PERFORMING VOICES IN IMMERSIVE THEATER: KINESTHESIS 2.0 FOR EMBODIED PERFORMANCE

Tilemachos MOUSSAS (Τηλέμαχος Μούσας)[1], **Natalia KOTSANI** (Ναταλία Κωτσάνη)[2], and **Anastasia GEORGAKI** (Αναστασία Γεωργάκη)[1]

[1]**National and Kapodistrian University of Athens**, Athens, Greece
[2]**National and Technical University of Athens**, Athens, Greece

ABSTRACT

Kinesthesis 2.0 is an innovative interactive tool designed for augmenting the voice of singers and actors, enabling them to dynamically modify their voice in real-time during music theatre performances. This tool enhances immersive theatre by integrating technology and artistic expression, allowing for an interplay between verbal and gestural elements. Kinesthesis 2.0 focuses on achieving embodiment, where actors use multi-sensory stimuli, integrating voice, facial expressions, and gestures to create a comprehensive portrayal of characters. This embodiment is facilitated by the tool's ability to capture and translate these modulations, offering a dynamic interaction between auditory and visual storytelling.

Technically, Kinesthesis 2.0 uses Python, Open Sound Control protocol, and Max/MSP to map the facial movements into predefined gestures. MediaPipe is used for face and hand detection, while OpenCV processes the webcam feed, capturing head movements, facial expressions, and hand gestures. Through the gesture and facial expressions detector, the artist controls various effects (pitch shifter, multi-channel vocoder, reverb, multi-tap delay and reverse delay). Artistically, Kinesthesis 2.0 enriches the expressive capabilities of performers by amplifying and augmenting their voice, thus enhancing their ability to convey emotions and narratives. As a case study, the tool was used by the singers and actors in the theatre performance "Rayman Scream".

1. INTRODUCTION

Kinesthesis 2.0 combines a facial expressions and a hand gesture detection system developed in Python using Mediapipe with a four-voice vocoder, a pitch-shifter, a reverb, a multi-tap delay and a reverse delay developed using MAX/MSP. By tracking movements and using them to control the parameters of the vocoder and delay effects, a performer can create expressive vocal effects combined with a digital choir, by moving face, head or hands in specific ways, creating complex vocal timbres and rhythms. Kinesthesis 2.0 was tested and used during the rehearsals from the composer, the actors and singers of the theatre play "Rayman Scream", since December 2023.

By capturing bodily movements, Kinesthesis 2.0 facilitates a form of embodiment, wherein the performer's bodily gestures and facial expressions directly influence the auditory output, using different sensory modalities. This integration of gesture, technology, and artistic expression aims to immerse the audience in a multisensory performance, where the interplay between physical movement and vocalization enriches the overall narrative experience.

2. RELATED RESEARCH

Gesture is a part of human communication, and it has been used as a form of non-verbal communication. It is a fundamental component of language that contributes meaningful and unique information to a spoken message and reflects the speaker's underlying knowledge and experiences. Theoretical perspectives of speech and gesture propose that they share a common conceptual origin and have a tightly integrated relationship, overlapping in time, meaning, and function to enrich the communicative context [1].

Gesture as an expressive tool has been a fundamental aspect of theatricality throughout history and constitutes a codified theatrical language performed not simply through speech [2]. Gestuality, is an extension of natural human behaviors, with actors using their bodies to convey meaning and emotion to their audiences. Gestures can be used to communicate a wide range of emotions and ideas, from joy and excitement to fear and sorrow, and they can be used to convey complex concepts and ideas that may be difficult to express through words alone.

For example on the southern coast of India, the Kathakali ritual theatre hand gestures called Mudras [3], are central to the performance. Divided into two primary components, Cholliyattam and Elakiyattam, the narrative unfolds through intricate gestures, body movements, and facial expressions. The significance of eyes in expressing emotions (Oculesics in Sathika Abhinaya) is emphasized, with early morning eye exercises shaping the actor's ability to convey emotional states. In Cholliyattam, enacted verses are brought to life through a sophisticated blend of Kinesics, Gait, Postures, and Gestures, complemented by the rhythmic repetition of sung verses. Elakiyattam, the silent pinnacle of Kathakali, showcases the actor's creative freedom to communicate without singing, utilizing a combination of non-verbal elements such as Kinesics, Gait,

Postures, Gestures, and Oculesics, along with characteristic grunts. These Mudras, with their precise and elaborate movements, convey profound meanings, emotions, and ideas.

Actors are using a series of movements and poses to convey meaning and emotion to their audiences. While some gestures are traditional schemata, there are also "less obvious" gestures that are just as important in conveying meaning on stage. These everyday gestures can be found at the heart of a scene, revealing an actor's character and emotions in subtle ways. In a study devoted to physical contact in tragedy, Maarit Kaimio [4] explores these less obvious gestures, including the gesture of "nursing" that appears in a scene from Euripides' Orestes. This gesture, which may seem insignificant at first glance, reveals much about the characters and their relationships.

In the 20th century, the use of gesture in theater underwent another significant evolution, as Jerzy Grotowski, Eugenio Barba, and Peter Brook explored new ways of using gesture to create powerful and immersive theatrical experiences. Grotowski, developed a technique known as "poor theater," highlighting the use of the body and voice to create intense and emotionally charged performances. One of the most interesting developments in recent years has been the use of motion-capture technology in theater, which allows actors to create complex and realistic movements that can be projected onto a screen or other digital medium [5].

In achieving expressive electroacoustic music, a deep understanding of the morphological relationships between physical gestures and their corresponding sonic outcomes is necessary. When a morphological relationship is deterministic, it means that there is a consistent and logical correspondence between the performer's intentions, expressive gestures, and the resulting sound. For example, a gentle short gesture of our fingers could produce a subtle and intimate sound, whereas a violent movement could create a more dramatic and intense sound. Research has shown that audiences heavily rely on physical gestures to understand the expressive intent of a performance [6] and that electronic music performances ought to demonstrate a coherent correlation between the magnitude of a control gesture and the ensuing acoustic outcome. [7].

3. MULTIMODAL INPUTS AND AUDIENCE PERCEPTION

In immersive theatre, "immersive" not only means "a state of being immersed", but also "providing information or stimulation for a number of senses". There are many types of sensory stimulation, including visual, aural, and tactile [8]. Kinesthesis 2.0 creates an immersive and multisensory experience for both performers and audiences. Research in multi-sensory processing and cross-sensory cognitive modes has shown how the brain constructs perceptions by combining congruent inputs and segregating incongruent ones. These studies have explored terms like "synesthetic association" or "cross-modal correspondence" to describe how different sensory dimensions correspond to each other [9].

For example, synesthetic correspondence involves non-redundant sensory dimensions, such as the association between sound frequency and brightness. By contrast, cross-modal correspondences include both synesthetic associations and those between redundant stimuli perceived through different sensory modalities, such as the auditory and visual length of an event [10]. Spectators are not merely listeners, they become observers of a synchronized dance where gestures become audible, creating an immersive encounter. Music perception is achieved through interaction with other sensory fields, as a result, the audience can activate memories, images and feelings [11]. An immersive experience can manifest in different states: "engagement, engrossment, and total immersion," each representing a gradient of immersive experiences. Interactivity and immersion are closely related in performance, interactivity can enhance immersive experiences by fostering deeper audience involvement [12, 13]. By applying sonic interaction design, this tool aims to explore the methods for engaging audiences with immersive experiences by extending the interactivity to the auditory environment.

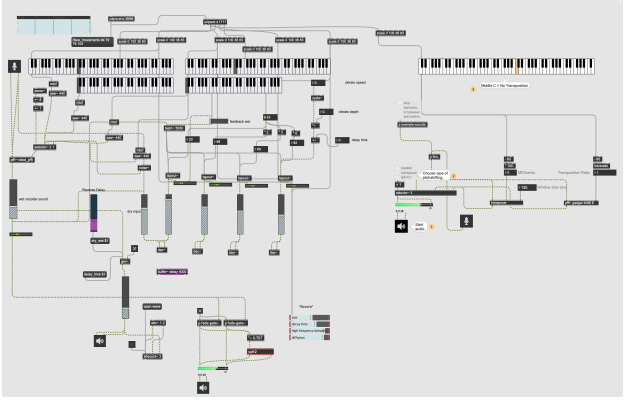
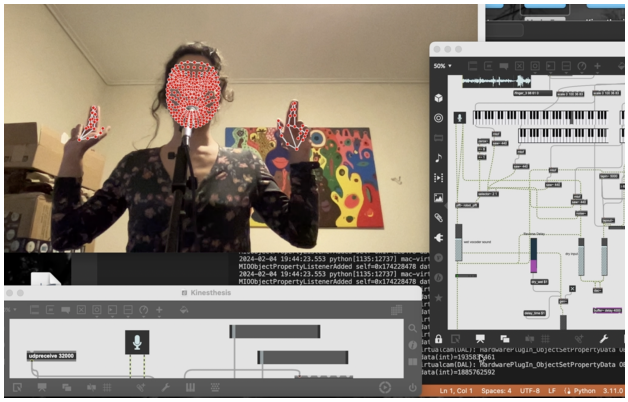


Figure 1. Kinesthesis 2.0 Max/MSP Patch

4. ARCHITECTURE OF THE SYSTEM KINESTHESIS 2.0

The Kinesthesis 2.0 simple interface, through the Max/MSP patch, offers various pitch-shifting options, including transposer, gizmo or bypass. Additionally, the user can choose a specific type of a MIDI instrument to track the voice, choosing between "harmonic", "in-between" or "percussive" mode. In the Kinesthesis 2.0 interface, the user can control through various parameters the vocal effects in real-time. This includes a four-voice harmony through four individual vocoders, assigning each vocoder to a specific voice within the harmony and modulating their parameters accordingly. The interface offers also various preset options for reverse delay customization (for the reproduction of recorded phrases in reverse playback), a multi tap delay and a parametric reverb effect.

The real-time control of the sound parameters is achieved by using hand gestures and facial expressions. The OSC (Open Sound Control) messages for facial movements, such as head tilts, eye movements, and mouth gestures (mapped to predefined gestures such as 'One' 'Two'



A Mezzo Soprano using Kinesthesis 2.0

'Three' 'Four' and 'Five') are sent to an IP address and port. The gestures are continuously analyzed and when a match is found, the corresponding OSC message is dispatched. The Max/MSP patch receives OSC messages transmitted over a network using the User Datagram Protocol (UDP). Subsequently, the received list of data is broken down into individual messages, with each message corresponding to one of the parameters intended for processing. This allows us to dynamically control various aspects, such as the pitch of each vocoder, the percentage of delay effect (specifically, multi-tap and reverse delay), as well as the size, decay time, high-frequency damping and diffusion of the reverb.

The Python code and the MAX/MSP patches of Kinesthesis 2.0 can be found in a public repository [1].

5. AESTHETIC ISSUES AND CONCLUSIONS

Kinesthesis 2.0 was used for the first time in the theatre performance "Reyman Scream" (March 2024, Theatre "Sfendoni", Athens), as a vocal augmentation system, offering to the performer the ability to improvise and composing music in real-time. On the rehearsals, the theatrical ensemble has used Kinesthesis 2.0 real-time voice automation and expressive features to dynamically modulate the soprano singer's and the main actor's vocal output, creating a harmonious blend and simulating a digital choir, synchronizing vocal elements with dramatic expressions. This integration demonstrated Kinesthesis 2.0 functionality and enriched the collaborative dynamics in the context of a theatre performance. A demo from the theatre play rehearsals is uploaded in a public repository [2].

One of the most challenging aspects of the Kinesthesis 2.0 interface is the use of real-time vocal input. While audiences generally respond positively to these moments, there are some issues that need to be addressed in order to optimize the use of this feature. For example, audiences may have a preconceived notion that "singing" will occur when the performer is using the microphone. This can lead to disappointment when the performer focuses on gesture control without live vocal input. To address this issue, performers can be trained to use vocal input more effectively and to communicate with the audience about what they are doing.

Another issue with real-time vocal input is that it can become difficult for audiences to understand what the vocalist is doing when the composition becomes dense. This is because the voice is traditionally perceived as a solo instrument, which should be heard above a background texture. The microphone is a key component of the Kinesthesis 2.0 interface, but it can also be a source of confusion for audiences. The microphone directs focus towards the spatial positioning of the voice and the organization of surrounding auditory elements. Nevertheless, within the Kinesthesis 2.0 interface, the function of the live voice deviates from established contextual models. In this framework, the vocal element can undergo substantial transformations, resulting in dense textural accompaniments, and the degree of processing may reach an intensity wherein the original voice becomes challenging to discern. To address this issue, composers can create contrasting sections of the composition where the texture becomes more minimal and the connections between gesture and sonic outcome are more transparent.

Furthermore, it is important to consider the potential ethical implications of using the Kinesthesis 2.0 interface in a performance context. For example, if the Kinesthesis 2.0 interface is used to process and manipulate the performer's voice in real-time, it could be argued that this violates the performer's right to control their own artistic expression. It is important to address these concerns and to ensure that the use of the Kinesthesis 2.0 interface is transparent and respectful of the performer's autonomy.

Music perception is achieved through interaction with other sensory fields. Musical meaning might be constructed largely by synesthetic processes, where the sensory associations from sound activate memories, images, and emotions [9]. As feedback is gathered and new technologies and techniques are developed, the interface will continue to evolve and improve. By embracing a spirit of experimentation and continuous improvement, the Kinesthesis 2.0 interface has the potential to revolutionize the way that vocal performances are created and experienced.

Figure 2. Performing with Kinesthesis 2.0

[1] https://github.com/Tilemachos88/Kinesthesis-2.0

[2] https://drive.google.com/file/d/1lMhyEjLZOvChCnfAloLseboFkz4Sjzu9/view?usp=sharing

6. REFERENCES

[1] S. Clough and M. C. Duff, "The role of gesture in communication and cognition: Implications for understanding and treating neurogenic communication disorders," *Frontiers in Human Neuroscience*, vol. 14, p. 323, 2020.

[2] H. Roisman, *The encyclopedia of Greek tragedy*, 2014.

[3] E. Barba and S. Sanzenbach, "The kathakali theatre," *Tulane Drama Review*, vol. 11, no. 4, pp. 37–50, 1967.

[4] M. Kaimio, "Comic violence in aristophanes," *Arctos–Acta Philologica Fennica*, vol. 24, pp. 47–72, 1990.

[5] J. Grotowski *et al.*, "Towards a poor theatre, ed," *Eugenio Barba (Copenhagen: Odin Teatrets Forlag*, 1968.

[6] G. Emerson and H. Egermann, "Gesture-sound causality from the audience's perspective: Investigating the aesthetic experience of performances with digital musical instruments." *Psychology of Aesthetics, Creativity, and the Arts*, vol. 12, no. 1, p. 96, 2018.

[7] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," *Computer music journal*, vol. 26, no. 3, pp. 11–22, 2002.

[8] J. Machon, *Immersive theatres: Intimacy and immediacy in contemporary performance*. Bloomsbury Publishing, 2017.

[9] G. F. F. Bragança, J. G. M. Fonseca, and P. Caramelli, "Synesthesia and music perception," *Dementia & neuropsychologia*, vol. 9, pp. 16–23, 2015.

[10] N. Bien, S. Ten Oever, R. Goebel, and A. T. Sack, "The sound of size: crossmodal binding in pitch-size synesthesia: a combined tms, eeg and psychophysics study," *NeuroImage*, vol. 59, no. 1, pp. 663–672, 2012.

[11] J. S. Snyder, C. M. Schwiedrzik, A. D. Vitela, and L. Melloni, "How previous experience shapes perception in different sensory modalities," *Frontiers in human neuroscience*, vol. 9, p. 594, 2015.

[12] E. Brown and P. Cairns, "A grounded investigation of game immersion," in *CHI'04 extended abstracts on Human factors in computing systems*, 2004, pp. 1297–1300.

[13] N. Mount, C. Chambers, D. Weaver, and G. Priestnall, "Learner immersion engagement in the 3d virtual world: principles emerging from the delve project," *Innovation in Teaching and Learning in Information and Computer Sciences*, vol. 8, no. 3, pp. 40–55, 2009.