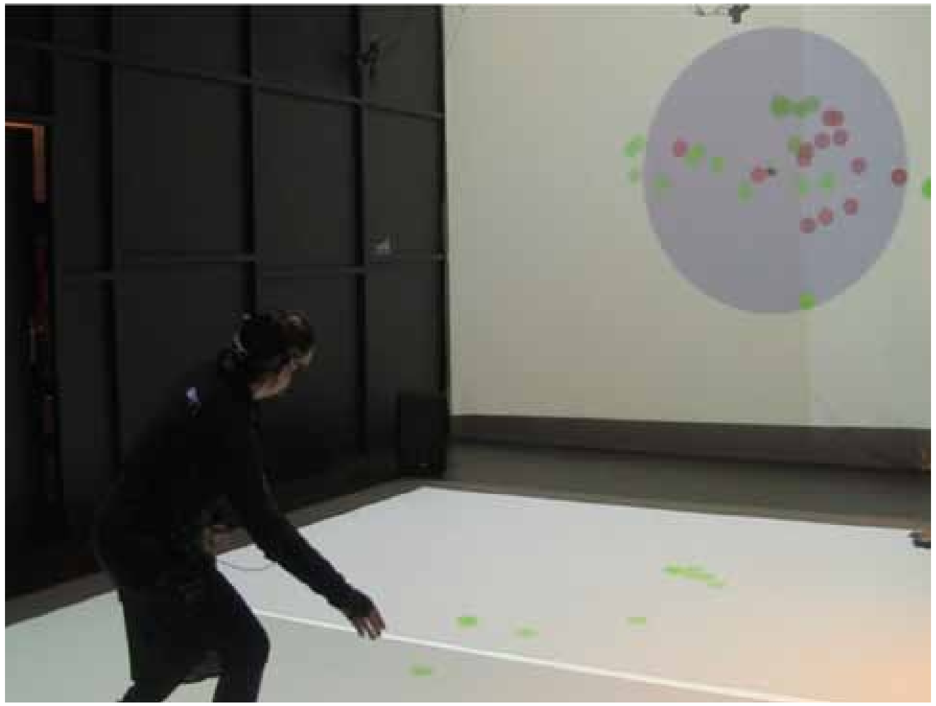


Fig. 46 User investigating a choreography path in the Dance use case as part of a dance workshop at Destelheide educational center in Belgium.

The Dance use case consisted of two parts, namely, a setup part and an exploration part. In the setup part, the data objects were set out in space by a user. Accordingly, the data objects represent dance movements in the sense that a sequence of these data objects illustrates a trace of the dance movements as an occupation of space. In the exploration part, the user is given the task to explore the trace of the initial movement. The exploring user had no visual representation of when the virtual objects creation took place within the choreography. Besides indicating the direction of the path through an increase in pitch, variation in the velocity of the original movement was conveyed by the variation of the pitch's increase. This strategy was strongly inspired on Stockhausen's take on Information Theory (see Section 2.3.3.1).

2.3.3.3 JOINDER, a software framework for interactive sonification

The developed software, applied in the above use cases, resulted from the need for generic data sonification tools for both research and applications [94]. The software aims at providing an infrastructure for integrating the concepts discussed above while addressing requirements related to auditory display, namely, portability, flexibility and integrability. For that aim, a Java framework has been developed, based on a functional division of the multimodal interaction realm into individual branches around a state representation (see Fig. 47). In this framework, the concrete implementation of a virtual world, its visual and auditory representations, and the human interfaces, can be defined according to the desired performance, access or functional needs of the intended use cases, using external libraries and platforms, such as Java3D, Supercollider 3, Max/MSP or Ableton Live (see Fig. 48).

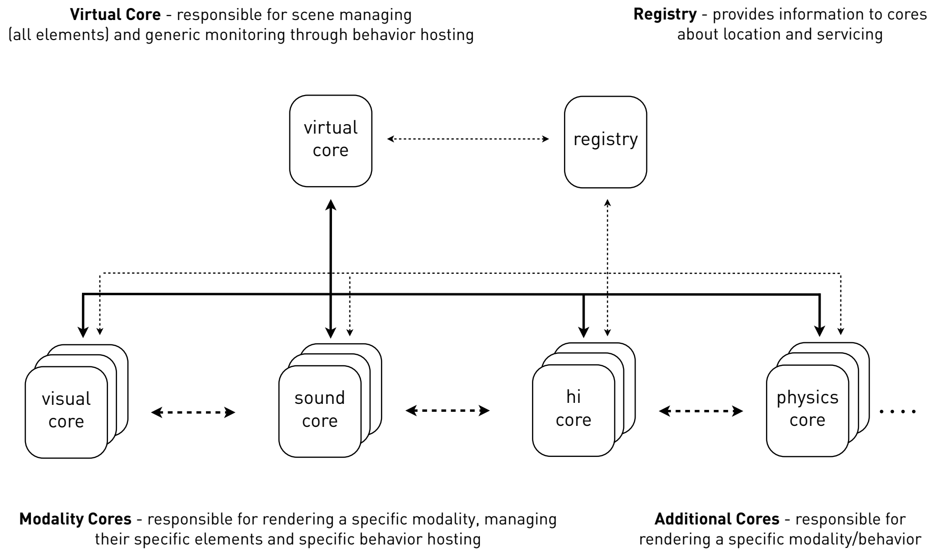


Fig. 47 Schematic representation of the developed system. The virtual core serves as a global hub for coordinating the elements rendered by the modality cores (visual, audio, human interface) and other additional cores (e.g. physics).

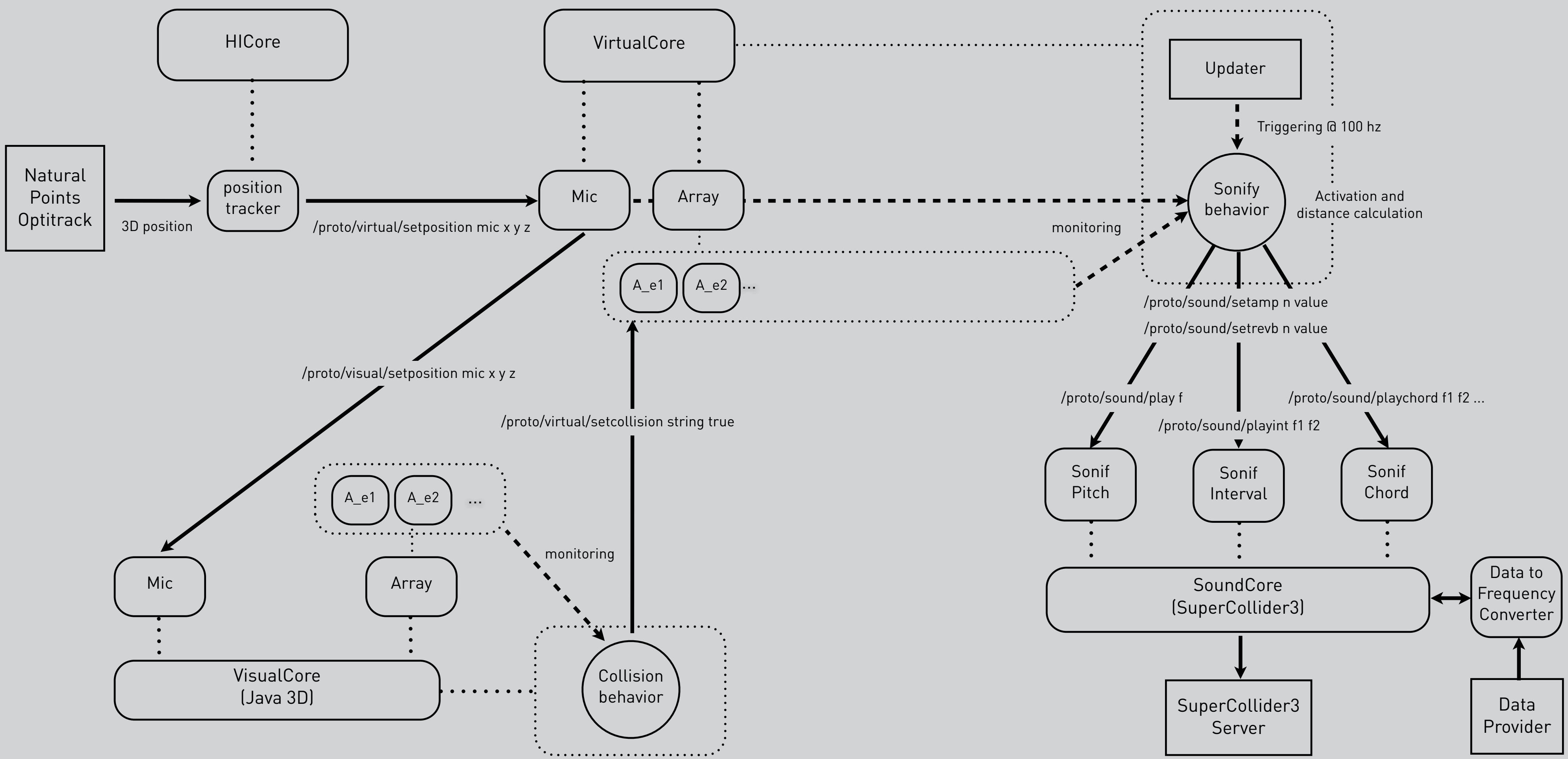


Fig. 48 Data flow between the system's different cores and tasks implemented for the first sonification use case.

2.3.4 Expansion of the system and strategies through its application within an artistic context: the SoundField project

Given the limitations of a laboratory setup, where the first interactive signification use case took place, a need arose for a broader, more ecological-based design. This would allow the implementation of features through a participatory design strategy [39]. In this approach, software development is done in collaboration with users, which typically leads to a focused and comprehensive bottom up creation of modules that complement the top down premisses concerning interaction metaphors, mediation paradigms, and other preliminary decisions (such as choice of different media technologies).

As a consequence, the SoundField project was conducted as a work in progress, artistic installation, involving a broad audience, who participated in the development, from October 2010 to May 2011 at the Destelheide Educational Center in Dworp, Belgium. This event was organized in close collaboration with the Flanders "Youth and Music" program. The focus of the project was mainly on the exploration of interactivity in new media arts through social interaction principles [30], extending the above mentioned virtual objects mediation approach. During the referred period, several implementations were developed according to the users' requirements, observations and evaluations, which were gathered during evaluation sessions. The requirements were incorporated (through rapid prototyping) and presented to the same user groups. After another set of tests and interaction sessions, the specifications of the given use case, and the experience of the users with the technology was recorded, in order to incorporate this knowledge in the following use cases.

In summary, this project combined new media art studies, interactive software design and participatory development. The Destelheide Educational Center provides workshops and courses in various disciplines such as dance, literature, music, theater, sculpture and audiovisual arts, throughout the year.

2.3.4.1 Expansion of the technological setup

The JOINDER framework was expanded in order to accommodate the requirements of the SoundField project. The following description presents an overview of the base setup used during this project (see Fig. 49). The Java-based framework constituted the core of the system concerning virtual scene state representation and monitoring. An Optitrack motion capture system was employed to capture the 3D position and orientation of IR reflective marker sets, which are typically attached to the human body. The data was transmitted through the NatNet protocol via the Arena software and converted to OSC by a custom driver. An ICubeX system was used for connecting various sensors (e.g. a bend sensor) in order to control the radius of the virtual inspection tool. The data is received

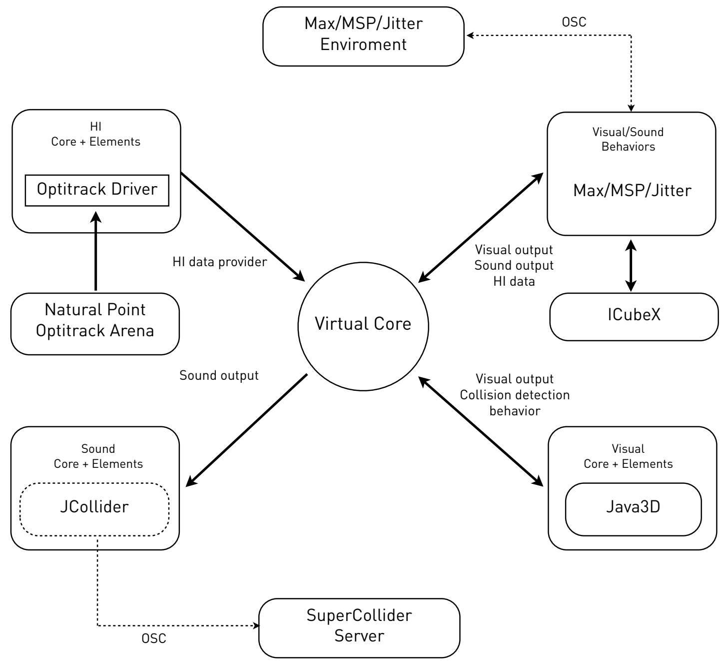


Fig. 49 An overview of the interconnection between the software and the hardware components used in the SoundField project.

in a Max/MSP patch. All the referred OSC formatted information is gathered in the virtual core and used to place and configure the virtual objects in the scene. Based on this information, collision detection and additional information (for example, the distance between virtual inspection tool and activated sound object) is calculated and sent in the OSC format to, for instance, a Max/MSP patch. The Max/MSP (together with Ableton Live through MaxforLive) and SuperCollider3 environments are responsible for the sound synthesis and/or spatialization based on the data transmitted by the Java framework. Java 3D and Max/MSP/Jitter based implementations are used for the rendering of the scene's visual projections. A summary of these components is presented in Fig. 49.

2.3.4.2 SoundField use cases

The use cases developed during the SoundField sessions encompassed various interaction strategies through multiple implementation scenarios. A prototype demonstration deployment, which expanded the interaction possibilities presented in the previous section, constituted the starting point for the development of the subsequent use cases. In this prototype demonstration, the envisioned social dimensionality, as a means of establishing cooperative interaction, was implemented through simple metaphors - point, line and plane (see Fig. 50). From there, 14 distinct deployments and additional variations were developed in close collaboration with their respective user groups. These use cases ranged from 3D sound manipulation to theatrical audiovisual augmentation and multimodal rehabilitation tasks. A total of 77 users actively participated in the full evaluation process.

Given the wide range of use cases developed during the SoundField project, the following description will be limited to the most relevant examples that are pertinent to this article's scope. A more in-depth presentation and analysis of the SoundField project in terms of user participatory development is under [40].

Fig. 50 Users interacting with the SoundField's demonstrator.

For our current purposes, these use cases constitute an extension of the methodology and interface design strategies that were presented in the above sonification use cases, described in section 2.3.3.2.

The SoundPath use case was implemented within a dance workshop environment. The interface consisted of an array of virtual objects that were positioned in the 3D environment by a dancer while performing a free form dance routine. This array of objects illustrated the path of the performance and held the

Fig. 51 Users performing within the SoundPath use case. The first user generated a path which was in turn explored by the second user.

information concerning the original movement, which was followed and accessed by another performer. It presents a variation of the Dance use case described earlier on. The main difference consists in the fact that the generation and exploration are condensed into one single stage, both occurring in real time. In addition, the trail would be limited to a predetermined number of objects, in order to avoid overpopulating the virtual scene. In practice, the SoundPath configuration was used in dance choreographies where contrapuntal movements between the two dancers are emphasized (See Fig. 51). The sonic output would be the result, the reinterpretation or answer to the first dancer's movement through the second dancer's activation of the performance's path. Also, the articulation between the use of personal and scene space, either in the creational or inspection tasks, created enough room for diversity of execution within this symbiotic setting.

Another variation of the interfaces previously used in the above sonification scenarios was SoundMaze and SoundGrid. These use cases consisted primarily in the recreation of a sequence of sounds (e.g. a melody) through the activation of virtual blocks, placed in a maze/grid-like way. These use cases originally targeted a rehabilitation workshop for patients with mental disabilities, and were valuable in evaluating the interaction paradigm within a constrained access. The results led us to speculate on other applications where the activation objects would have a dynamic form and placement according to the temporal enrolling of a given performance.

The SoundObjects use case was based on the initial idea of SoundField, namely, to create an environment where collaboration of participants is a fundamental condition to unveil new form of interactions and sonic output. In this use case, each user can manipulate a virtual spherical object in the three dimensional space. Using the position of the objects in relation to the other objects, it is possible to configure the morphology and sonic response of two additional objects attached to the spheres, namely a 2-point line and a 3-point plane.

Fig. 52 The SoundMovie use case. Using absolute position, participants controlled several attributes of their movie while an auxiliary object would determine their participation in the final mix (e.g. the wall projection) through relative distance.

The sonification parameters of the virtual objects mimicked the tuning parameters of a real string (2-point line) or a real membrane (3-point plane) and their respective central pitch was based on, respectively, the line length or plane area. A user can then trigger the objects in real time through collision detection. Other use cases mainly explored the three dimensional nature of the interface space through the placing and movement of the user-controlled virtual objects. For example, SoundSculpture used a three dimensional, absolute position to control the sonic parameters. SoundStage, SoundScene and SoundTales used relative distance between user controlled and static interface objects to manipulate sonic landscapes. Similarly, the SoundMovie (See Fig. 52) and SoundTheater variants implemented audio/video selection and mixing to support scenic performances.

2.3.4.3 Results

The users' participation was organized in different in different stages, which followed the same development pattern in every use case. After the presentation of selected demonstrators, a pre-interaction meeting was conducted in order to access the requirements in relation to the workshop's goals. This would be succeeded by an implementation of the gathered demands and the respective interaction session. Afterwards, a post-interaction evaluation collected feedback information concerning this session's results through questionnaires, group interview and debate.

This series of iterations aimed at gathering new usages of the mediation paradigm through an assessment of the user's preliminary evaluation concerning early stage prototypes that implemented such new approaches. Technologically speaking, the main idea was to test the adaptability of the framework's structure to rapid prototype development challenges. The results were to be used in incrementing the software code base (e.g. new device drivers) of the JOINDER framework, as well as in circumscribing possible refactorization areas (e.g. class structure, thread management, etc.). Additionally, this user participation strategy envisioned to collect new applications of the interface paradigms offered by the system. These applications were proposed by the prospective users after a brief first contact with -often- no previous experience concerning multimodal/immersive technologies. Furthermore, these prototypes often resulted from one day's implementation effort. Consequently, they frequently contained faults, due to non-thorough testing, which at times influenced the aesthetics, accuracy and performance evaluation. Nevertheless, our main drive was to evaluate the suitability and viability of these new interaction modes proposed within an artistic context into the interface's (and underlying software framework) range of possibilities. As such, the participation of the users and their level of satisfaction contributed to valuable insights concerning the needs and the fulfillment of expectations with regards to the used interface strategies.

Concerning the interface paradigm, performance and requirements fulfillment, several tendencies were observed. Throughout the various use cases, the users' apprehension of the interaction paradigms was swift, as the visual and sonic components of the interface were usually considered of a self-explanatory nature. As the virtual objects manipulation and behaviors triggering/monitoring persisted as a base paradigm across the demo and prototype testing phases, the user found that interface affordances could be discovered without problems. This allowed the exploration of the use cases' specific features in a comfortable manner. Consequently, requirements and implementation were related satisfactorily and the majority of the users rated the developed prototype positively, within the prototype stage boundaries.

As a result, the configuration possibilities of the interface were extended as the participants of the workshops assimilated its underlying paradigms in a fairly straightforward manner. Within this scope, the JOINDER software framework functioned as foreseen, providing a scalable progression between single and multiple users environments as well as between simple and complex behavior handling. The framework's modular design and scalable technology provided an adequate platform for the requirements with regards to the expansibility of the project. Both new software environments (e.g. Max/MsP) as hardware devices (e.g. iCubeX bend sensor, WiiMote) were incrementally incorporated without any repercussion to the already used technologies, as well as functions and data management facilities already available in the JOINDER framework. This made it possible to make iterative transitions between setups and to consolidate design choices.However, the SoundField iterations did lead to some concerns regarding the interface concept's application. First, limitations regarding the number of participants in social interaction scenarios were pinpointed during the evaluation of use cases with more than three participants due to technological constrains of the motion capture system. More specifically concerning the trackable area and usable number of rigid bodies. Additionally, when several users have the same virtual objects manipulation mapping to sound, the discernibility component of the interface is reduced and interaction becomes confusing. When there are more then the three critical participants, the outcome of a user's own actions would often become cluttered by the global scene sonic output. As such, the contribution of the visual output has proven to be, in many cases, as important in terms of discoverability [111] as the sonic feedback. One can attribute this to two facts. First, the visual layer constitutes the first contact with the interface (as in the only perceptual output immediately before the interaction). Second, the sonic output was a consequence of the user's manipulation (given that the artistic research within this interactive art research is centered on the emergence of the artistic content only upon interaction [30]). Also, in terms of accuracy and suitability [111], the reliance of the interface on an adequate representation of the virtual object's real space occupation was highlighted as an important factor to account for (being essential to, for example, usable activation through collision detection).

In summary, these observations provide both a preliminary confirmation that the developed interactive sonification metaphors (and related software) can be used in artistic contexts. Consequently, these metaphors can provide useful guidelines for the development of music production immersive scenarios.

2.3.5 Validation of the developed sonification tools

In this section, the tools developed in the previous stages will be compared to Smalley's compositional theory of Spectromorphology, Spatiomorphology [134] and Space-form framework [137]. This comparison aims at accessing the possible use of new interactive multimodal interfaces, such as the ones that JOINDER framework provides, in the domain of electroacoustic music composition and performance.

2.3.5.1 Spectromorphology

According to Smalley, the sounding body is activated through energy-motion trajectories, which conveys its spectromorphological identification. This principle is based on a gestalt-theoretical insight that is also relevant to our framework for interactive sonification. For example, while interactively investigating a dataset, the user's intentionality may be transmitted through gestures that provide energy-motion trajectories. By describing the user's exploration of data, these trajectories may contain information that can be used to fine-tune the way in which the user explores his goals or reveals unexpected phenomenon in the data.

However, this approach assumes an interface that can represent the data on a "human scale", that is, a scale reachable by the user. In addition, we have to acknowledge that an electroacoustic gesture is more than the sum of its components. As in the tonal tradition, the use of chord as an identity mark (e.g. Tristan chord) goes beyond the function that is dictated by its constituent pitches, as the chord may serve as a pointer or trigger for an association that is external to the chord. In sonification, exploring a dataset can be a very difficult task by merely supplying auditory information about every sample. Therefore, the sonification process should provide hierarchical levels of feedback and tools for their interpolation. During the inspection stage, interactive control should be made available to allow the switch between the different levels of sonification in a dynamic way and to allow the user to establish relations between the dataset's elements (e.g. spatial selection/activation of elements, sample grouping, etc.). It is this combination of gestural framing (actions upon the data) and sonic texture setting (scope of the actions) that can pave the way for a spectromorphological analysis of a dataset through an interactive and embodied sonic exploration.

As stated above, we believe that the concept of gestural framing, and its related sonic motion, constitutes a meaningful concept for interactive sonification. Gestural framing entails a multidimensional spectral occupancy that, in agreement with Schaeffer's dualistic classification of the sound object and its enclosing structure [27], is composed by multiple layers or levels. The transposition of Smalley's motion classification (unidireccional, cyclic, etc.) [134] to the interactive sonification domain is supported by the virtual object's morphology and mapping configuration. Our explorations in different use cases suggest that an embodied use of these morphologies and mappings can facilitate the discovery of group patterns within the dataset's samples. For instance, the configuration of the virtual inspection window enables iterative, granular or sustained texture motions by configuring the amount of samples or selecting different types of virtual object (e.g. array or a line), using the own body as point of reference.

Through this framework, embodied involvement can be supported through an interactive scope activation (via multiple virtual inspection tools) that offers a spatial control strategy. As such, the user can control mappings in a hierarchical way, for example, from body center (larger scope) to arm (medium scope) to hand (small scope). In addition, these different scopes are typically complemented by the assignment of hierarchical sonification levels. When in use, the user's active listening establishes a dialogue where expectations concerning the behavior of the data are reflected (or not) in the sonic result of the exploratory movement and adapted through a continuous adjustment of the gesture. The detection of meaningful data can be related to the presence of a "musical" function. However, since the presence of such "musical" functions is often undefined, interactivity offers a tool for disambiguation, applicable wherever it may be found to be relevant.

The basic stages of a gesture (onset, continuant, termination) are determined here by the user's action, which controls both the activation and ceasing of the sonic activation. In addition, the relative position of the user within the interactive interface framework, typically modulates the level of participation in the total sonic output, creating the spectromorphological body of the elements that are explored in the interaction. By shifting the activation pattern, the motion and growth characteristics of the dataset (e.g. the temporal evolution of ambient temperature data in a given location) can be exposed. Both motion and growth patterns convey directional information, from which structural functions (e.g. sonic markers) can be attributed regarding expectation. As such, the dynamic scopes of the interface aim at maximizing the user's objective, by providing multiple perspective views on the evolution of the dataset. For example, in the Dance use case, not only directionality, but also the rate of change was conveyed to the user, which created the grounds for a task execution that was based on musical expectation. It illustrates how the application of Smalley's classification concerning directionality and density of sound sources were used for guiding the user during his task.

In summary, by moving away from the temperate scale-based granularity and the fixed timbre classification scheme (e.g. instruments) that are typically used in sonification projects, it seems that the spectromorphological framework encompasses a fuller scope of auditory display possibilities. Such a framework considers a spectral content that is defined by the gestural behavior of users. Its temporal evolution provides the basis for an identity assessment of the data in relation to gestural control and sonic feedback.

In this approach to gestural framing, motion and growth are used as a reference for the implementation of interactive sonification tools that interconnect hierarchical sonic levels to interactive, dynamic scope tools. They established a methodology that is rooted in embodied interactions (that use the own body as reference), which can enable the interpretation of discrete, low-level elements, through higher-level relations. As such, individual samples of a dataset can be grouped through configurable scope adjustment, allowing perceptual interpolation. Having clear analogies with the physical inspection of real world objects, the virtual microphone metaphor affords a correspondence with real world exploration. One can easily compare the browsing through an array of virtual objects representing a dataset with hitting or scraping the surface of a set of objects.

Concerning the two components of the spectromorphological framework, namely the spectral and the morphological component, we can say that in our approach, the spectral component is portrayed by the multilevel mapping whereas the morphological component is rendered through the gesture-based interaction. Both are interconnected through the virtual object mediation and the variable scope strategies.

In terms of the spectral domain, the different classifications of source type (note, note collectives, and noise) are incorporated into the sonification strategy. Here, the main component is scope variation. This mechanism allows the simultaneous activation of different samples. This perceptual interpolation, along with the sonification levels, conveys a variation of granularity of the sound sources that allows the user to, for example, hear large chunks of data (directional, colored noise) and zooming in to a specific group (note collectives) or single value (note) when alerted to the presence of a given numerical relation (ex. interval) between the elements of a dataset.

Additionally, the higher levels of sonification reveal a distinct spectral signature that can help the user to pinpoint a region of interest within the dataset. For that purpose, the implemented alerts in the above described dataset sonification are based on pitch, as intervalic relationships are deeply embedded in our cultural background. As stated by Smalley, "Even with a pitch-drone, the actual pitch (as opposed to the fact that it is high, low, or a stable reference-point, or its motion is textured) is unimportant if intervallic pitch relationships are absent" [134, pp. 14]. By confronting the highlight of external relations (abstract language) [48] with the relations contained on the dataset (abstracted language), this affords a multilevel inspection by the user. Furthermore, relative anchoring plays a fundamental role in assessing the identity of the sonic artifact and thus the represented dataset.

Referred by Smalley as "rooting", sonic markers highlight the variation in the dataset by establishing a spectral framing upon which one can infer the behavior of the analyzed data. As such, these elements may be used for comparison purposes. By analyzing different points of a dataset, through multiple inspections tools, the user establishes a sonic ground from which the relevance of the presented data can be inferred by comparison. For example, one window with the average daily temperatures of a given month may provide auditory rooting to analyze the variance concerning the hourly temperatures of a given day. These sonic markers help to foreground dataset behaviors in a similar way that the presence of key sound objects can be used to establish musical functions within the discourse of a work.

Concerning the spectral space and density (e.g. the rate and quality of the spectrum's occupancy), both the concepts of sound object perception (as a result of sonic description through time) and gesture (as an user's action) are closely linked within the dataset representation strategy and interactive scope activation. These attributes are incorporated as descriptors of the dataset's sonic identity and behavior. The first is related with the already discussed motion and growth types. Density is transposed as a guideline for a detailed assessment while overviewing a dataset. Again, the higher levels of sonification function as an alert for searched relations and as a marker for exploration guiding, as they pierce through the high density feedback of wide scope activation. Furthermore, density parsing is also related with spatial placing of the dataset's virtual objects. This allows the user to step back and forth and to reposition the listening focus within multiple perspectives. As stated by Smalley, "Spectral density is related to distance perspective" and "same density set further back will free space closer to the listener so that it can be occupied by other spectromorphologies" [134. pp. 15].

2.3.5.2 Spatiomorphology and Space-form

This section addresses the concept of Spatiomorphology [134] and Space-form [48], and its correspondence to the developed interactive sonification model. Spatiomorphology deals with the categorization of changes in the spectral space, used by a sound object in a given environment. Since this spectral occupation is time dependent, it is intimately related to density and dynamic levels as well as motion and growth processes. In addition, spatiomorphology also addresses the listening space in which a work is physically diffused. In this sense, "spectromorphology becomes the medium through which space can be explored and experienced" [134, pp. 16].

In our sonification framework, this definition is applied by making the virtual world a physical space of interaction. As such, multiple perspectives of the sound space are afforded to the user by the relative position and distance of the user's body to the virtual dataset in space. For example, in the Dance use case, the real space becomes populated with virtual elements, resulting from the dance movements. This space is then transformed into a composed space and accessed by the user through spatial exploration and user-centric scope variation. As the composed and listening space become one, the user investigates the listening space by direct interaction with the composed space. The affordance of multiple perspectives is a base component of sonic perception and cognition as "listeners can only become really aware of the variants if they have had an opportunity to compare perceptions of the same work under different listening conditions" [134, pp. 16]. The Dance use case is an example of a trail that first transforms the dancer's gesture into a spatial trajectory by conveying directionality, energy and velocity information about the sound object. The inspection process then allows the user to apply further concepts such as image definition in which scope variation becomes a means to reduce the listening space to a certain region and alternate between different spatial textures of the dataset.

The acknowledgement of space as a primary component for stimuli integration and structure analysis is central in Smalley's classification framework. In the space-form theory [137], the spectral content-driven classification is extended as the composed and listening space are given a center role in the acousmatic interpretation and production framework and are subject to a more thorough investigation. One of the key considerations is that time can be subservient of space. As such, immersive virtual reality and real world geometry interfaces can be of service here. Given the fact that music is a time dependent construction, the interface action space can serve as a placeholder for time-stamped samples of one given dataset. Time is, therefore, represented in space and the creation of source-bonded spaces is based on the user's direct manipulation of the entities at play. The application of this theoretical viewpoint can be seen in the interactive sonification use cases, where the co-inhabitance of the same space by user and virtual objects allowed the user to explore an interface, based on space as a time-browsing device. This configurable representation, founded on placement of virtual objects within the interaction space, conveys a suitable environment for transmodal perception and actuation. As mentioned previously, the sound's spectral evolution, as well as its motion and placement within a given space, is deeply related with the physical act of sound making in both natural/cultural environments. Visual and sonic phenomena are intimately connected with our proprioceptive capabilities and gesture making. This puts the human agent at the core of the perception activity, where meaning inferences direct us to acknowledge any gestural information that a sound might suggest as we virtually mimic the necessary conditions and actions to reproduce the received auditory event. In this sense, the "materialization" of sonic elements through spatial occupancy combined with human centered scope interaction merge virtual/real world elements and behaviors as equally important components of the same process.

Acknowledging the crucial role of embodiment and cross-modality in the assimilation of sonic stimuli, Smalley defines space-form as a framework for classifying multimodal stimuli through the spectromorphological content of sound sources and their evolution's relationships. This classification borrows a proxemic spatial definition from Hall [68], which encompasses four human centric volumes, namely: intimate, personal, social and public human centric volumes In addition, it encompasses the idea that our perception is based on a dynamic shifting of these ranges. Through expansion and contraction, the inner, proprioceptive scope of the individual is contrasted with the social arena.

This process allows a conversion of the sonic landscape to our human scale. Again, by magnifying spectral content and gestural behavior, the microphone space is referenced as a primary tool for interconnecting the visual and aural elements. Our application of a virtual microphone metaphor with a variable scope range puts this facility at the user's command. Here, the concept of scale is fundamental in the creation of changes in perspective, such as overture versus enclosure, within an immersive space.

This feature, along with positioning, affords the definition of multiple vantage points and this perspectival articulation allows for the listening user to establish an interconnection between the attentional focus (prospective space) and the surrounding aural environment (circumspace). Consequently, the incorporation of proper technological tools facilitates the exploratory task by affording the accumulation of information from multiple viewpoints, resolutions and scales to be processed at a high level of perception. As such, it facilitates the extraction of the visual, sound and spatial elements' identity and, consequently, their comparison.
In this approach, spectral content is regarded as a provider of perspective and movement. Through resolution, quality and dynamic level manipulation, proximal and distal spaces can be connected in order to relate position, movement and behavior of sonic sources. According to Smalley, these attributes are assigned to a note/noise continuum and are bounded to our pitch based musical cultural basis. Nevertheless, they represent a point of departure for a more global analysis, which comprises concepts such as gravitation (an extension of the cadence compositional device) through the representation of attraction poles and directionality (e.g. glissandi) and the creation of spectral lines and planes. Often used to represent musical tension and relief, the concepts can be helpful in highlighting events and behaviors of interest that direct the listening user's attention.

Generically speaking, the application of these methods are useful in generating interface objects that co-inhabit the spectral, virtual and physical space. Also, the concept of circumspace entails the possibility of multiple/simultaneous subspace creation. Combined with processes related to positioning and scope variation, they allow a more versatile "polyphonic" assessment of the elements at play. As stated by Smalley, circumspectral spaces enable "greater flexibility in the use of perspectival shifts to create spatial forms that seem to turn in three dimensions" [137, pp. 17].

All of the above mentioned strategies contribute to a dynamic selection of vantage points during user's analysis, operation or performance. Furthermore, Smalley highlights immersive spaces as a type of circumspace that appeals to such type of inspection as "the listener gains from adopting, and is encouraged to adopt, different vantage points (...) freeing aural elements from continuous, concentrated scrutiny" [137, pp. 18]. Our proposed methodology of interaction, based on virtual objects mediation and immersive exploration, conforms with this view. It offers the user the necessary freedom and means to better apply his expertise and intentions. Through active exploration, "acousmatic space-form is centered on us, not only as receivers and perceivers, but also as producers and inhabitants of space" [137, pp. 21].

Within this framework, the above-mentioned concepts have been incorporated into our interaction design through real space geometry virtual objects (which enables the use and configuration of geometric planes/string, as in the SoundObjects use case), through multi-level sound mapping of virtual objects (by allowing the sonic highlighting of points of interest) and variable scope definition (in facilitating the activation selection of the scene elements via mechanisms as proximate vectorial wipe and distal interpolation).

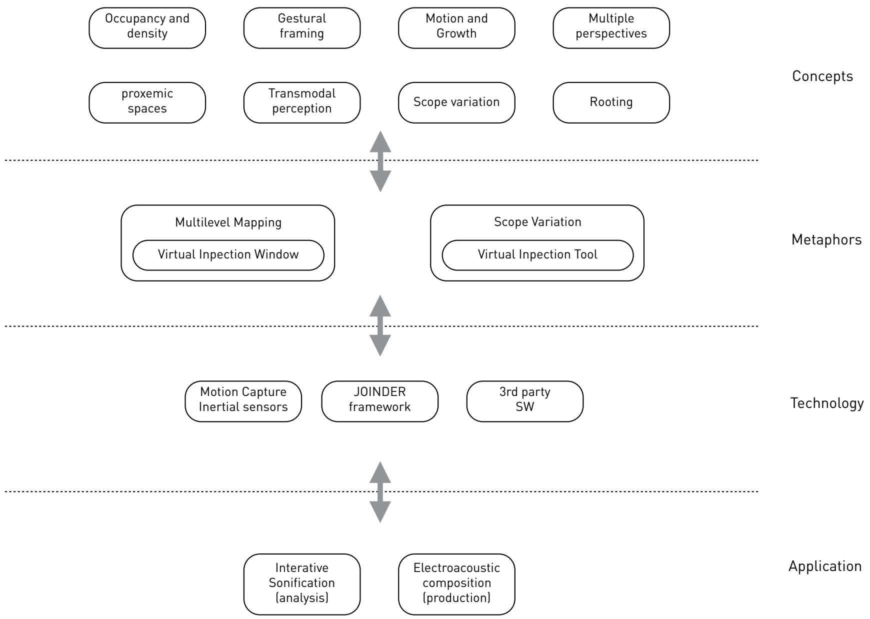


Fig. 53 Summary of the Spectromorphology and Space-Form concepts and their relationship with the interaction metaphors, technological components and application areas.

In conclusion, the perception and reasoning processes based on spectral energies and shapes, along with their motion and growth processes, are in conformity with the developed methodology and tools applied in the interactive sonification use cases. These concepts were implemented through a framework design that highlights multiple perspective spatial interaction and user defined scope activation. As such, the musical functions of singularity and regularity were transposed to data exploration and monitoring, which were then translated into sonic feedback. As hinted during the SoundField project, it is our claim that the same functional strategies and related implementation are applicable within the electroacoustic composition and performance realm since they rest on enabling the user with the appropriate tools that adequately support his scientific expertise as well as his artistic intentionality. A summary of the concepts addressed in this section is depicted in Fig. 53.

2.3.6 Conceptual Framework for composition environment development

In electroacoustic music, the concept of musical gesture as a materialization of the composer's inner musical intention has always been present at different levels of conception, both within the non-realtime compositional and realtime performance levels. For example, one can point out the expressive use of the mixing board's faders in both the mixing and spatialization processes. It is a trivial, but nevertheless good example of an embodiment-based discourse that incorporates the physical factor in the creational process. Even more interesting is the fact that this process can be used across different levels of granularity throughout the musical work, ranging from individual amplitude envelope of a sound object to post-production panning of entire sections.

2.3.6.1 Expansion of developed framework within a compositional scenario

In relation to the above faders example, consider the extension of an embodied driven interface to a full body scale, through the application of direct manipulation and object oriented interface concepts. The use of physical faders is a significant step towards direct action from obtaining the same results by, for example, drawing amplitude envelops in a 2D DAW interface software. The same principle can be obtained by the user following a path object representing an audio track in a immersive 3D environment while drawing the respective amplitude envelop (in which amplitude is encoded by distance to the object of 3D point markers which are created at a certain rate or on demand and then interpolated) [4]. The immersive 3D environment combined with these techniques enables the use of simultaneous, multiple control which doesn't require visual focus. For example, both the hands can be used for concurrent control of different actions as probing a specific region of the dataset with one hand and adjusting its amplitude in the global mix or modulation rate with the other. Such an environment also stimulates cooperative analysis as multiple users can be "around" the dataset(s) as if they were around a meeting table, simultaneously analyzing multiple documents.

The virtual inspection window concept can be used for accessing sound objects and/or functions through a time-based, dynamic palette of, for example, activation and modulation of sound samples or sound effects. Taking as a reference the previously described sonification use cases, the selection and activation of sonically represented data can be expanded to sound object (sample) activation, allowing the control of both synthesis (additive, subtractive, spectral, granular, physical model, ...) and modulation algorithms (effects, filtering, envelopes, ...) in a corporeal way. For example, the user's left hand could select a sound object (composed by an individual or a set of sample/synthesis, depending on the radius size of the virtual inspection tool) while the right hand would select a modulation algorithm (e.g. sound effect) for applying to the selected source. As in the interaction sonification cases, the physical distance between the user (virtual inspection tool) and activation points (virtual inspection window) would provide the mixing of both source amplitude and effect participation in the mix. This methodology goes in the direction of facilitating a better connection between user and sound by establishing a bridge between physical behavior, musical gesture and sound object. As advocated in [151], it puts the user in close contact with the compositional material where the latter is morphologically configurable in order to maximize gestural interaction within a embodied, real-time interaction scenario.

Additionally, by founding this interaction on the manipulation of virtual objects in space, one can take the further advantage of everyday life interaction metaphors. It allows the extrapolation of real world interaction by applying it to

the manipulation of objects that would be impossible to handle in real life (for example, a tree size drum stick hitting a pool size snare). These fictitious performance scenarios are a good example of the exclusive possibilities afforded by the use of immersive virtual environments.

2.3.6.2 Preliminary categorization and example applications

In this final section, we will present two main proposals for the categorization of immersive interfaces based on the Space-form categorization, namely space centered and user centered, and a related example of their application. This division is routed on the interaction relationship between space and user. These two categories can support both individual and collaborative application scenarios. In either case, the interface provides the means to convey the interpretation previously addressed in section 2.3.5.1 of spectromorphological classification of gestural/aural elements towards their application in immersive interface development.

The first category "space centered" applies to the use cases where users share and operate a common set of interface elements. As such, it emphasizes the social and public spaces of interaction. Related applications of this category can be found in some of the previously presented use cases of SoundField, namely SoundPath and SoundMaze (see section 2.3.4.2). The virtual objects are distributed in space and are accessed by the user by means of exploration and occupancy (through scope variable, virtual microphones). Consequently, the interaction space, constituted by the superposition of the virtual scene onto the real room, can be regarded as an enacting prone, composed circumspace that affords perspectival shifts. It conveys the possibility of the user to enjoy multiple views and vantage points of the visual and aural scene. Lastly, it emphasizes the application of scaling processes to allow the creation of microphone spaces, aurally magnifying the sound objects under the user's scrutiny.

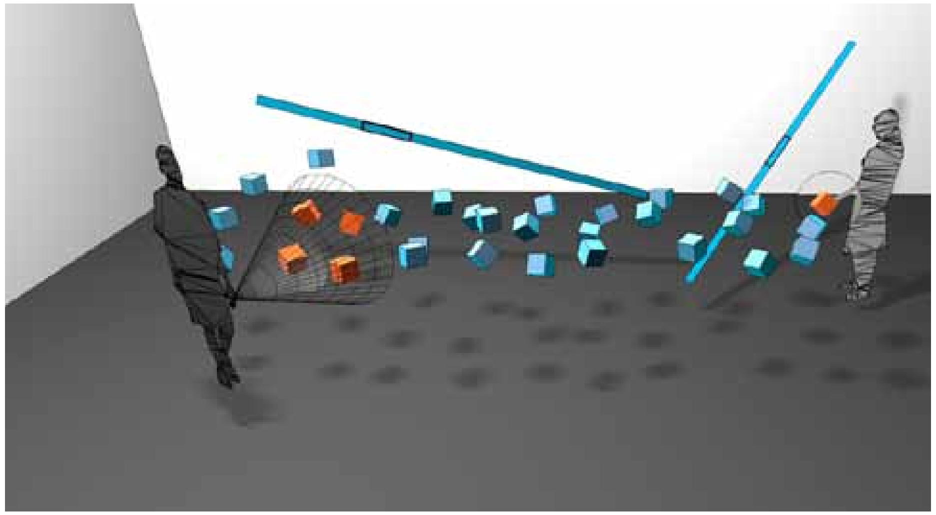


Fig. 54 Space centered example use case.

As an example, consider an interface that is built around a set of virtual object types that are placed on a three dimensional grid (See Fig. 54). Together they construct a latent composition that can be explored both globally or in detail using a variable scope or microphone space. The change of perspective of each individual user defines the sonic outcome. Users, changing the scope, blend objects within their reach into the overall sound-palette. They can further be controlled in detail by grabbing or releasing the individual objects.

There are two types of virtual objects available in the environment, namely cubes and beams. Cubes represent monophonic sounds that can be manipulated by rotation, dynamically changing three parameters attached to the X, Y and Z axis of the object. Apart from the rotation, the objects can be repositioned in the environment, enabling the clustering or dispersion of the rendered sounds. The beams, available in the environment, represent stereo sounds, which behave similar to the cubes. However, their orientation distributes the sound over a pair of opposing speakers. A user defines the actual position of the sound by moving his along the virtual object.

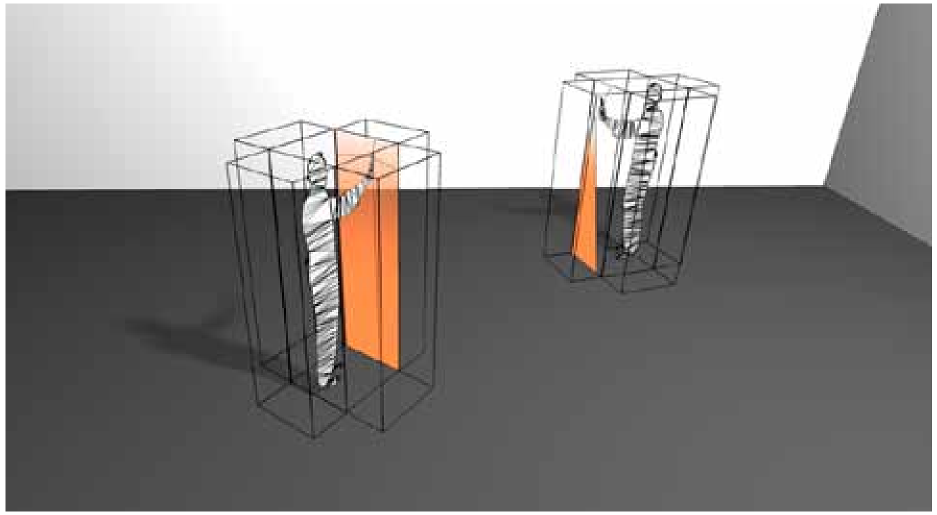


Fig. 55 User centered example use case.

The second category, "user centered", applies to the use cases where users operate an individual interface in common space. A prototype of this category can be found in the previously presented SoundObjects use case (see section 2.3.4.2). The virtual objects are placed at different angles around the user and follow his position throughout the interaction process. Furthermore, the virtual objects can change their morphology according to the user's intentions and are accessed through limited scope, virtual microphones within his reach. This category emphasizes the use of the personal and ensemble space as well as a proprioceptive based interaction. It allows the creation of two distinct levels of interaction, one concerning the manipulation by each user of their respective individual interface, and a second related to the relative position of the users within the interaction room. In this second example, each user can add four layers of a composition (See Fig. 55). The four sound-layers are mapped onto an open cube's Z and X planes. By touching the plane in front of him, the user activates one or two possible X/Y-controllers (resembling the functionality of a trackpad). A user would change orientation to face whatever plane he wants to access. For this, the open cube's orientation stays fixed. A user has two ways of interacting with the sound in front of him; either reaching out with the hand faced down, enabling the control of two parameters controlled by moving in the X and Z axis, or by reaching out with his palm facing away from him, enabling the control over two additional parameters in the X and Y axis. With the palm faced down, the user controls the panning of the sound, moving the hand from left to right and the amplitude, moving the hand, and thus the sound, further away or closer to him. The other set of controls, determines the timbre of the sound using a resonant filter, controlling both cutoff frequency (X-axis) and resonance (Y-axis). Furthermore, the bounding box with the four layers follows the user when he repositions himself, making a spatial positioning possible. In this way, each individual user gains direct control over a subset of the individual layers composing the overall sound, while their relative position adds a second more global dynamic level of control.

2.3.7 Conclusions

Based on the commonalities between electroacoustic music theory and interactive exploration of data through sonification techniques, we have presented a framework that allows for the exploration of user-centered interfaces for immersive musical composition. Following a musicological exploration of the electroacoustic practice of Schaeffer, Stockhausen, Wishart and Smalley, we identified relevant representation strategies and interaction metaphors based on multilevel mapping and scope variation. These constituted the point of departure for the design and implementation of a software framework that supports immersive interface prototyping within an embodied music cognition approach.

The software framework satisfies useful requirements pertaining performance, integrability and scalability, that have been identified as key for the interactive sonification field of research. Following a user centered iterative approach, the interaction metaphors were expanded and successfully applied in an interactive art domain.
This led to an in-depth validation of the developed solutions from a musical production/ performance point of view, based on Smalley's Spectromorphological and Space-Form classification framework. As a result, we obtained a preliminary framework for musical interfacing, based on direct manipulation, spatial mapping, and collaborative participation.

chapter 3 /

conclusion

3.1 Discussion and Future Work

In this thesis, a conceptual framework and related technological implementation for non speech sound-based communication has been presented. The main purpose of this work is to propose a bridge between sound design, interaction strategies and software tools. By combining these domains into a consistent framework, this proposal discloses the ability to both prototype multimodal, interactive interfaces as well as interconnecting the rendering backends within a unified conceptual and technological middleware development approach.

Rather than focusing on one single aspect of the framework (for example, a usability study regarding all the morphological variations of the virtual inspection tool metaphor), the discussed results have been achieved by following a broader development plan that interconnects several aspects of sound design, interaction strategies and software tools. As such, this work represents a point of departure for the development of sonic spatial interaction where both the interaction metaphors and the technological realizations proposals offer room for further expansion and improvement.

The main achievements of this thesis can be summarized as follows:

- A conceptual framework for the representation and access of sound objects has been developed. This conceptual organization was inspired by an exploration of the theories and compositional practices developed by Schaeffer, Stockhausen, Wishart and Smalley regarding electroacoustic music's analysis and production. Through an analysis of the relationships between the sonic, spatial and human elements, interaction metaphors were identified and formulated based on the concepts of multilevel mapping and scope variation, namely the virtual inspection window and tool. Following an embodied music cognition approach, we developed a methodology for sonically representing data within an interactive, multimodal scenario.

- The described use cases in interactive sonification and interactive art aim at illustrating the potential of the developed interaction metaphors. As such, they constitute a demonstration of the versatility and applicability of these metaphors within multiple interaction scenarios. Further variations of the virtual inspection window and tool can be explored in terms of morphology (until now, the virtual objects' morphology is confined to basic geometric forms - sphere, cube, line, etc.), and in terms of spatial positioning and layering (for multilevel mapping). As such, immediate future work can take the cases presented here and expand them through the study of gesture to sound correspondence based of electroacoustic music realizations. For example, one can imagine a scenario where a user defines the morphology of a virtual inspection tool by gesturally interpreting musical examples of Smalley's spectromorphogical classification.

Then, the user can assign the targeted dataset to the one that, according to theuser's expertise, maximizes the performance of the exploratory or monitoring tasks. Additionally, further investigation can address the perceptual added value in the application of spatialization techniques used in electroacoustic practice concerning the manipulation of virtual elements in space.

- A technological framework, JOINDER, has been developed. This software framework constitutes a foundation for multimodal and interactive interface prototyping through a user centered development methodology. This software solution delivers data-management facilities and complies with the research fields' requirements of multimodal interfacing, heterogeneous technologies integration and real-time performance. Furthermore, in order to accommodate future extensions, the design, implementation and founding technology are in conformity with the main requirements for auditory display, namely: scalability (e.g. distributed multicore setups) and portability (e.g. multi-platform compatibility).

- Both the development methodology (definition of context requirements, requirements driven technology choice and top-down contextual approach) as the implementation strategy (user centered prototyping and bottom-up component integration) have proven to be successful choices. However, by valorizing architectural design over exhaustive codebase development, the current number of features offered by JOINDER is lower than the one supplied by mature, single modality specialized tools, such as the systems presented in the state of the art. Nevertheless, by laying down an infrastructure that embraces the different modalities (vision, sound, proprioceptive human interfaces) in an equalitarian view, it provides a versatile hub for interconnecting the referred technologies and an integrated platform for multimodal prototyping. As such, future developments can follow along several paths. They can range from incorporation of new input devices to web servicing access (see discussion and future work in JOINDER dedicated section). Furthermore, specific issues concerning the trackable area and the usable number of rigid bodies that were used during the SoundField iterations as well as the suitability between hardware devices and the properties of the virtual elements they control (e.g. the use of a bend sensor controlled the scope radius of the virtual inspection tool) can be further investigated and expanded. In addition, the standardization of the user centered development methodology combined with the extension of the application domain into other areas beyond interactive sonification and art installations can further strengthen the maturation of the technology presented in this thesis. These areas can include, for example, interactive multimedia presentation, multimodal technology development, interactive advertising, gaming development and entertainment industry.

3.2 Conclusion

In this thesis, a conceptual and software framework for non-verbal sound communication within the domains of interactive sonification and musical composition has been developed. Driven by the theories of electroacoustic music and embodied music cognition, the importance of interaction in top-down and bottom-up cognitive processes constitutes the primary directive for this work. As such, a human-centered, integration-oriented methodology was presented that provides an alternative to previous developments in data to sound translation tools.

As such, the development of human/space multimodal symbiosis expressed in this work is expected to provide a human centered foundation for the design and implementation of more efficient tools within the auditory display and musical production community.

publications

- Coussement P, Leman M, Diniz N, Demey M (2008) The Heart as an Ocean exploring meaningful interaction with biofeedback. Proceedings of Audio Mostly, pp. 132-134.

- Deweppe A, Diniz N, Coussement P, Lesaffre M, Leman M (2010) Evaluating Strategies for Embodied Music Interaction with Musical Content. Proceedings of Conference on Interdisciplinary Musicology "Nature versus Culture" (CIM10), pp. 23-24 of July, 2010. Sheffield; United Kingdom.

- Diniz N, Demey M, Leman M (2010) An Interactive Framework for Multilevel Sonification. Proceedings of ISon 2010, 3rd Interactive Sonification Workshop, pp. 65-68, KTH, Stockholm, Sweden, April 7, 2010.

- Diniz N, Deweppe A, Demey M, Leman M (2010) A Framework for music-based interactive sonification. Proceedings of 16th International Conference on Auditory Display (ICAD 2010), pp. 345-351, George Washington University, Washington DC, USA, June 9-15, 2010.

- Deweppe A, Diniz N, Coussement P, Leman M (2011) A methodological framework for the development and evaluation of user-centered art installations. Journal of Interdisciplinary Music Studies. Nature versus Culture special issue. Spring 2011, Vol. 5, No. 1, pp. 19-39.

- Diniz N, Coussement P, Deweppe A, Demey M, Leman M (2012) An embodied music cognition approach to develop multilevel interactive sonification. Journal of Multimodal User Interfaces, pp. 1-9, Springer Verlag, 2012.

- Diniz N, Deweppe A, Coussement P, Leman M. JOINDER: A software framework for networked multimodal interface development. (Submitted)

- Coussement P, Diniz N, Deweppe A, Leman M. SoundField: an artistic research place for embodied social interaction. (Submitted)

- Diniz N, Coussement P, Deweppe A, Leman M. Multilevel immersive interfaces for electroacoustic music composition and performance. (To be submitted)

- Deweppe A, Diniz N, Coussement P, Leman M. SoundField: adopting Living Lab methodologies for active exploration of creative opportunities in embodied mediation technologies for cultural applications. (In preparation)

references

1. Adeney, RW (2011) The HarmonyGrid: music, space and performance in grid music systems. PhD Thesis, Queensland University of Technology.

2. Almada A, Lopes G et al. (2009) YVision: A General Purpose Software Composition Framework. Human-Computer Interaction. New Trends, pp. 779-788, Springer.

3. Anderson J (2000) A provisional history of spectral music. Contemporary Music Review, Vol. 19, No. 2, pp. 7-22, Taylor & Francis.

4. Barrass S, Barrass T (2006) Musical creativity in collaborative virtual environments. Virtual Reality, Vol. 10, No. 2, pp. 149-157, Springer.

5. Barri T (2009) Versum: audiovisual composing in 3d. In Proceedings of the 15th International Conference on Auditory Display, pp. 18-21, Copenhagen, Denmark.

6. Bauer M, Hilliges O et al. (2003) Integrating studierstube and dwarf. International Workshop on Software Technology for Augmented Reality Systems.

7. Beilharz K, Ferguson S (2009) An Interface and Framework Design for Interactive Aesthetic Sonification" in Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark, May 18-22, 2009.

8. Bernard JW (1994) Voice leading as a spatial function in the music of Ligeti. Music Analysis, Vol. 13, No. 2/3, pp. 227-253, JSTOR.

9. Berthaut F, Desainte-Catherine M, Hachet M (2010) DRILE: an immersive environment for hierarchical live-looping. Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), Sydney, Australia.

10. Bierbaum A, Just C et al. (2001) VR Juggler: A virtual platform for virtual reality application development. Proceedings Virtual Reality, pp. 89-96, IEEE.

11. Blattner MM, Sumikawa DA, Greenberg RM (1989) Earcons and icons: their structure and common design principles. Human Computer Interaction, Vol. 4, No. 1, pp. 11–44.

12. Bottoni P, Faralli S et al. (2006) CHAMBRE: integrating multimedia and virtual tools. Proceedings of the working conference on Advanced visual interfaces, pp. 285-292, ACM.

13. Bouchet J, Nigay L (2004) ICARE: a component-based approach for the design and development of multimodal interfaces. CHI'04 extended abstracts on Human factors in computing systems, pp. 1325--1328}, ACM.

14. Boulez P (1971) Boulez on Music Today. Translate by Susan Bradshaw and Richard Rodney Bennett. Cambridge, Mass. Harvard University Press.

15. Bovermann T, Hermann T, Ritter H (2006) Tangible data scanning sonification model, Proceedings of the 12th International Conference on Auditory Display, London, pp. 77-82, UK June 20 - 23, 2006.

16. Brattico E, Sassanelli F (2000) Perception and Musical Preferences in Wishart's Work. Journal of New Music Research, Vol. 29, No. 2, pp. 107-119, Taylor & Francis.

17. Bregman AS (1994) Auditory scene analysis: the perceptual organization of sound. MIT Press, Cambridge.

18. Brindle RS (1966) Serial composition. Vol. 66, Cambridge University Press.

19. Brindle RS (1987) The new music: the avant-garde since 1945. Oxford University Press.

20. Bruegge B, Klinker G (2005) DWARF Distributed Wearable Augmented Reality Framework. Chair for Applied Software Engineering, Munich, Germany, Technische Universitat Munchen.

21. Cadiz RF, Lipscomb SD (2004) A perceptual map of electro-acoustic sounds based on spectromorphological features. Proceedings of the 8th International Conference of Music Perception and Cognition, Evanston, Illinois, USA.

22. Camilleri L (1993) Electro-acoustic music: analysis and listening processes. Sonus Contemporary Music Materials, Vol. 1, No. 3, pp. 26.

23. Camurri A, Trocca R, Volpe G (2002) Interactive systems design: a KANSEI-based approach. Proceedings of the conference on new interfaces for musical expression, pp. 1-8. National University of Singapore, Singapore.

24. Caramiaux B, Fdili Alaoui S et al. (2011) Proceedings of the 14th International Conference on Digital Audio Effects (DAFx-11), Paris, France, September 19-23.

25. Chang D, Nesbitt KV (2006) Developing gestalt-based design guidelines for multi-sensory displays. Proceedings of the 2005 NICTA-HCSNet multimodal user interaction workshop, vol 57. Australian Computer Society, Sydney, pp. 9–16.

26. Childs E (2003) Musical Sonification Design. MA thesis, Dartmouth College, 2003.

27. Chion M (1983) Guide To Sound Objects. Pierre Schaeffer and Musical Research, (Trans. John Dack and Christine North, 2009), http://www.ears.dmu.ac.uk/.

28. Clarke M (1998) Extending Contacts: The Concept of Unity in Computer Music. Perspectives of New Music, Vol. 36, No. 1 (Winter, 1998), pp. 221-246.

29. Coenen A (1994) Stockhausen's Paradigm: A Survey of His Theories. Perspectives of New Music, pp. 200-225, JSTOR.

30. Coussement P, Demey M, Leman M (2010) Dynamic mapping strategies for interactive art installations: an embodied combined HCI HRI HHI approach, in: Proceedings of the Artech, 5th International Conference on Digital Arts "Envisioning Digital Spaces", Guimarães, Portugal.

31. Coussement P, Diniz N, Deweppe A, Leman M (2012) SoundField: an artistic research place for embodied social interaction", Art and Research. (Submitted)

32. Darken R, McDowell P, Johnson E (2005) Projects in VR: the Delta3D open source game engine. Computer Graphics and Applications, Vol. 25, No. 3, pp. 10-12, IEEE.

33. Davachi S (2011) Aesthetic Appropriation of Electronic Sound Transformations in Ligeti's Atmospheres. Musicological Explorations, Vol. 12, No. 0, pp. 109-110.

34. David MF (1993) The history and future of direct manipulation. Behaviour & Information Technology, Vol. 12, No. 6, pp. 315-329, Taylor & Francis.

35. Davison BK, Walker BN (2007) Sonification Sandbox Reconstruction: Software Standard for Auditory Graphs," in International Conference on Auditory Display, Montreal, Canada.

36. De Campo A, Frauenberger C, Holdrich R (2004) Designing a generalized sonification environment. Proceedings of the International Conference on Auditory Display 2004.

37. Delalande F (2007) Towards an Analysis of Compositional Strategies. Circuit : musiques contemporaines, Vol. 17, n 1, p. 11-26.

38. Deliege C (1989) On form as actually experienced. Contemporary Music Review, Vol. 4, No. 1, pp. 101-115, Taylor & Francis.

39. Deweppe A, Diniz N, Coussement P, Leman M (2011) A methodological framework for the development and evaluation of user-centered art installations. Journal of Interdisciplinary Music Studies. Nature versus Culture special issue. Spring 2011, Vol. 5, No. 1, pp. 19-39.

40. Deweppe A, Diniz N, Coussement P, Leman M. "SoundField: adopting Living Lab methodologies for active exploration of creative opportunities in embodied mediation technologies for cultural applications. (In preparation)

41. Deweppe A, Diniz N, Coussement P, Lesaffre M, Leman M (2010) Evaluating strategies for embodied music interaction with musical content. Proceedings of the 10th international conference on interdisciplinary musicology, Sheffield United Kingdom.

42. Di Scipio A (1998) Compositional Models in Xenakis's Electroacoustic Music. Perspectives of New Music, pp. 201-243, JSTOR.

43. Diniz N, Coussement P, Deweppe A, Demey M, Leman M (2012) An embodied music cognition approach to multilevel interactive sonification. Journal on Multimodal User Interfaces. Springer Berlin / Heidelberg, Vol. 5, No. 3, pp. 211-219.

44. Diniz N, Demey M, Leman M (2010) An interactive framework for multilevel sonification. Proceedings of the Interactive Sonification Workshop "Human Interaction with Auditory Displays", pp. 65-68.

45. Diniz N, Deweppe A, Demey M, Leman M (2010) A Framework for music-based interactive sonification. Proceedings of the 16th international conference on auditory display, Washington, DC, USA, pp. 345-351.

46. Dumas B, Lalanne D, Oviatt S (2009) Multimodal Interfaces: A Survey of Principles, Models and Frameworks. Human Machine Interaction. Lecture Notes in Computer Science, Vol. 5440, pp. 3-26, Springer Berlin / Heidelberg.

47. Eckel G (1999) Applications of the cyberstage spatial sound server. AES 16th International Conference on Spatial Sound Reproduction.

48. Emmerson S (1986) The relation of language to materials, pp. 17-39, Macmillan London.

49. Emmerson S (1989) Composing strategies and pedagogy. Contemporary music review, Vol. 3, No. 1, pp. 133-144, Taylor & Francis.

50. Emmerson S (1998) Aural landscape: musical space. Organised Sound, Vol. 3, No. 2, pp. 135-140, Cambridge University Press.

51. Endres C, Butz A, MacWilliams A (2005) A survey of software infrastructures and frameworks for ubiquitous computing. Mobile Information Systems. IOS Press. Vol. 1, No. 1, pp. 41-80.

52. Ferey N, Nelson J et al. (2009) Multisensory VR interaction for protein-docking in the CoRSAIRe project. Virtual Reality, Vol. 13, No. 4, pp. 273-293, Springer.

53. Fielding RT, Taylor RN (2002) Principled design of the modern Web architecture. ACM Transactions on Internet Technology (TOIT). Vol. 2, No. 2, pp. 115-150. ACM.

54. Figueroa P, Green M, Hoover HJ (2002) InTml: a description language for VR applications. Proceedings of the seventh international conference on 3D Web technology, pp. 53-58, ACM.

55. Fineberg J (2000) Spectral music: history and techniques, Vol. 19, Routledge.

56. Flowers JH (2005) Thirteen years of reflection on auditory graphing: promises, pitfalls, and potential new directions. Proceedings of the International conference on auditory display (ICAD), Limerick, Ireland.

57. Forte A (1973). The Structure of Atonal Music. New Haven and London: Yale University Press.

58. Freed A, Schmeder A (2009) Features and future of open sound control version 1.1 for NIME. Proceedings of the 9th Conference on New Interfaces for Musical Expression. Carnegie Mellon School of Music, Pittsburgh, Pennsylvania, USA.

59. Fry B (2008) Visualizing data:[exploring and explaining data with the processing environment. O'Reilly Media, Inc.

60. Gaver WW (1986) Auditory icons: using sound in computer interfaces. Human-Computer Interaction, Vol. 2, No. 2, pp. 167-177.

61. Gaver WW (1993) What in the world do we hear?: An ecological approach to auditory event perception. Ecological psychology, Vol. 5, No. 1, pp. 1-29, Taylor & Francis.

62. Gibson JJ (1979) The ecological approach to perception. Haughton Mifflin, Boston.

63. Godoy RI (2004) Gestural imagery in the service of musical imagery. Gesture-Based Communication in Human-Computer Interaction, pp. 99-100, Springer.

64. Gossman J (2010) From metaphor to medium: Sonification as extension of our body. Proceedings of the16th International Conference on Auditory Display, pp. 145-152.

65. Gossmann J (2005) Towards an auditory representation of complexity. Proceedings of 11th Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9.

66. Grisey G, Fineberg J (2000) Did you say spectral?. Contemporary music review, Vol. 19, No. 3, pp. 1-3, Taylor & Francis.

67. Gröhn M, Takala T (1994) Magicmikes multiple aerial probes for sonification of spatial datasets. Proceedings of the 2nd International Conference on Auditory Display, pp. 271–272, Santa Fe, New Mexico, 1994.

68. Hall E (1990) The Hidden Dimension. New York: Anchor Books.

69. Harley J (2002) The electroacoustic music of Iannis Xenakis. Computer Music Journal, Vol. 26, No. 1, pp. 33-57, MIT Press.

70. Harvey J (1975) The music of Stockhausen: an introduction. University of California Press, Berkeley.

71. Harvey J (1981) "Mortuos Plango, Vivos Voco": A Realization at IRCAM. Computer Music Journal, Vol. 5, No. 4, pp. 22-24, JSTOR.

72. Hermann T (2008) Taxonomy and definitions for sonification and auditory display. Proceedings of the 14th international conference on auditory display, Paris.

73. Hermann T, Hunt A (2005) Guest editors' introduction: an introduction to interactive sonification. IEEE Multimedia, Vol. 12, No. 2, pp. 20–24. IEEE.

74. Hermann T, Neuhoff J, Hunt A (2011) The Sonification Handbook. Logos Verlag, Berlin, Germany.

75. Hughes CE, Stapleton CB et al. (2005) Mixed reality in education, entertainment, and training. Computer Graphics and Applications, Vol. 25, No. 6, pp. 24-30, IEEE.

76. Jacquemin C (2004) Architecture and experiments in networked 3d audio/graphic rendering with virtual choreographer. Proceedings of Sound and Music Computing conference (SMC'04).

77. Jaimes A, Sebe N (2007) Multimodal human-computer interaction: A survey. Computer Vision and Image Understanding. Vol. 108, No. 1-2, pp. 116-134, Elsevier.

78. JavaOSC. http://www.illposed.com/software/javaosc.html. Last accessed: April 24th, 2012.

79. Javolution. http://javolution.org/. Last accessed: April 24th, 2012.

80. JBullet. http://jbullet.advel.cz/. Last accessed: April 24th, 2012.

81. JEPP http://jepp.sourceforge.net/ Last accessed: July 18th 2012.

82. JMonkey engine http://jmonkeyengine.com/ Last accessed: August 20th 2012.

83. JPype http://jpype.sourceforge.net/ Last accessed: July 18th 2012.

84. JPython http://www.jython.org/ Last accessed: July 18th 2012.

85. Kaper HG, Wiebel E, Tipei S (1999) Data sonification and sound visualization. Computing in science & engineering, Vol. 1, No. 4, pp. 48-58. IEEE.

86. Kaplan J, Yankelovich N (2011) Open wonderland: Extensible virtual world architecture. Internet Computing, No. 99, pp. 1-1, IEEE.

87. Kato H (2002) ARToolKit: Library for Vision-based Augmented Reality. IEICE, PRMU, pp. 79-86.

88. Kendall GS (1991) Visualization by ear: auditory imagery for scientific visualization and virtual reality. Computer Music Journal, Vol. 15, No. 4, pp. 70-73. JSTOR.

89. Kendall GS (2010) Meaning in electroacoustic music and the everyday mind. Organised Sound, Vol. 15, No. 1, pp. 63-74, Cambridge University Press.

90. Kiefer C, Collins N, Fitzpatrick G (2008) HCI methodology for evaluating musical controllers: a case study. Proceedings of the conference on new interfaces for musical expression, Genova, Italy, pp. 87-90.

91. Kohl J (1983) The Evolution of Macro-and Micro-Time Relations in Stockhausen's Recent Music. Perspectives of New Music, Vol. 22}, No. 1/2, pp. 147-185, JSTOR.

92. Kohl J (1990) Into the Middleground: Formula Syntax in Stockhausen's Licht. Perspectives of New Music, pp. 262-291, JSTOR.

93. Konig WA, Radle R, Reiterer H (2010) Interactive design of multimodal user interfaces. Journal on Multimodal User Interfaces, Vol. 3, No. 3, pp. 197-213, Springer.

94. Kramer G, Walker B et al (1997) Sonification report: status of the field and research agenda. Available at: http://dev.icad.org/node/400.

95. Kramer JD (1978) Moment form in twentieth century music. The Musical Quarterly, Vol. 64, No. 2, pp. 177-194, JSTOR.

96. Kryonet. http://code.google.com/p/kryonet/. Last accessed: April 24th, 2012.

97. Kuck R, Wind J et al. (2008) Improving the avango vr/ar framework: Lessons learned. Virtuelle und Erweiterte Realitat, Vol. 5, pp. 209-220.

98. Langner G (1997) Temporal processing of pitch in the auditory system. Journal of New Music Research, Vo. 26, No. 2, pp.116–132.

99. Leman M (2008) Embodied music cognition and mediation technology. MIT Press, Cambridge.

100. Leman M, Lesaffre M, Nils L, Deweppe A (2010) User-oriented studies in embodied music cognition. Musicae Scientia, Vol. 14, No. 2/3. Université de Liège, Belgium.

101. Leman M, Naveda L (2010) Basic gestures as spatiotemporal reference frames for repetitive dance/music patterns in samba and Charleston. Music Percept 28, pp. 71-91.

102. Leonard J, Ng K (2011) Music via Motion: A distributed framework for interactive multimedia performance. The 17th International Conference on Distributed Multimedia Systems, Florence, Italy.

103. Leonardi G, Kohl J (1998) Inori: Microcosm/Macrocosm Relationships and a Logic of Perception. Perspectives of New Music, pp. 63-90, JSTOR.

104. Lodha SK, Beahan J, Heppe T, Joseph A, Zane-Ulman B (1997) MUSE: A Musical Data Sonification Toolkit. International Conference on Auditory Display, Palo Alto, California.

105. Looser J, Grasset R et al. (2006) OSGART-A pragmatic approach to MR. University of Canterbury. Human Interface Technology Laboratory.

106. Maconie R (1972) Stockhausen's Mikrophonie I: Perception in Action. Perspectives of New Music, Vol. 10, No. 2, pp. 92-101, JSTOR.

107. Max/MSP/Jitter http://cycling74.com/products/max/ Last accessed: August 20th 2012.

108. Mulder A, Fels S, Mase K (1997) Mapping virtual object manipulation to sound variation. IPSJ SIG Notes, Vol. 97, No. 122, pp. 63–68.

109. Nielsen J, Landauer TK (1993) A mathematical model of the finding of usability problems. Proceedings of ACM INTERCHI'93 conference, pp. 206–213, Amsterdam, The Netherlands.

110. Nielsen J, Molich R (1990) Heuristic evaluation of user interfaces. Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people, CHI '90, pp. 249-256.

111. Norman D (2010) Natural user interfaces are not natural. Interactions, Vol. 17, No. 3, pp. 6-10. ACM.

112. Ohlenburg J, Herbst I et al. (2004) The MORGAN framework: enabling dynamic multi-user AR and VR projects. Proceedings of the ACM symposium on Virtual reality software and technology, pp. 166-169, ACM.

113. Olwal A, Feiner S (2004) Unit: modular development of distributed interaction techniques for highly interactive user interfaces. Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia, pp. 131-138, ACM.

114. Open Frameworks http://www.openframeworks.cc/ Last accessed: Last accessed: August 20th 2012.

115. OpenJMS. http://openjms.sourceforge.net/. Last accessed: April 24th, 2012.

116. OpenNI http://openni.org/ last accessed: August 20th 2012.

117. Paine DG (2004) Gesture and musical interaction: interactive engagement through dynamic morphology. Proceedings of the 2004 conference on New interfaces for musical expression, pp. 80-86, National University of Singapore.

118. Paine G (2002) Interactivity, where to from here?. Organised Sound, Vol. 7, No. 03, pp. 295-304, Cambridge University Press.

119. Papagiannakis G, Singh G, Magnenat-Thalmann N (2008) A survey of mobile and wireless technologies for augmented reality systems. Computer Animation and Virtual Worlds. Wiley Online Library. Vol. 19, No. 1, pp. 3-22.

120. Park JS (2011) AR-Room: a rapid prototyping framework for augmented reality applications. Multimedia Tools and Applications, Vol. 55, No. 3, pp. 725-746, Springer.

121. Pauletto S, Hunt A (2004) A toolkit for interactive sonification. Proceedings of the International Conference on Auditory Display 2004.

122. Polfreman R (2009) Frameworks 3d: composition in the third dimension. In Proceedings of the 9th International Conference New Interfaces for Musical Expression, Pittsburgh: Carnegie Mellon University, USA, pp. 226-229.

123. Ponder M, Papagiannakis G. et al. (2003) VHD++ development framework: Towards extendible, component based VR/AR simulation engine featuring advanced virtual character technologies. Proceedings of Computer Graphics International, pp. 96-104, IEEE.

124. Pousset D, Fineberg J, Hyacinthe R (2000) The works of Kaija Saariaho, Philippe Hurel and Marc-Andre Dalbavie—Stile concertato, stile concitato, stile rappresentativo. Contemporary music review, Vol. 19, No. 3, pp. 67-110, Taylor & Francis.

125. Pressnitzer D, McAdams S (2000) Acoustics, psychoacoustics and spectral music. Contemporary Music Review, Vol. 19, No. 2, pp. 33-59, Taylor & Francis.

126. Pure Data/GEM http://puredata.info/ Last accessed: August 20th 2012.

127. Reitmayr G, Schmalstieg D (2005) OpenTracker: A flexible software design for three-dimensional interaction. Virtual Reality, Vol. 9, No. 1, pp. 79-92, Springer.

128. Rose F (1996) Introduction to the pitch organization of French spectral music. Perspectives of New Music, pp. 6-39, JSTOR.

129. Scaletti C (1989) Composing Sound Objects in Kyma. Perspectives of New Music, Vol. 27, No. 1, pp. 42-69.

130. Schaeffer B, Goudeseune C (2003) Syzygy: native PC cluster VR. Proceedings Virtual Reality, 2003, pp. 15-22, IEEE.

131. Schmalstieg D, Fuhrmann A et al. (2002) The studierstube augmented reality project. Presence: Teleoperators and Virtual Environments, Vol. 11, No. 1, pp. 33-54.

132. Serrano M, Nigay L et al. (2008) The openinterface framework: a tool for multimodal interaction. CHI'08 extended abstracts on Human factors in computing systems, pp. 3501-3506, ACM.

133. Sheridan J, Sood G et al. (2004) Soundstudio4D - A VR interface for gestural composition of spatial soundscapes. Proceedings of The International Conference on Auditory Display, pp. 1-7.

134. Smalley D (1986) Spectro-morphology and structuring processes. The language of electroacoustic music, pp. 61-93, London: MacMillan Press.

135. Smalley D (1996) The listening imagination: listening in the electroacoustic era. Contemporary Music Review, Vol. 13, No. 2, pp. 77-107, Taylor & Francis.

136. Smalley D (1997) Spectromorphology: explaining sound-shapes. Organised sound, Vol. 2, No. 2, pp. 107-126, Cambridge University Press.

137. Smalley D (2007) Space-form and the acousmatic image. Organised Sound, Vol. 12, No. 1, pp. 35-58, Cambridge University Press.

138. Smalley R (1974) 'Momente': Material for the Listener and Composer: 1. The Musical Times, Vol. 115, No. 1571, pp. 23-28, JSTOR.

139. Smalley R (1974) 'Momente': Material for the Listener and Composer: The Musical Times, Vol. 115, No. 1574, pp. 289-295, JSTOR.

140. Stockhausen K (1965) Mikrophonie I, fr Tamtam, 2 Mikrophone, 2 Filter und Regler. Stockhausen, Texte zur Musik 3, Cologne: Verlag M. DuMont Schauberg, p. 5765.

141. Stockhausen K (1989) Stockhausen on Music. Lectures and Interviews compiled by Robin Maconie. London and New York: Marion Boyars.

142. Stockmann L, Berndt A, Rober N (2008) A Musical Instrument based on 3D Data and Volume Sonification Techniques. Proceedings of the Audio Mostly Conference 2008, pp. 72-79, Piteå, Sweden.

143. Sun Java Real-Time System. http://java.sun.com/javase/technologies/realtime/index.jsp. Last accessed: April 24th, 2012.

144. Sziebig G, Ruda I, Demiralp M, Mastorakis N (2009) Achieving total immersion: Technology trends behind Augmented Reality - A survey. Proceedings of WSEAS International Conference. Mathematics and Computers in Science and Engineering. No. 5.

145. The Computer Language Benchmarks Game http://shootout.alioth.debian.org/ Last accessed: August 20th 2012.

146. Thompson J, Kuchera-Morin JA et al. (2009) The Allobrain: An interactive, stereographic, 3D audio, immersive virtual world. International Journal of Human-Computer Studies, Vol. 67, No. 11, pp. 934-946, Elsevier.

147. Toop R (1999) Gyorgy Ligeti. Phaidon Inc Ltd.

148. Underkoffler J. G-Speak http://oblong.com/ Last accessed: August 21st 2012.

149. Vaggione H (1996) Articulating microtime, Computer Music Journal, Vol. 20, No. 2, pp. 33-38, JSTOR.

150. Valbom L, Marcos A. (2005) WAVE: Sound and music in an immersive environment. Computers & Graphics Vol. 29, No. 6, pp. 871-881, Elsevier.

151. Vaughan M (1994) The human-machine interface in electroacoustic music composition. Contemporary Music Review, Vol. 10, No. 2, pp. 111-127, Taylor & Francis.

152. Vecmath. http://java.net/projects/vecmath. Last accessed: April 24th, 2012.

153. Vickers P, Hogg B (2006) Sonification Abstraite/Sonification Concrete: An´ Aesthetic Perspective Space For Classifying Auditory Displays In The Ars Musica Domain. Proceedings of the 12th International Conference on Auditory Display, London, UK June 20 - 23.

154. Wakefield G, Smith W (2011) Cosm: A Toolkit for Composing Immersive Audio-Visual Worlds of Agency and Autonomy. Proceedings of the International Computer Music Conference (ICMC 2011), University of Huddersfield, England.

155. Wanderley MM, Orio N (2002) Evaluation of input devices for musical expression: borrowing tools from HCI. Computer Music Journal, Vol. 26, No. 3, pp. 62-76.

156. Weissmann S, Gunn C et al. (2009) jReality: a java library for real-time interactive 3D graphics and audio. Proceedings of the 17th ACM international conference on Multimedia, pp. 927-928, ACM.

157. Wilson CM, Lodha SK (1996) Listen: A Data Sonification Toolkit. International Conference on Auditory Display, Palo Alto, CA.

158. Winberg F, Hellstrom SO (2001) Qualitative aspects of auditory direct manipulation: a case study of the Towers of Hanoi. Proceedings of the 7th International Conference on Auditory Display, Citeseer.

159. Windsor L (2000) Through and around the acousmatic: the interpretation of electroacoustic sounds. Music, electronic media and culture, pp. 7-35, Ashgate.

160. Wishart T (1986) Sound Symbols and Landscapes. The Language of Electroacoustic Music (S. Emmerson, ed.). Macmillan, London, pp. 41-60.

161. Wishart T (1996) On sonic art. Amsterdam: Harwood Academic Publishers.

162. Wollrath A, Riggs R, Waldo J (1996) A distributed object model for the Java system. Computing Systems, Vol. 9, pp. 265-290, University of California Press.

163. Worrall D, Bylstra M, Barrass S, Dean R (2007) SoniPy: The design of an extendable software framework for sonification research and auditory display. Proceedings of the 13th International Conference on Auditory Display, Montreal, Canada, June 26.

164. Xenakis I (1966) The Origins of Stochastic Music. Tempo (New Series), Vol. 3, No. 78, pp. 9-12, Cambridge Univ Press.

165. Zattra L (2005) Analysis and analyses of electroacoustic music. Proceedings of the sound and music computing, Salerno, Italy.

appendix

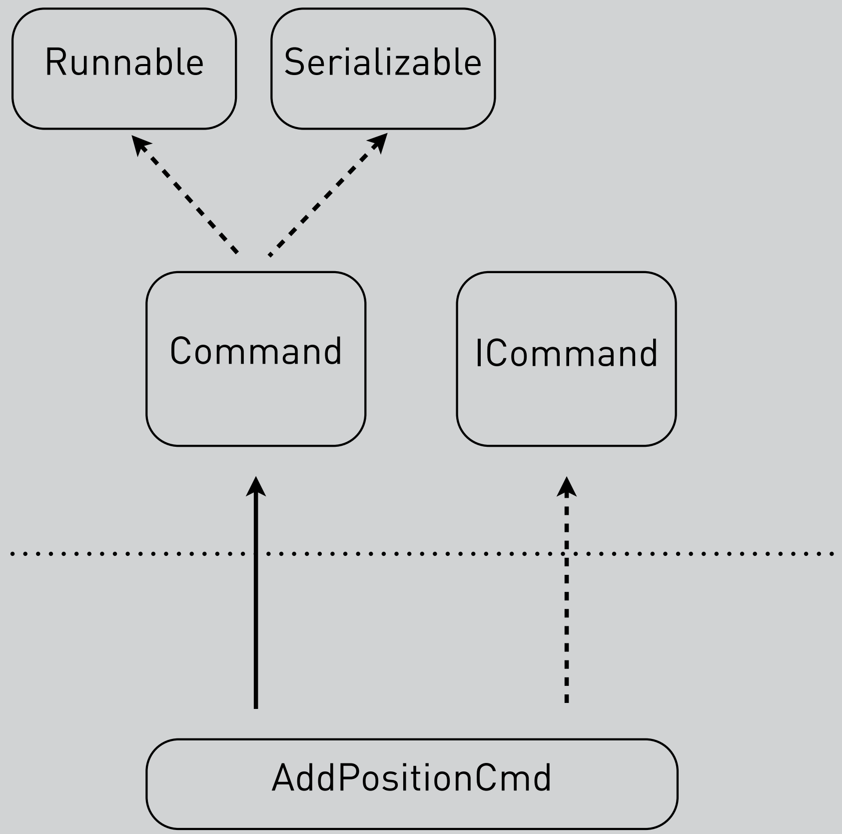


Fig. 1 Joinder components / Commands

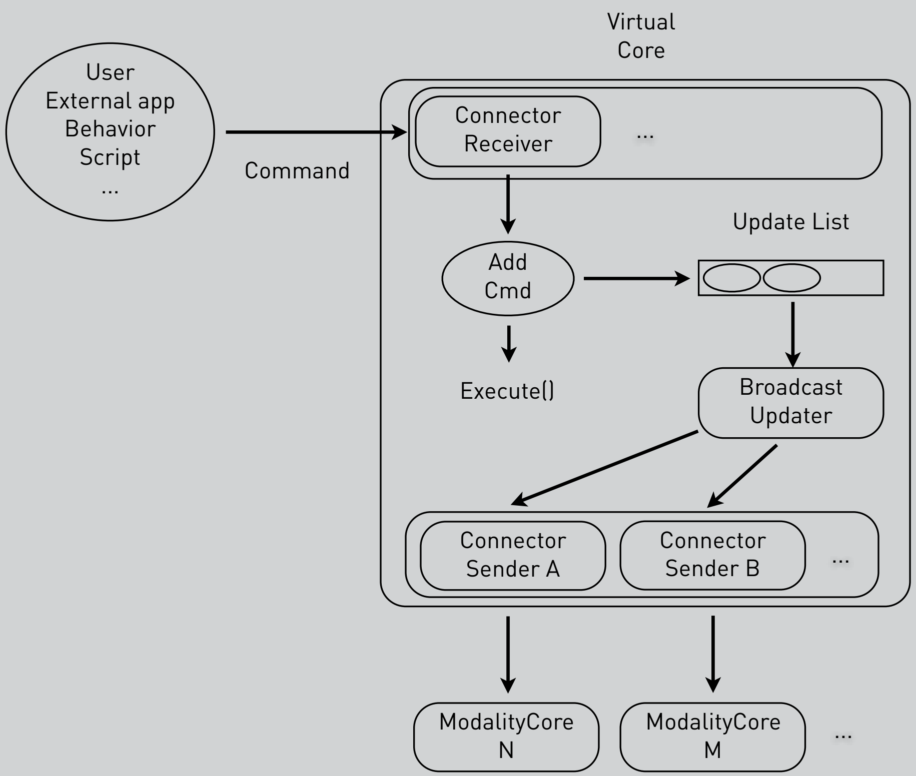


Fig. 2 Joinder components / Broadcast behavior 1

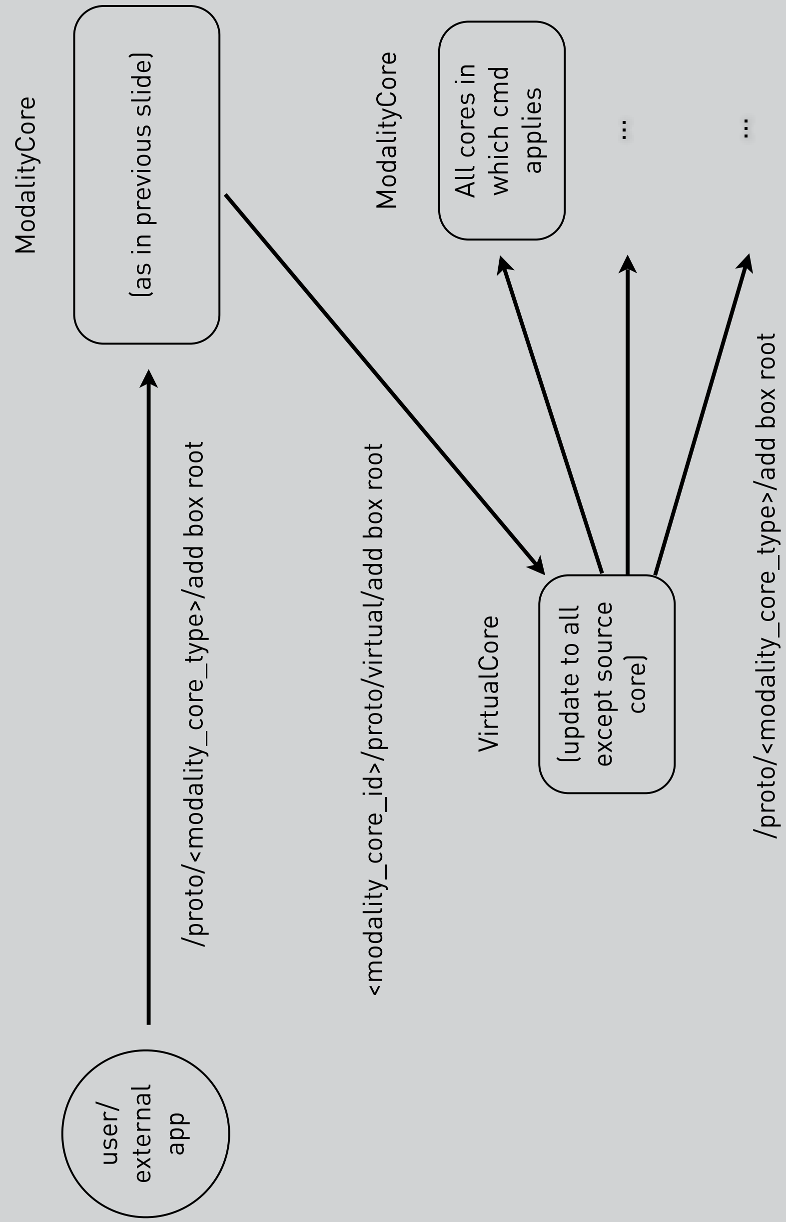


Fig. 3 Joinder components / Broadcast behavior 2

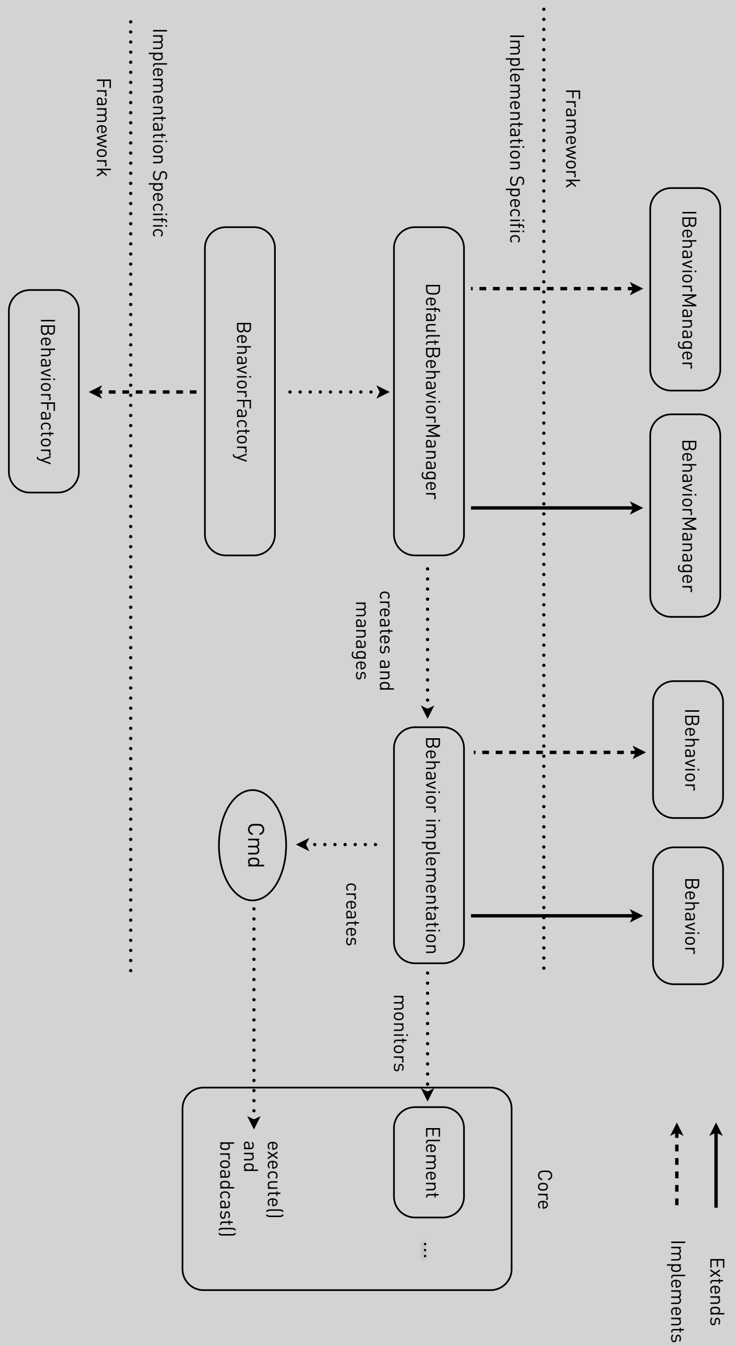


Fig. 4 Joinder components / Behaviors

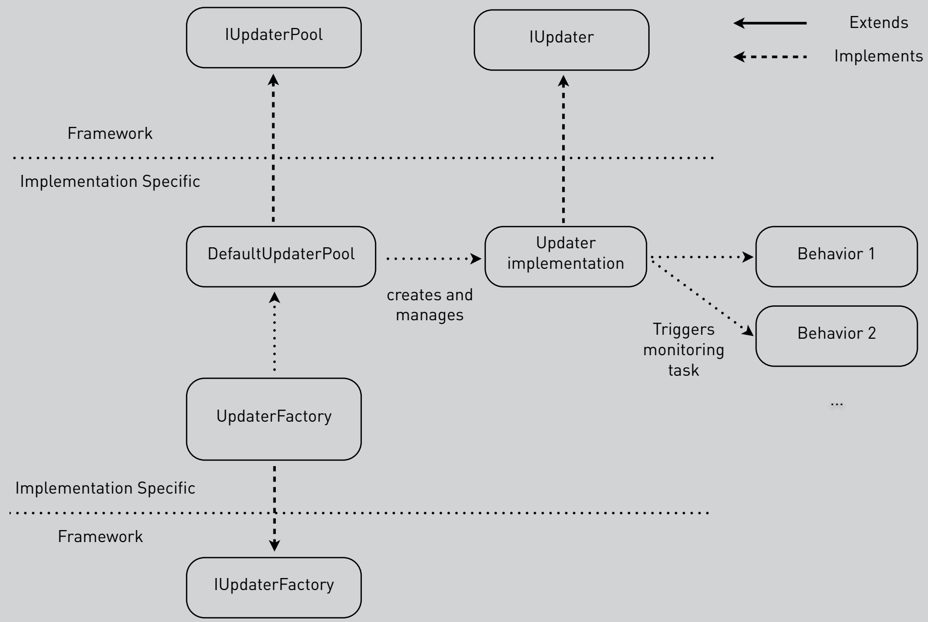


Fig. 5 Joinder components / Updaters

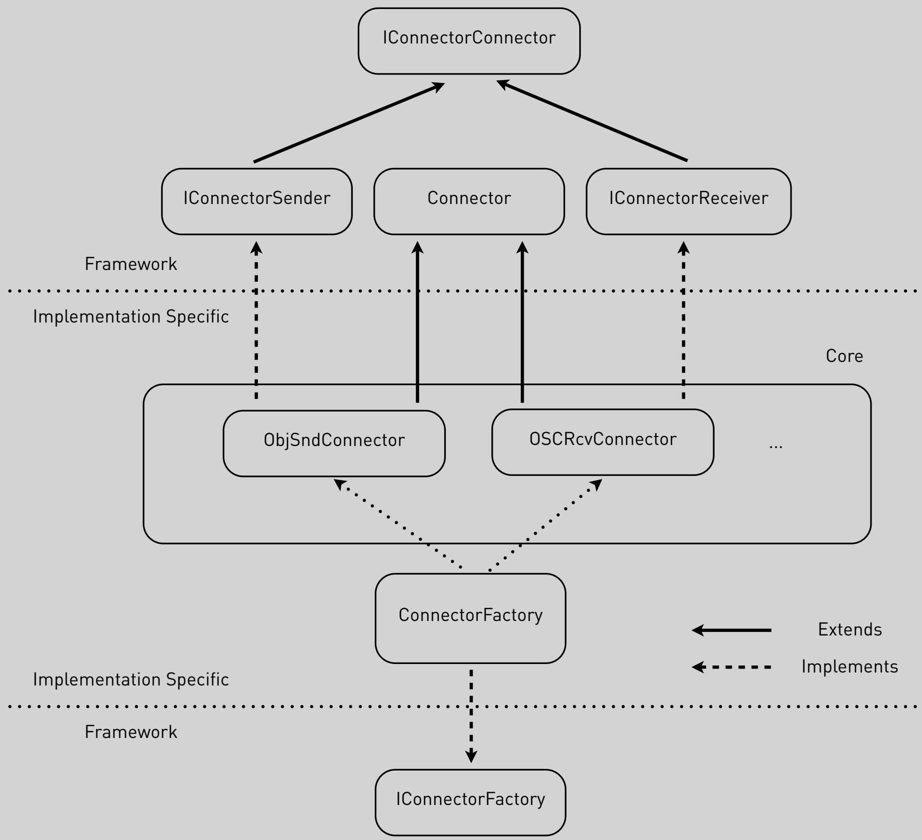


Fig. 6 Joinder components / Connectors

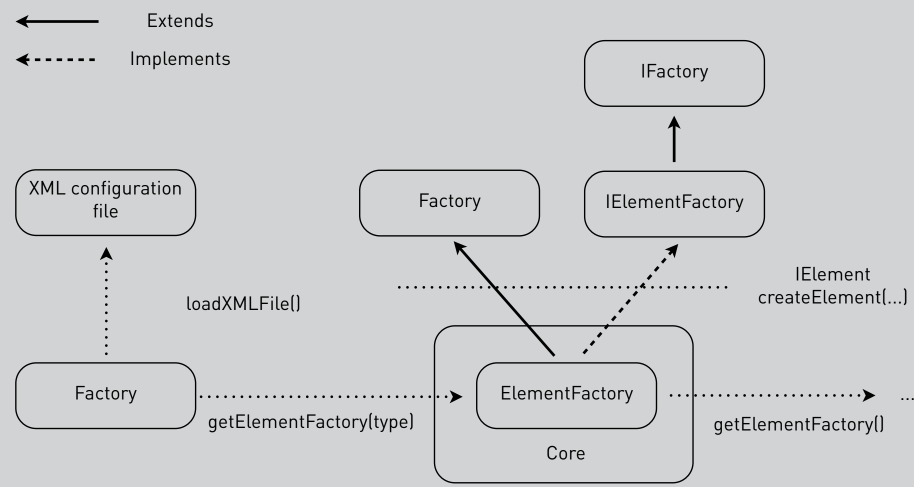


Fig. 7 Joinder components / Instance Factories Ex. Element

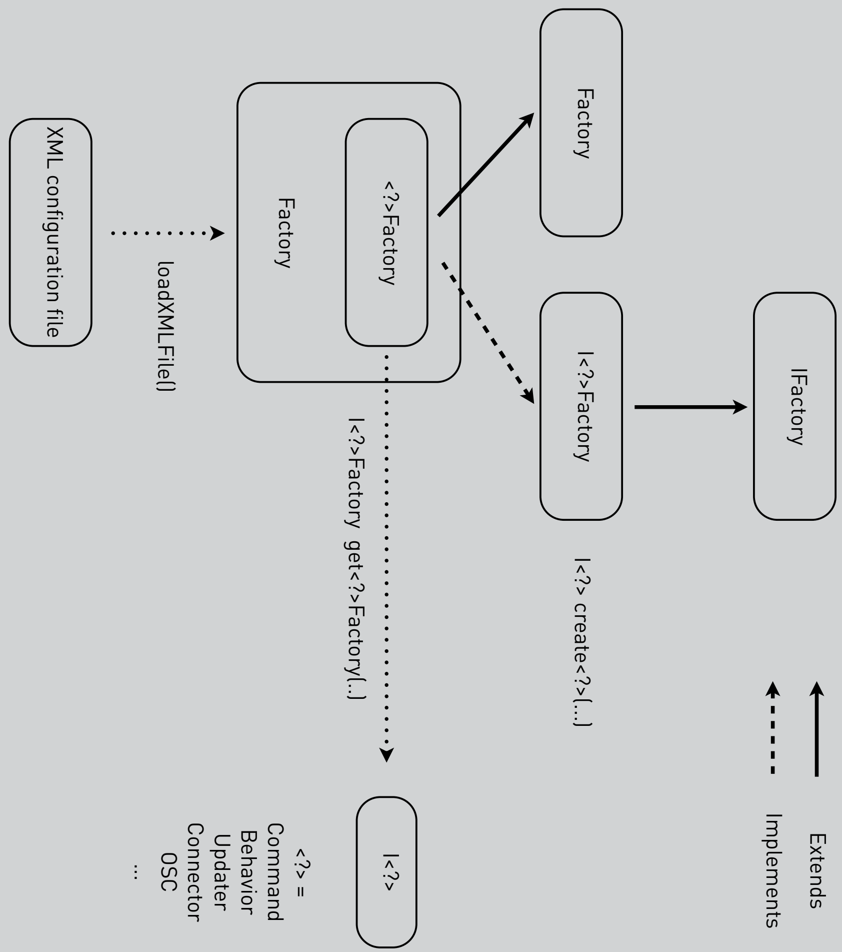


Fig. 8 Joinder components / Static Factories

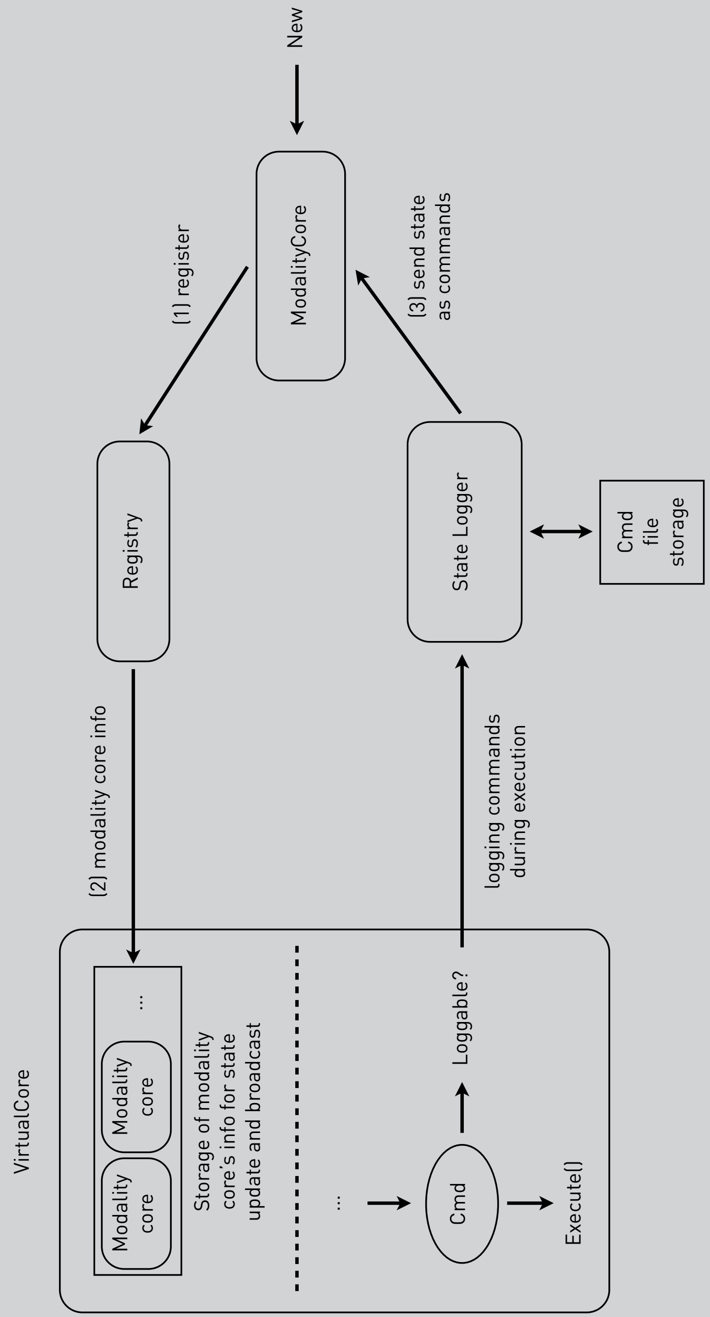


Fig. 9 Joinder components / State Logger

JOINDER

nuno.diniz@gmail.com