

Transhuman Ansambl - Voice Beyond Language

Lucija Ivšić
Sensilab, Monash University
900 Dandenong Road
Caulfield, Victoria
Lucija.Ivšić@monash.edu

Jon McCormack
Sensilab, Monash University
900 Dandenong Road
Caulfield, Victoria
Jon.McCormack@monash.edu

Vince Dziekan
Sensilab, Monash University
900 Dandenong Road
Caulfield, Victoria
Vince.Dziekan@monash.edu

ABSTRACT

In this paper we present the design and development of the *Transhuman Ansambl*, a novel interactive singing-voice interface which senses its environment and responds to vocal input with vocalisations using human voice. Designed for live performance with a human performer and as a standalone sound installation, the *ansambl* consists of sixteen bespoke virtual singers arranged in a circle. When performing live, the virtual singers listen to the human performer and respond to their singing by reading pitch, intonation and volume cues. In a standalone sound installation mode, singers use ultrasonic distance sensors to sense audience presence. Developed as part of the 1st author's practice-based PhD and artistic practice as a live performer, this work employs the *singing-voice* to explore voice interactions in HCI beyond language, and innovative ways of live performing. How is technology supporting the effect of intimacy produced through voice? Does the act of surrounding the audience with responsive virtual singers challenge the traditional roles of performer-listener? To answer these questions, we draw upon the 1st author's experience with the system, and the interdisciplinary field of voice studies that consider the voice as the sound medium independent of language, capable of enacting a reciprocal connection between bodies.

Author Keywords

Human-computer interaction, Voice studies, Singing-voice, Agency, Musical interface, Live performance

CCS Concepts

•Applied computing → Sound and music computing; Performing arts; •Human-centered computing → Sound-based input / output; •Information systems → Music retrieval;

1. INTRODUCTION

Given the universality of the human voice, and its musical and en masse creative potential, it has long been of interest

in human-computer interaction [9, 5, 19] and at NIME [12]. This significance of voice interaction reveals itself wherever screen-based interaction is ineffective or limited. Recent advancements in technologies such as voice recognition, speech synthesis, digital signal processing (DSP), machine learning (ML) and artificial intelligence (AI), are at the core of popular tools such as Amazon Alexa, Apple's Siri, and Google Home. Yet, they all understandably use language-based voice interaction. On the other hand, although less present in the field of more experimental and expressive interactive works, even there language-based voice interaction still takes primacy [6].

Rather than analysing or conceiving the voice as the mere language carrier, and therefore using it for clear commands to receive feedback, this paper looks at the alternative yet numerous poetic and expressive possibilities of the *singing-voice*. More specifically, while the technological choir performs songs that contain lyrics, the focus of our investigation is directed towards the singing-voice as a medium in itself along with the spatiality of the work and how it enhances immersion in the act of listening and feeling the voice, both for the performer and the audience. By surrounding the human performer and the audience with an artificial choir, our investigation extends beyond the solely human voice, considering the "voices" of machines with their own autonomy, exploring contemporary ways of live performance.

Our work is influenced by recent philosophical, scientific and critical studies, particularly Voice studies that focus on the voice as a medium in itself, outlined further in the next section. Section 2 offers a brief review and showcase of voice in interactive art and performance, accompanied by relevant examples that inspired our work. Then we present the *Transhuman Ansambl*, its system, concept, technical features and modes of interactions. In Sections 3.3 and 3.4 we demonstrate the work in two complementary modes – as a collective for live performance and a standalone interactive sound installation (See also Making of video). We briefly address the cultural context and overall significance of the intangible cultural heritage preservation realm that this work implies. Finally, in Section 6, we analyse our observations and propose prospective applications.

2. BACKGROUND AND RELATED WORK

2.1 Voice Studies

The continuous development of sophisticated AI and ML systems, along with their ever-increasing presence in all aspects of our lives, urges us to reconsider the relationships we develop with non-human entities. The affect of performativity carried through the contemporary human voice now extends beyond the sole exclusivity of the body. This con-



Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

NIME'24, 4–6 September, Utrecht, The Netherlands.

cept of performativity, a term that initially arose within J.L. Austin’s language philosophy, was powerfully introduced into feminist theory e.g [4] and applied to help think about each of those aspects as ongoing processes, rather than fixed and non-performative representations. Performativity is essential to a voice that evokes and moves from and between bodies and things – it is this performative aspect of voice that brings affect to the foreground and establishes emotions, connectivity, and attachments between individuals and things. Hence, understanding voice and its performance in technology and the arts within a posthumanist context that de-centres humans is driving the rapidly developing interdisciplinary field of *voice studies* [21].

In this section, we outline a small but representative number of projects selected primarily due to their approach to voice and their direct impact on our work. While these works differ from each other in concepts and ways of working with voice, their common denominator is questioning the material qualities of voice through artistic exploration that goes well beyond words. Rather than analysing or conceiving the voice as the mere language carrier, and therefore using it to receive feedback and introduce interactivity, these works utilise the sculptural and spatial possibilities of the voice. The works that follow are not limited to their way of presenting themselves to the audience – the list includes interactive musical interfaces, installations, and performances – the only limitation, if it can be called as such, is their emphasis on the singing-voice that invokes emotions.

One of the first works that drew our attention due to its explicit and reciprocal relation to voice is Alvin Lucier’s renowned work *I Am Sitting in a Room* (1970). Lucier recorded his own voice in a reverberant room, played this original recording in the same room to achieve another recording with the same resonances, repeating the procedure until the original voice was covered by the resonance and reverberation produced by the room. Although seemingly not resembling our *Transhuman Ansambl* which is among other things performed live, what they have in common is treating the voice as a pure sound material. More specifically, while both works contain text/lyrics, when played back in a spatial environment, the semantic value of the language either loses its shape entirely, turning into a textural soundscape as in the case of Lucier’s work, or falls into the background as is the case with our work.

When spatialised and expanded into a surround environment in the form of installations, the voice (as a sound medium) starts to interplay with sculpture [14]. *The Forty Part Motet* is one of the acclaimed examples of a sound installation, consisting of forty loudspeakers playing back a recording of Spem in Alium (1573) by Thomas Tallis. While going through available audio-video documentation of the work, we have noticed how despite being generated in simple loudspeakers, voices used in Cardiff’s work retained the ability to affect viewers and evoke emotions.

Further into the development of our work, attempting to answer questions regarding the interface between the human performer and the virtual singers, we narrowed the search by looking only at the works that use a microphone as an interface. One such pioneering work is the *Singing Tree*, created by a team of researchers at MIT in the late 90s. Designed both for a personal interactive experience and as part of a larger project titled “Brain Opera”, the participant interacts with the *Singing Tree* by singing into a microphone [17]. During singing, the system analyses the voice for parameters such as the pitch and rewards the participant with audio and visual feedback if they have maintained a steady pitch. The more recent musical interface developed by Stefano Fasciani and Lonce Wyse uses vocal gestures as an

alternative to traditional, physical controllers. Their voice interface mapped the dynamic aspects of vocal sound to the synthesizer’s parameters [8].

3. TRANSHUMAN ANSAMBL

“...everyone is a listener of others and a performer to others.”

— Atau Tanaka, quoted in [20]

Transhuman Ansambl is an interactive technological choir consisting of 16 non-human autonomous entities called *virtual singers* that sense their environment and respond to a human artist’s voice and audience presence using vocalisations (explained in detail in Section 3.2.3). Each one of these 16 presence-sensitive singers is an individual, self-contained agent: a small and delicate object with an embedded loudspeaker, ultrasonic distance sensor, LED ring, and custom electronics developed specifically for the project. Inspired by the traditional Croatian ring-shaped social formations such as *kolo*, which are defined by a strong sense of shared identity, and *bećarac*, an emergent minimal vocalic chant performed as an interaction between an individual and the group, the virtual singers are circularly distributed around the performance space (Fig. 2). Designed for both live performance and as a standalone sound installation, the audience is always invited to experience the work from inside the ring, surrounded by the singers (Fig. 6).

Conceptually, the work re-imagines the first author’s work as a musician, lead vocalist and performer through technology, while exploring the multiplicity of her multicultural identity (a Croatian woman now living in Australia). Given her extensive practice as a singer, the work is centered around the singing-voice – considering both the material qualities embedded in the physical voice, so much as its role in the identity-making process. Stemming from the advent of migration from Croatia to Australia, traditional Croatian songs were used as starting points for creative departure and development. Although these initial musical pieces we have created with the virtual singers borrow vocal techniques found in Croatian folklore, such as the two-part singing and call-and-response method [15], they are original compositions rather than a prescribed performance of a renowned traditional piece. Rather than introducing additional disjuncture from tradition, these hybrid musical compositions allow the first author to create space for artistic subjectivity and potentially new aesthetics.



Figure 1: Standalone installation version of *Transhuman Ansambl*, with people interacting with the virtual singers. Melbourne, January 2024.

3.1 Cultural context and Ethnographic Study

To obtain a more genuine understanding and richer knowledge about *kolo* and *bećarac*, throughout June 2023, the 1st author carried out ethnographic research that focused on observing two official associations dedicated to the preservation and performance of traditional Croatian folklore. During the ethnographic study, several active performers and both art directors (also acting leaders of the ensembles and once performers) were interviewed individually. The data acquired from these interviews, as well as the accompanying audio-visual observations of the groups during their rehearsals mostly consisted of information about the group's dynamic, interactions before and during performance, and level of individual autonomy.

Conversations with the artistic directors revolved around the processes of defining and appointing roles to each of the performers within the group, degrees of their autonomy, and definitions of the *voice* within the group - both on an individual and group level. What are the main characteristics of a successful traditional folklore performance group? Is there a feature such as a group's lead singer, and if so, how is it defined and appointed?

On the other hand, questions directed to individual performers focused on ways they interact with the audience, among themselves, and also if there is any room for individuality. What was apparent as a shared element, and as such considered a key observation extracted from both the performers' and artistic directors' answers is the highly desired uniformity, in both the aural and visual aspects of the performance.

Expanding on previous research from the field of Intangible Cultural Heritage (ICH), structural elements such as the layout and order of the performers, as well as the reasons to sometimes form smaller, internal formations (duets or triplets), were then used as a basis to constitute system stories for the *virtual singers* in the *ansambl*, as explained in the next section.

Consisting of the immaterial expressions of one's culture, ICH serves to help constitute the cultural identity of its creators [13]. Continuously living and evolving, this heritage is based on collective memory and is a powerful device for telling stories and reaffirming one's identity [10]. Nowadays, new technologies represent a vehicle for the safeguarding and transmission of ICH and as such are yet to be fully explored and implemented through practice [1].

3.2 Virtual Singers

3.2.1 Conceptual definition

Acting as key entities in the artificial life of the *ansambl*, ones that *react* to their environment, the audience, and the performer, virtual singers are considered to be *agents*. The notion of agency can be defined through the mere technical lens as the capability of perceiving the environment through sensors [18], and responding in a timely fashion to changes that transpire – something virtual singers are capable of doing. To be more precise, in most artificial life systems, what constitutes an agent are two important features: i) level of autonomy which implies the use of sensors and constitutes the relation between perception and response, and ii) level of adaption to changes in the surrounding environment [2]. For virtual singers, this structure is apparent through the following features:

- *Stereo pairs*. Traditionally, within the Croatian folklore group, each performer has their own pair with whom they form a duet. This was translated and con-

stituted in our work through the creation of stereo pairs, giving each virtual singer a corresponding counterpart;

- *Audio-visual uniformity*. As observed throughout the ethnographic research mentioned in the previous section 3.1, one of the key elements of a successful folklore group is uniformity on stage during live performance. This uniformity is seen as a sign of discipline and collectiveness, transpiring to the audience both through the visual content (dress, body language, height, movement) as well as the aural (singing). Hence, the hardware design of virtual singers is uniform and the same across all 16 singers (Fig. 3);
- *Limited autonomy*. When performing traditional choreographies, each performer has a strictly dedicated role which requires rigorous adherence. Yet when not performing as a collective, individual expressiveness that varies from performer to performer is welcomed. When applied to our *ansambl* and its standalone installation mode, each of the virtual singers has some minute internal decision-making mechanism that determines what and when a sound will be produced.

Therefore, virtual singers have the following characteristic properties: while they are all identical at their core, each one of them belongs to a set of pre-defined types that conditions certain behaviours. These inherent ontologies then present specific attributes – how each singer relates to the others, their relation to the human performer, their likeliness to interact, and their level of activity.

3.2.2 Technical Details

Each individual singer is equipped with a speaker, a custom-made RGBW LED light ring, and an ultrasonic distance sensor (HRLV-MaxSonar-EZ0), see Fig. 4. Singers are cabled back to a central amplifier and an Arduino MEGA microcontroller. The microcontroller receives data from the distance sensors and controls the LED lights around each speaker (individual LEDs within each speaker are addressable). A central computer manages sound distribution to each speaker and controls the Arduino. A vocal microphone is used by the performer to interact with the system. The physical form of the singers was fabricated using 3D printing and the virtual singers were hand assembled in our lab before being deployed for live performance.

A singing-voice dataset was developed as part of the core software framework for the work. The dataset consists of over a hundred recorded samples of the artist's voice and was used to build a "vocal matrix" of sounds that are triggered by the system in response to both performer vocalisations and audience behaviour. Across the vocal dataset, three main vocal techniques were used, where each one was performed twice, in first and second voice:

- *Falsetto*. A vocal technique is used for hitting higher notes than what the singer is commonly able to achieve. Results in a voice that sounds airy and soft;
- *Belting*. Defined by a strong but warm voice where the singer combines their chest and mixed voice;
- *Musical phrasing*. The vocal technique allows the vocalist to create a sequence of notes that would allow phrase expression (e.g., lyrics, spoken word, vocalisations).

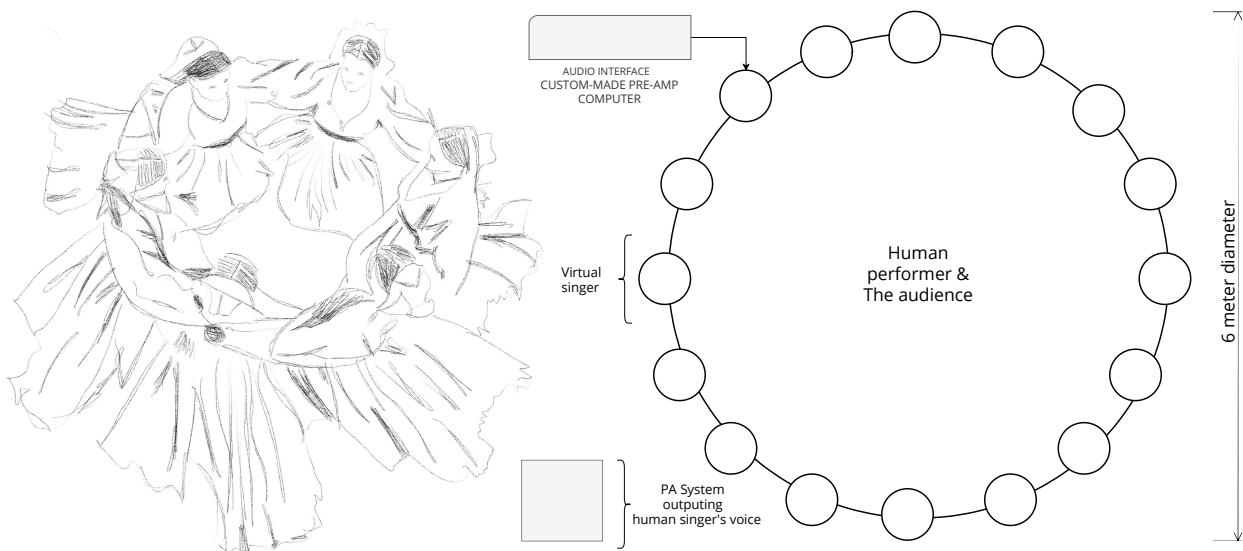


Figure 2: Installation layout sketch (left) showing the inspiration for the circular shape of Transhuman Ansmbl from traditional Croatian singing. The diagram on the right shows the layout and key elements used in the work.



Figure 3: Excerpt from the live performance showing the uniformity in the hardware design of the virtual singers along with the position of the human performer.

The dataset was developed based on traditional methods of Croatian composition, using two-part singing [15]. This dataset was then analysed and categorised manually according to pitch, vocal technique, length, and volume, then distributed among the singers, dividing them into two equal groups of eight, where one group sings the first voice, and the other one the second voice.

3.2.3 Modes of Interaction



Figure 4: Closeup image of the individual singer, showing the location of the speaker, ultrasonic distance sensor, and LED light ring.

As mentioned at the very beginning of Section 3, the *ansmbl* exists in two modalities – as a live performance with a human performer, and as a standalone sound installation. The installation layout and modes of interaction were inspired by research on the human voice beyond language, considering voice as an extension of the body: something that has both a physical and spatial dimension [7, 16].

The work was developed in MAX MSP, visual program-

ming software for music and multimedia. Each *singer* consists of four key and connected objects that determine their state and the current mode of interaction. Those objects are the same and standardised across all 16 singers and are as follows:

- *State indicator*. Based on a simple toggle between active (1)/inactive (0) states, this object determines whether the singer is singing. This state is dependent on the artist’s vocal pitch and incoming ultrasonic distance sensor data.
- *Playlist*. List of individual vocal samples created by the artist. They are assigned to each singer according to the voice group they belong to (first or second voice).
- *Ultrasonic distance data list*. Object used to register data collected from the ultrasonic distance sensor in real-time. This data is registered in the form of numbers from 1 - 10, representing the proximity of the closest object to the singer’s speaker. This element was introduced as a way to give virtual singers agency apparent through the ability to “sense” the audience at a close distance.
- *Serial RGB LED ring lights controller*. Object controlling LED ring lights, whose activity corresponds to the singer’s *State indicator* object. This feature was added during later stages of development to help detect which virtual singer is active (when acting individually), but also to amplify the engaging effect the circle has on the audience when live performing.

As shown in Fig. 5, the artist’s vocal audio is sent to the computer via a microphone, where it is analysed in real-time. This analysis includes the registration and interpretation of vocal features such as volume, pitch, and voice attack which is measured as the range of energy heard in an accordion frequency band (i.e., high dynamic results are classified as short or strong attacks while those with lower dynamics are classified as long and soft). In parallel, data from the ultrasonic distance sensor is registered and sent to the computer, affecting the output of the individual speaker by changing its volume or chosen vocal technique. To avoid the activation of the system while the singer is silent, or speaking instead of singing, certain spectral frequency limits were set by sampling the artist’s voice in various situations during development and noting its corresponding spectral signature. Once we have determined the spectral frequencies within which singing occurs, the software can recognise the artist’s voice when singing and differentiate singing from speaking. The received datasets are then registered and analysed for aforementioned parameters; if their values are within the predetermined margins, corresponding scenario activates and accordingly sends out information to all 16 singers (Figure 5). Each singer then responds accordingly, based on singing pitch and pitch length, selecting a vocal sample to play based on this information. This creates a type of “call and response” interaction between the performer and singers who respond to the performer’s voice with their own voices.

3.3 Performing with the Ansambl

Transhuman Ansambl was exhibited and tested through a series of performances in late November of 2022 while still in an early prototype stage Fig. 6. The audience (of approximately 100 people) expressed curiosity and wonder towards the singers as they were initially perceived as common

speakers until becoming active. Additionally, the invitation to experience the live performance within the boundaries of the circular ring (i.e. being surrounded by the singers) surprised the audience and evoked more interest in “who” the singers are and what they “do”. Audience members noticed how their physical distance from individual singers affected the choir’s singing. This was noted as an important aspect of interaction that needs further exploration since it demonstrated a potentially new way of interaction between the audience, singers and the performer. At that point, the audience became active performers, having the ability to make changes in singers’ performance based on their distance from them. For example, approaching a singer resulted in changing their voice from signing to falsetto to whispering.

Consequently, the act of “enclosing” the audience within the boundaries of the ring in close proximity to the performer created an intimate space that connects the audience and performer. This was evident through many oral accounts expressed immediately after performances by both the artist and the audience. As mentioned in Section 2.1, by doing so, the intersubjective and affective quality of voice was brought to the foreground. Transhuman Ansambl suggests a newly formed human-machine assemblage that employs technology to challenge the future of traditional roles, such as that of performer-listener [11], and to make us more perceptive to our surroundings, and of ways we affect each other through sound and physical presence.

The second rendition of the live performance mode is currently in the finishing stages of development and although not yet exhibited to the public, it has been tested separately with three experienced singers and the formal case study of the system is underway.

Apart from the substantial expansion of the singing-voice dataset assigned to each of the virtual singers, the greatest addition is the option for the human performer to create an infinite, stackable feedback loop together with all 16 virtual singers. The human performer is now able to instantly record and playback their singing through each of the speakers. While live looping is a well established and popular technique, our implementation is innovative due to the following features:

- *Partial control*. While the human performer is in control of what they will sing, each of the virtual singers is deciding (within a limited range) which part of the singing will be recorded and then playback through that particular audio channel. In essence, this loss of control over each “singing loop” creates space for surprise and improvisation, yet provides the virtual singers with an additional sense of agency and autonomy.
- *Spatial dimension*. Human performers can (re)arrange the ongoing feedback loop simply by standing closer to the virtual singers whose playback they prefer. By doing so, the other 15 virtual singers will echo the “chosen” one and perform as a collective. Ultimately this allows the human performer to then use the *ansambl* as an accompanying choir.
- *Immersiveness*. The human performer is not required to use any screen-based interface or device other than a microphone to perform with the virtual singers. Additionally, the spatiality of the work apparent through the 16 virtual singers that surround the human performer, reinforces their presence and as such strengthens the immersion.

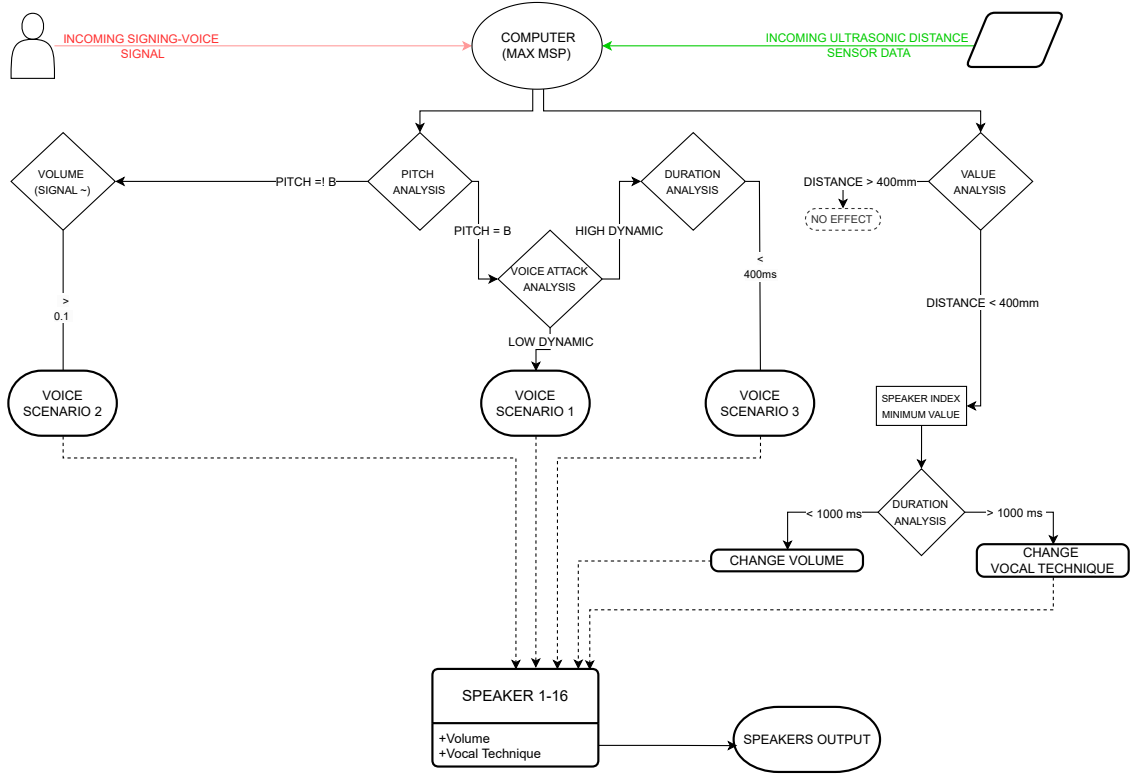


Figure 5: Transhuman Ansambl’s simplified schematic diagram illustrating the interaction process and the interrelationship between the singer’s vocal input and variables from the ultrasonic distance sensor on the speakers final output.

3.4 Standalone Sound Installation

The *ansambl’s* exhibition mode as a standalone sound installation (i.e. whenever not performing live with a human performer) is in its final stage of development. Conceptually, this rendition was envisioned as a technological translation of an intermission period in between performances among a group of performers. Moments when performers relax, engage in small talk, or prepare for the performance, which are usually hidden from the public are now brought closer to the audience. As such, it allows the audience to get closer to each of the virtual singers (Fig. 7), and freely explore them. The human interaction with the singers is made possible through either mere movement within the circle, or by closely inspecting one of them, listening to the voices they produce, individually or jointly. Technically, this is where the ultrasonic distance sensor comes into play, sensing the viewer’s movement and proximity. At moments the viewer can directly affect and provoke a particular singer by approaching it closely, but this won’t necessarily be a rule. As explained in Section 3.2, virtual singers have a certain degree of autonomy which allows them to decide to respond or not.

Hence, the modes of interaction between the audience and the virtual singers are intentionally not obvious at the beginning of the encounter; the singers produce sound but inconsistently, aiming to provoke exploration and capture the viewer’s attention.

During the aforementioned ethnography study in Section 3.1, we observed the behaviour of the performers during

those short periods of break and noticed a variety of common actions. Breathing exercises, vocal warm-ups, casual chatter, laughter, hydration, quick phone calls, and stretching were of interest to us as the most common ones. Rather than focusing on the meaning of each action or the reasoning behind it, we brought our attention to the sounds that their voice emanated and its spatial dimension (e.g. contagious laughter that spreads across the room, vocal warm-ups done in duets or triplets, loud abrupt breaths during breathing exercises). We then recreated those sounds using the 1st author’s voice, where the spatial dimension was achieved with both the physical layout of the virtual singers (large circle), and with spatial sound panning methods.

While this rendition of the work has not yet been shown to the public, it underwent several testing rounds where people could experience the installation, interact with the system (Fig. 7) and provide informal feedback. Almost all provided oral feedback mentioned a strong sense of the artist’s presence in the room, so as the feeling of intimacy with some of the virtual singers. The combination of the ability to physically interact and provoke a response in each one of the singers, as well as the act of attentive listening, viewers felt engaged in a reciprocal, and genuine interaction.

4. CONCLUSION

In this paper we have presented *Transhuman Ansambl*, a novel interactive work that features the use of the human voice as the primary means of communication between hu-



Figure 6: Transhuman Ansambl installation and live performance, Melbourne, November 2022. The audience and performer are surrounded by a circle of 16 autonomous virtual singers who collaborate and interact with the performer by listening and responding to the performer’s voice.



Figure 7: People interacting with the virtual singers.

mans and machines, using vocal articulations and song as the medium of interaction and creative exchange of human and machine agency [3]. Rather than conceptualising voice interaction in terms of commands or linguistic exchange, the *Ansambl* requires a far more creative and open-ended approach to the interface. Developed primarily as a live performance, although some modes of the system resemble “live looping”, what differentiates the two are distinct features such as i) partial control, ii) spatial dimension, and iii) immersion. To perform effectively with this system demands something of the performer as well as the virtual singers while both hold a certain level of autonomy. When singing with the *ansambl*, the human performer can only decide on what to sing, while each virtual singer decides on the playback section. Consequently, the loss of absolute control over the playback creates space for improvisation and surprise. Additionally, the spatial dimension of the work creates an intimate space even when placed within a public live performance space. The act of surrounding the audience with all 16, presence-sensitive virtual singers and thus closer to the human performer, suggests a new way of employing technology to challenge the roles of performer-listener, and of ways we affect each other through sound and physical presence. The standalone sound installation mode of the work that allows the audience to interact with the virtual singers without the presence of the human performer is currently in the final stages of development.

5. REFERENCES

- [1] M. Alivizatou-Barakou, A. Kitsikidis, F. Tsalakanidou, K. Dimitropoulos, C. Giannis, S. Nikolopoulos, S. Al Kork, B. Denby, L. Buchman, M. Adda-Decker, C. Pillot-Loiseau, J. Tillman, S. Dupont, B. Picart, F. Pozzi, M. Ott, Y. Erdal, V. Charisis, S. Hadjidimitriou, L. Hadjileontiadis, M. Cotescu, C. Volioti, A. Manitsaris, S. Manitsaris, and N. Grammalidis. Intangible Cultural Heritage and New Technologies: Challenges and Opportunities for Cultural Preservation and Development. In M. Ioannides, N. Magnenat-Thalmann, and G. Papagiannakis, editors, *Mixed Reality and Gamification for Cultural Heritage*, pages 129–158. Springer International Publishing, Cham, 2017.
- [2] P. Beyls. Interaction and self-organisation in a society of musical agents. 2007.
- [3] O. Bown and J. McCormack. Creative Agency: A Clearer Goal for Artificial Life in the Arts. In G. Kampis, I. Karsai, and E. Szathmáry, editors, *Advances in Artificial Life. Darwin Meets von Neumann*, Lecture Notes in Computer Science, pages 254–261, Berlin, Heidelberg, 2011. Springer.
- [4] J. Butler. Critically Queer. *GLQ: A Journal of Lesbian and Gay Studies*, 1(1):17–32, Nov. 1993.
- [5] L. Clark, P. Doyle, D. Garaialde, E. Gilmartin, S. Schlögl, J. Edlund, M. Aylett, J. Cabral, C. Munteanu, J. Edwards, and B. R. Cowan. The State of Speech in HCI: Trends, Themes and Challenges. *Interacting with Computers*, 31(4):349–371, 09 2019.
- [6] A. Desjardins, A. Psarra, and B. A. Whiting. Voices and voids: Subverting voice assistant systems through performative experiments. In *Creativity and Cognition, C&C ’21*, New York, NY, USA, 2021. Association for Computing Machinery.
- [7] M. Dolar, Stockholm Inst Of Transition Staff, S. iek, and A. Zupancic. *A Voice and Nothing More*. MIT Press, Cambridge, UNITED STATES, 2006.
- [8] S. Fasciani and L. Wyse. A voice interface for sound generators: adaptive and automatic mapping of gestures to sound. 2012.
- [9] W. W. Gaver. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2):167–177, 1986.
- [10] C. Goulding and D. Domic. Heritage, identity and ideological manipulation: The case of Croatia. *Annals of Tourism Research*, 36(1):85–102, 2009.
- [11] C. Heim. *Audience as Performer: The changing role of theatre audiences in the twenty-first century*. Routledge, London, Aug. 2015.
- [12] R. Kleinberger, N. Singh, X. Xiao, and A. v. Troyer. Voice at NIME: a Taxonomy of New Interfaces for Vocal Musical Expression. In *International Conference on New Interfaces for Musical Expression*, June 2022.
- [13] F. Lenzerini. Intangible Cultural Heritage: The Living Culture of Peoples. *European Journal of International Law*, 22(1):101–120, 02 2011.
- [14] O. Louvel. *The Sculpted Voice: an exploration of voice in sound art*. PhD thesis, D, 2019.
- [15] I. Milačić. Sveučilište u Zagrebu Filozofski fakultet Odsjek za fonetiku. 2021.
- [16] N. Neumark, B. LaBelle, R. Gibson, T. Van Leeuwen, T. Y. Levin, V. Madsen, J. Potts, T. Senft, M. Thomas, and M. Morse. *Voice: Vocal Aesthetics in Digital Arts and Media*. MIT Press, Cambridge,

UNITED STATES, 2010.

- [17] W. Oliver, J. Yu, and E. Metois. The Singing Tree:: design of an interactive musical interface. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, and techniques - DIS '97*, pages 261–264, Amsterdam, The Netherlands, 1997. ACM Press.
- [18] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach, 4th US ed.* 1995.
- [19] B. G. S. B. Stefania Serafin, Bill Buxton. *Auditory Interfaces*. Focal Press, Boston, Mass., 2023.
- [20] A. Tanaka. Sensor-Based Musical Instruments and Interactive Music. In R. T. Dean, editor, *The Oxford Handbook of Computer Music*, page 0. Oxford University Press, Apr. 2011.
- [21] K. Thomaidis and B. Macpherson. *Voice Studies: Critical Approaches to Process, Performance and Experience*. Taylor & Francis Group, London, UNITED KINGDOM, 2015.