Intergestura: a gestural agent based on sonic meditation practices

Kieran Maraj
DisPerSion Lab
York University
Toronto, Canada
kmaraj@yorku.ca

Doug Van Nort
DisPerSion Lab
York University
Toronto, Canada
vannort@yorku.ca

Abstract

This paper presents the *Interdependence Gestural Agent (Intergestura)*, an electroacoustic music-performing multi-agent system whose design is based on sonic meditation principles, adapted to incorporate principles of gestural listening. The *Intergestura* system is comprised of a human performing real-time granular synthesis on a digital drawing tablet, and a pair of software agents that behave according to the rules of a certain sonic meditation piece. The agents behave in a call and response manner, listening for both the physical and sonic gestures of one another and of the human performer. Through this behaviour, performance with the agent affords an experience in which the participant deeply focuses their own attention, awareness and listening, similar to the conditions produced when performing the original text score with a group of human performers.

Introduction

The use of algorithmic processes has changed the way we operate in all areas of musical practice, from making instruments, to composing, performing, improvising, and listening (Magnusson 2019). The line between what can or should be considered an instrument or interactive composition has been blurred. Much contemporary research has focused on endowing machines with creative agency, further complicating these distinctions (Pasquier et al. 2017).

This paper presents the *Interdependence Gestural Agent (Intergestura)* system. *Intergestura* sits at the intersection of interactive composition and improvisational performance system,and facilitates human-machine co-creation and partnership. The system is designed around two key design consideration: i) the sonic meditation works of composer Pauline Oliveros (Oliveros 1974), and ii) an embodied view on machine listening that includes gestural action. While Oliveros' text pieces focus on listening and responding to sonic gestures (Van Nort 2009), *Intergestura* adapts this across modalities to develop a call/response relationship that is based on both sonic and physical gestures, which results in sound-based output. We first discuss relevant prior works before discussing the system in more detail.

Related Work

Sonic Meditations

Composer Pauline Oliveros published her first set of sonic meditations in 1974. The meditations are a highly influential collection of text-based scores meant for performance by musicians and non-musicians alike, and have inspired many subsequent sonic meditation pieces in the decades since (Jensen 2009). Through sounding and listening, these pieces foster diverse modes of training, coordinating and synchronizing attention and awareness (Oliveros 1984). The enhancement and development of aural attention and awareness are some of the explicit goals of these pieces. In performing a sonic meditation, all persons present are meant to take part in the piece - audience is performer and performer is audience. In this way, "Oliveros is more interested in the social, psychological and even physiological aspects of music making than in its product" (Gioti 2020).

The meditations present a structure which engender a meditative engagement with collective sounding and listening – something we build upon by exploring such structures in the context of interactive agents. We also build upon a more recent piece, *dispersion.eLabOrate*, that explicitly augments another of Oliveros' sonic meditations. The *eLabOrate* project features a room-scale ecosystemic augmentation of the *Tuning Meditation* (Hoy and Van Nort 2019). In *eLabOrate*, a group of human participants are joined by a room-scale agent which listens to the collective through a microphone array, analyzes what it hears in software, and then generates sounds according to the meditation instructions. As the original sonic meditations are co-created through a process of blurring sonic boundaries between participants, *eLabOrate* is co-created by and blurs boundaries between human participants and technological system. We expand this research into agent-based augmentations via *Intergestura*, which shifts the focus from environmental interactions towards an embodied, gesture-centric approach to listening and interaction.

Musical Agents and Co-creation

A classic example of an agent-based performance system is George Lewis' 1988 *Voyager* system. *Voyager* is a software system with roots in Black American and African diasporic cultures intended for use in real-time improvisation settings

(Lewis 2000). The system can perform completely on its own or it can perform with other human players, "listening" to them either through a MIDI interface or a machine listening algorithm. *Voyager* is made up of many constituent elements that Lewis conceives of as "players" in an improvising orchestra, but they could equally be thought of as individual agents in a multi-agent system (MAS). Players are grouped together and given tasks, including melody generation, choice of pitches, rhythm, tempo, and interval range, amongst others. Lewis views *Voyager* as its own entity and understand interactivity with it as a process of dialogue. He repeats that the system is non-hierarchical - in performance, it can follow, or act as leader, it can choose to respond or not, just as a human improviser might be faced with making the same set of choices. To play with *Voyager* is to engage in an act of human-machine co-creation.

Intergestura: Overview

The behaviour of *Intergestura* takes its structure from the piece *Interdependence*, from the collection *Four Meditations for Orchestra* (Oliveros 1997). There are two roles for performers in *Interdependence* - sender and receiver. Senders play a short, staccato pitch and receivers respond to short pitches with a short pitch. Performers can switch between these two roles at will. There are three variations on this base call and response structure, as seen in Table 1. Performing *Interdependence* with a group of human performers requires attentive listening and quick responses. In performance one does not know what role other performers are taking at any given time - one must be ready to respond at a moment's notice and as instantaneously as possible. In addition to the heightening of collective sonic awareness, from these simple instructions arise sonic textures that often start out sparse and pointillistic and morph into dense masses of sound. Oliveros notes that "correct player reactions can create an atmosphere of electricity that runs through the ensemble in a rippling effect." Successful performance of the piece requires and results in both awareness - of the overall sound field - and attention - to the inner details of that field and the cues contained within it. It is both the enhancement of collective listening practices and the aesthetic possibilities of the piece that have inspired us to adapt the structure into an agent-based system, blurring the line between interactive instrument/composition/partner, much as the sonic meditation pieces themselves blur the line between audience and performer, composition and meditation exercise. It is important to note that while *Interdependence* structures the agent behaviour, the system is not an attempt to directly recreate the meditation - the system can certainly be used to perform *Interdependence*, but can additionally be steered to perform other forms. We now discuss details of the system design.

System Description

The *Intergestura* system is built in the Max/MSP programming environment. It is comprised of an instrumental aspect in which Wacom and MIDI-based gestural inputs are mapped to granular synthesis-based sound processing and an agent component that responds to and co-performs with

Variation	Respond To	Respond With
0	Short pitch	Short pitch
1	Short pitch	Short pitch or long tone
2	Short pitch or end of long tone	Short pitch or long tone
3	Short pitch or end of long tone	Short pitch, long tone, or long tone with gliss

Table 1. The four variations for receiver behaviour in *Interdependence*. Sender always plays a short pitch, at any time and at any dynamic.

this instrumental system. The system uses gestures as its fundamental unit. References to 'pitch' or 'tone' in the instructions are replaced with 'gesture' in *Intergestura's* structure. For instance, the base version becomes 'respond to a short gesture with a short gesture.' The agents have access to two corpora to structure their behaviour - one is a running memory of human input gestures and the other is a collection of human-segmented gestures, similar to the design of the FILTER system, which draws upon running episodic and semantic performance memories (Van Nort, Oliveros, and Braasch 2013). The agents, like human performers, can choose either role to play in and can switch between these roles, as well as between the different variations.

Gesture

Human input into the system primarily takes place through gestural input with a drawing tablet. Physical gestures on the tablet are captured and through various mapping processes are connected to a granular synthesis module, finally producing an audible sonic gesture. Additional control of the system takes place through a standard MIDI controller.

Gesture is one way that meaning is produced and understood in music-making (Leman 2010). Through physical and sonic gesture, meaning is made directly and indirectly - Leman writes "gesture appears as a mediator for music-driven social interaction or as the vehicle through which a 'me-to-you' relationship is established". This view on gesture is a rich space for thinking about the development of agential systems, building upon the view of action/sound gestures as a point of interaction design discussed in (Van Nort 2009). In *Intergestura*, agent behaviour is based on call and response, listening and reacting. Gesture is how the agents understands their roles, and through gesture the human-machine relationship is established. Via gestural interaction, the agents participate in the co-creative process.

From an instrumental performance system perspective, *Intergestura* is inspired by elements of the *greis* system (Van Nort, Oliveros, and Braasch 2013), such as tablet interaction, parallel granular engines, and semantic and running memory structures. We discuss these components, followed by the agent modules which interact with them.

Synthesis and Mapping Design

Sound in *Intergestura* is produced through a trio of identical granular synthesis modules. The human performer plays one module and the agents play the others. Each module contains a pair of granular engines that run in parallel to each other. Each engine provides a different quality of sonic granulation and the performer can crossfade between engines or have them both sounding simultaneously. Sound sources are grouped into sets according to similar sonic qualities, predetermined and selected by the human performer. In performance these are navigated by the performer, who also selects a given set for each agent, thereby determining the higher-level constraints for their general sound palette. While the agents can choose to randomly switch between sounds in a set, they cannot navigate between different sets.

Gestural control data from stylus actions on the tablet are mapped into the granular module to produce sound. While stylus pressure is always mapped to output volume, there are a number of mapping modes that can be selected and layered. This ranges from direct parameter mappings (e.g. stylus y-position mapped to both grain rate and size, x-position mapped to scrubbing through the sound source) to a self-organizing map (SOM) (Kohonen 1982). In this latter mode, many granular synthesis parameters are two-dimensionally organized across the surface of the tablet and are navigated via stylus coordinates. Scrubbing through the sound source is done through a time-based method inspired by (Van Nort, Wanderley, and Depalle 2014). A final mapping mode makes use of a multilayer perceptron (MLP) (Pal and Mitra 1992) to map stylus coordinates as well as velocity, acceleration, and jerk magnitudes to various granular synthesis parameters. The MLP is trained so that faster, chaotic gestures produce noisier sonic output while slower, smoother gestures produce smoother sonic output.

Agent Modules

Intergestura uses the cognitive concepts of semantic and episodic memory to form the agent's corpus, in combination with a reactive rule-based system, inheriting the instructions from the *Interdependence* meditation. Each form of memory constructs its own corpus. The semantic memory constructs a hybrid corpus, containing control gesture, sonic gesture, and machine analysis of these. The episodic, or running, memory constructs a corpus, consisting only of control gesture data. Semantic memories are sonic gestures that are explicitly segmented (via stylus button) by the human performer, and are used to structure the behaviour of the agent in sender mode. The gesture, containing control data and audio, is added to the semantic memory and analysis performed on the audio component of the gesture. There are up to 10 semantic memories at any point, with new memories taking the place of the oldest memory. The running memory stores only control gestures, which are used by the agent in both modes. The running memory consists of the last 20 gestures that have been made by the human performer. These gestures are segmented automatically - starting when the stylus makes contact with the tablet and ending when it is lifted. Again, new memories replace the oldest memory.

At the beginning of sender mode behaviour, the agent randomly selects one of the gestures from the bank of semantic memories. This gesture is used to structure the sender behaviour and generate a sequence of gesture playback triggers. The agent looks at the onsets detected in the analysis of the sonic component of the gesture to create a sequence of triggers and determines how fast to step through that sequence. Next, a random gesture is chosen from the running memory. The agent then begins to step through the sequence and at each trigger plays the control data from the chosen running memory into its granular module. After playing through the sequence, the agent makes a choice on whether to choose a new semantic memory and repeat the sender process, or to enter receiver mode.

On entering receiver mode, the agent randomly selects a number of gestures to respond to. It listens for the ends of incoming gestures - from the human performer and from the other agent. It specifically waits for a message that the stylus has stopped making contact with the tablet, whether from the human performer directly or from the other agents' played back gesture. On lift of the stylus, the agent decides to respond or not. In variations 0 and 1, the agent only responds to short input gestures. In variations 2 and 3 the agent will respond to input gestures of any length. If it chooses to respond, the agent selects a random gesture from the running memory. The agent will temporally compress or stretch the recalled gesture, based on the rules of the current variation. The third variation also gives the agent the option to add random, narrow-interval glissandi to the played back gestures. It plays back the transformed gesture into its granular synthesis module and returns to listening. After playing back its predetermined number of gestures it has the option to remain in receiver mode or return to sender mode. If the agent chooses to remain in receiver mode, it also selects which variation it will perform and returns to the top of the task.

Each agent has some autonomy over the sound source selected in the granular module. While it is constrained to the folder selected by the human, each time the agent receives an input gesture, it also looks at the standard deviation of the velocity of that gesture. If above a threshold, the agent will randomly select a new sound source from its selected folder.

Meta-Controls

Each agent has a set of meta-controls that can be engaged with by the human performer - a probability of role (sender or receiver), a probability of variation, and an on/off switch. The first meta-control influences the likelihood of the agent staying or switching between sender and receiver modes. At either extreme, it will remain in the respective mode until the control is changed. The second meta-control maps to parameters that exclusively influence the behaviour of the agent in receiver mode. Towards one end, the agent will be more likely to play variations 0 or 1 and towards the other, it will be more likely to play variations 2 or 3. At the centre, it is equally likely to select any of the variations. The final control is a button that enables or disables the agent.

Discussion

From our experience in performing this sonic meditation numerous times in diverse ensembles, attempting to play *Interdependence* with the *Intergestura* system feels similar to performing with a group of humans, but responses feel immediately instantaneous - in performing with humans there is some build up as performers attain focus and attention, but with the digital system the immediacy of response is there from the start. The pair of agents are highly responsive and playing with them feels like interacting with mirrored versions of oneself: recently played sonic gestures return but are transformed in some fashion by the agent's listening/sounding decision process. In performing sonic meditations with a group of humans, the boundaries between one's own actions and those of fellow performers become perceptually blurred - it can become challenging to tell where one's own actions end and an other's begin. Similarly, performing with *Intergestura* blurs the line between oneself and the agents. In a back and forth of many quick gestures in succession, it can be difficult to tell which gestures the human produced and which came as reflections from the system.

This quick back and forth style of playing often results in the feeling of rippling electricity described by Oliveros, especially when the agents are playing in variations 0 or 1. Furthermore, by interacting with the meta-controls, it is possible for the human performer to play with the system but move away from directly playing *Interdependence*. For example, one could choose to have only a single agent active, have it constantly in receiver mode, and constantly playing variation 2. In this state, the agent would respond to every human input gesture, acting as an augmenter of performance and allowing for sounds and shapes not possible for the human performer alone. Through these meta-controls the human performer can choose to perform *Interdependence* explicitly or to morph to other styles of playing.

Playing with *Intergestura* is a process of human-machine co-creation. (Brown 2012) and (Ravikumar, McGee, and Wyse 2018) argue for human-machine systems that embody co-creativity through partnership and shared tasks. Brown argues that metacreative systems should be conceived of as partnerships with generative systems, with human participants being viewed as components of the overall system, as opposed to viewing the system simply as a tool or autonomous machine. Ravikumar et al. additionally argue that co-creation should happen through a process of co-experience, where co-creation happens between participants who experience togetherness by working towards a shared goal. Performing with *Intergestura* embodies this notion of co-experience. Whether playing *Interdependence* directly or not, to engage with the system - becoming part of the larger human/machine system - is to engage in a partnership with the agents. It is clear that the agents on their own cannot take part in a process of creation, and the possibilities of what the human can achieve without the agents are greatly reduced. The human and agents come together to perform a shared instrumental system and to create the system in its totality. The system and partnership affords an experience of playing together, working towards a common task, and co-creating. The network of listening and sounding, grounded in simple rules of attention and action, that is fostered by sonic meditation style pieces is a powerful framework for emergent human/machine co-creativity - it is greater than the relatively simple rule sets the pieces (and agents) are based on.

Future Work and Conclusion

We are currently developing an alternate version of this system in which all interaction takes place through sound. As in dispersion.eLabOrate, our goal is for a system which can be used to augment traditional sonic meditation practices involving multiple participants. The system will be able to perform within an ensemble of humans with diverse instrumentation and engage in co-creation purely through sound, opening up interaction with the system to a wider range of people. This requires a different strategy than the current 'gestural listening' approach, and we intend to evaluate the implications of this design change in a future comparative study.

While it is changing, the paradigm of co-creation still stands in contrast to dominant conceptions of human-machine interaction which often view computers as tools to be used. We believe it is important to deeply explore new forms of interaction afforded by digital systems as they may lead to new ways of creating, thinking, and being in the world. *Intergestura* is one approach towards this larger goal, drawing on novel forms of human-human co-creation that we believe offer a rich space of possibility in the context of computational creativity for musical performance.

Author Contributions

K. Maraj developed the system and lead the writing of the manuscript. D. Van Nort supervised the project and contributed to system design, scholarly contextualization and paper revisions.

Acknowledgments

This research received support from the Social Sciences and Humanities Research Council, the Ontario Early Researcher Award and the Canada Research Chairs Program.

References

Brown, A. R. 2012. Creative partnerships with technology: How creativity is enhanced through interactions with generative computational systems. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Gioti, A.-M. 2020. From artificial to extended intelligence in music composition. *Organised Sound* 25(1):25–32.

Hoy, R., and Van Nort, D. 2019. An ecosystemic approach to augmenting sonic meditation practices. In *14th Int. Symposium on Computer Music Multidisciplinary Research*, 318.

Jensen, M., ed. 2009. *Deep listening anthology: Scores from the community of deep listeners*. Deep Listening Publications.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics* 43(1):59–69.

Leman, M. 2010. Music, gesture, and the formation of embodied meaning. In *Musical Gestures*. Routledge. 138–165.

Lewis, G. E. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* 10:33–39.

Magnusson, T. 2019. *Sonic writing: technologies of material, symbolic, and signal inscriptions*. Bloomsbury Academic.

Oliveros, P. 1974. *Sonic Meditations: March - November 1971*. Smith Publications.

Oliveros, P. 1984. *Software for people: Collected writings 1963-80*. Station Hill Press.

Oliveros, P. 1997. Four meditations for orchestra.

Pal, S., and Mitra, S. 1992. Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks* 3(5):683–697.

Pasquier, P.; Eigenfeldt, A.; Bown, O.; and Dubnov, S. 2017. An introduction to musical metacreation. *Computers in Entertainment (CIE)* 14(2):1–14.

Ravikumar, P. T.; McGee, K.; and Wyse, L. 2018. Back to the experiences: empirically grounding the development of musical co-creative partners in co-experiences. In *6th Int. Workshop on Musical Metacreation. 9th Int. Conference on Computational Creativity, ICCC*, 1–7.

Van Nort, D.; Oliveros, P.; and Braasch, J. 2013. Electro/acoustic improvisation and deeply listening machines. *Journal of New Music Research* 42(4):303–324.

Van Nort, D.; Wanderley, M. M.; and Depalle, P. 2014. Mapping control structures for sound synthesis: Functional and topological perspectives. *Computer Music Journal* 38(3):6–22.

Van Nort, D. 2009. Instrumental listening: sonic gesture as design principle. *Organised sound* 14(2):177–187.