Exploring “Synekinian Pairs”: Manual-Vocal Gesture Integration in Experimental Contexts

[redacted]

[redacted]

ABSTRACT

This study investigates the intuitive relationship between manual gestures and vocal syllables, grounded in shared neurobiological and motor systems. Participants were tasked with producing "Synekinian pairs," consisting of vocal syllables paired with instinctive gestures. Through qualitative and quantitative analyses, the findings reveal consistent patterns in manual-vocal associations, offering insights into gesture-based sound design, vocal synthesis, gesture notation and human-computer interaction. Future work aims to refine the experimental methodology, expand datasets, and validate these observations across diverse cultural and linguistic contexts.

1. INTRODUCTION

The parallels between manual and vocal gestures originate from shared neurobiological mechanisms and cognitive integration. Both modalities are controlled by interconnected neural networks, including Broca’s area, which plays a critical role in language processing and the coordination of communicative manual gestures [1]. This overlap points to an evolutionary link where manual gestures may have preceded vocal communication, shaping neural circuits that later supported spoken language [2].

Moreover, manual and vocal gestures share dynamic motor patterns and exhibit cross-modal integration. This integration enables correspondences across sensory modalities, such as associating rounded trajectories with soft sounds and angular trajectories with sharp sounds [3, 4]. The shared sensory and motor pathways underlying these associations optimize cognitive efficiency through common representations [5]. These findings underscore the profound connection between manual and vocal gestures, rooted in shared neural, motor, and perceptual systems [3].

Building on these principles, this paper proposes an experimental study to investigate the intuitive links between manual and vocal gestures. The goal is to develop musical and performative interfaces that are natural to learn and use. The paper presents the theoretical and practical framework for this study, followed by an experimental protocol used to gather an initial corpus of data. While we acknowledge the limitations of this exploratory phase in terms of universality and reproducibility, it provides a foundation for further quantitative studies based on an expanded data corpus. Additionally, ongoing work includes the automatic training of neural models to connect speech sounds to manual gestures. Finally, we explore potential applications of this research in fields such as gesture-based sound design, vocal synthesis, gesture notation, and multimodal interaction.

2. SPEAKING WITH THE HANDS

2.1 Describing sounds with the hands

The use of hand gestures to describe speech sounds relies on various strategies that make abstract phonetic concepts more tangible. One key approach is **articulatory mimicry**, where hand movements replicate the actions of speech organs. For example, bringing the hands together can simulate the closure of the lips for bilabial sounds like /m/, while pointing gestures can represent dental sounds like /t/ [6]. This strategy directly connects gestures to the physical production of sounds, providing a clear visual aid for understanding articulation.

Another method involves **iconic gestures** that evoke acoustic characteristics of sounds. Prolonged gestures, such as flowing hand movements, can represent sustained sounds like /s/, while sharp, abrupt hand motions can symbolize explosive sounds like /p/ or /k/ [7]. Similarly, **phonetic schematization** uses gestures to encode distinctive features of phonemes, such as the height of vowels or the place of articulation. For instance, high gestures can correspond to high vowels like /i/, whereas low gestures can depict open vowels like /a/ [8]. Iconic and schematic gestures also extend to **sound metaphors**, as in the depiction of rising tones with ascending gestures or rustling noises with rapid hand movements, linking gesture to auditory perception [9].

Finally, rhythmic gestures and **spatial representations** are employed to illustrate prosodic or phonological relationships. Rhythmic hand movements can represent the timing and stress of phonemes or syllables, an effective tool in pronunciation teaching [10]. Spatial mapping uses the gesture space to differentiate phonetic categories, such as grouping voiced and voiceless consonants into distinct areas [11]. These strategies highlight the multimodal nature of speech and the cognitive link between gesture and sound.

2.2 Gesture description

To describe a gesture, one can analyze its **spatial characteristics** (e.g., amplitude, direction, trajectory, and position in space) [8]. **Temporal aspects**, such as duration, rhythm, and phases of execution, are equally important [7]. The **dynamic qualities**, including force, fluidity, and speed, provide insight into the gesture's intensity and movement style [12]. Additionally, understanding the **intentionality and expressiveness** of a gesture—whether functional or symbolic—reveals its purpose and emotional undertones [13]. Lastly, the **relationship with the body and environment**, such as the involved body parts or interaction with objects, contextualizes the gesture within its physical and social framework [14]. This multidimensional approach enables a comprehensive description of gestures.

3. THE SYNEKINE PROJECT

For the past ten years, through residencies and artistic creations, *the Synekine Project*[1] has invited performers to question the intimate relationship between vocal gestures and manual gestures, through the manipulation of scenic devices based on new technologies [15-24]. Metaphorically, the preeminent neuromotor link between voice and gesture is closed by "creative prostheses" joining the capture of movement to the transformation of the voice by artificial intelligence. Linking space and time through movement then transforms the search for sound into a scenic exploration. By analogy in synesthesia, the neologism "Synekinesia" would reflect our capacity to associate two or several sense motors.

3.1 The "synekinian pair"

The debate on whether vocal production is based on phonemes or syllables reflects two complementary perspectives in speech processing. Phonemes, the smallest units of sound distinguishing meaning (e.g., /p/ vs. /b/ in French), are central in linguistic models [25] and are reflected in alphabetic writing systems and speech errors. Syllables, larger prosodic units combining vowels and consonants (e.g., /pa/), are supported by articulatory data, rapid repetition tasks, and neuropsychological evidence, suggesting they are natural units of motor planning. Modern approaches view speech production as hierarchical: phonemes are abstract units of encoding, while syllables function as integrated motor units. Both perspectives are necessary to fully capture the complexities of human vocal production.

In our approach, the syllable is the atomic unit of articulatory movement in vocal emission. Our aim is to observe the hand gestures that accompany these vocal atomic gestures. We deliberately set aside the linguistic information of language, even though we are dealing with the production of speech. Phonesthesia (the study of the intuitive connection between the sounds of speech and the sensory or emotional qualities they evoke) and symbolic sign language are left aside to focus on the relationship between gesture and sound, at the level of their production. To this end, we've set aside all semantic aspects of language, discarding meaning at the linguistic, phonetic and symbolic-gestural levels. As no intentionality or object is involved, we focus on the spatial, temporal and dynamic characteristics of the gesture.

Thus, *Synekinian pairs* describe meaningless syllables expressed simultaneously by the vocal apparatus and by non-symbolic manual gestures accompanied by body movements.

3.2 Grapevine Game

In 2014, during a residency in Paris with Lab//SEM[2], a group of musician-performers and dancers who combine artistic practices, three small group games were played to explore the relationship between voice and gesture. They were asked to produce "Synekinian pairs", i.e. pronounce a syllable while involving a body movement or hand gesture that instinctively and expressively accompanies the vocal sound emitted, meaning set aside.

The grapevine game, often referred to as telephone, is a classic party or group activity where participants pass a message from one person to the next through a line or circle, usually by whispering. The goal is to see how much the message changes as it travels through the group. Here, the message consists of a short sequence of one to three *Synekinian pairs*. The first person shows a new sequence to the second, who repeats it to the third, and so on. The last person transmits again to the first. Finally, the first person shows the whole group the final pair, then the initial given pair. The roles then change incrementally.

Surprisingly, it seems that some *Synekinian pairs* are more robust than others. Inter-individually, some pairs are more resistant to communication, while others find themselves unrecognizable at the end of the chain, in terms of both gesture and sound. Intra-individually, the person who created the initial sequence may find it difficult to reproduce the final version, as if the pairs composing it are archetypally ingrained in the memory or internal representation of their author.

3.3 Syllables Game

In this other game with the same group lined up, a syllable written on a sheet of paper is revealed. As soon as the syllable is exposed, each member of the group is

[1] [redacted]014/01/synekine-project/

[2] https://www.ensemble-labsem.fr/

invited to quickly create, produce and repeat the corresponding *Synekinian pair*, without looking at the others. In the repetition, a certain acoustic fusion takes place (pitches and onsets align, for example), but also, and sometimes, a gestural fusion. For some syllables, from the very first exposure, the same synekinian pair appears in different participants. For example, /if/ seems expelled, with a gesture going from the inside towards the outside whereas /mu/ seems round, with a gesture of gathering, from the outside inward. This simple observation allows us to hypothesize the existence of common *Synekinian pairs*, without providing any information on their origin, whether universal or cultural.
To further test this hypothesis, we decided to take the analysis a step further by offering the game to different people, from different cultures, and in isolation (one at a time), to avoid group fusion phenomena.

4. EXPRESSO GAME

In 2020, in the context of the creation of a piece of musical theater involving the student choir of the HfMT Hamburg, 16 singers, men and women of diverse origins (asia, europe, north and south america and africa) but of the same age (around 22 years old) were invited to participate in an individual session of the now named “Expresso Game”, including a recording for research and creation purposes.

4.1 The rules of the game

Here is the information given to participants at the start of the recording session: “You will see a table with syllables. One by one, in order, you will pronounce these syllables, making the gesture that seems most appropriate. Don't think too much, the important thing is to produce the gesture at the same time as you pronounce the syllable. You can repeat each syllable until the associated gesture is stabilized (e.g., 5 to 10 times). Then simply move on to the next syllable. You can involve the whole body (movement, posture, hand movement, head movement...).”

BE	RA	IH	SO	EN	SA
JE	AL	WI	IN	AN	HE
ER	TE	RO	DE	LE	LO
HA	EI	IC	NE	CA	CHE
ME	ES	FE	GE	AU	ZE
PO	KE	PU	SSE	UN	HÖ
KA	EL	A	DI	MO	IT
WE	DA	BE	HE	SE	PA
LI	NI	SI	AS	ZU	RE

Table 1. Syllables used in the Expresso Game

The set of syllables selected artisanaly in Table 1 is based on a simple phonological structure - /V/, /CV/ or /VC/ only, on the use of non-nasal voiced vowels and on a variety of consonants whose writing is intended to be universal (which turned out not to be the case.). The /BE/ and /HE/ syllables were deliberately duplicated to test the intra-individual robustness of the *Synekinian pairs* produced.

4.2 Data acquisition and filtering

Recording sessions took place weekly over a total period of three months, and enabled the acquisition of 16 (participants) times 54 *Synekinian pairs*, each repeated between 4 and 10 times with the following technical set-up:

- video front and sound - camera Sony fisheye
- video side - camera Sony HD
- skeleton, depth and sound - kinect V2 studio
- sound - Neumann KM184 stereo set XY

An initial hand-crafted editing phase, performed in the Ableton Live 10 audio/video sequencer, resulted in 54 videos, one for each syllable. Each video shows in sequence the last occurrence of the given syllable by each participant. An example of Video for the syllable /RO/ and /IT/ is available[3].

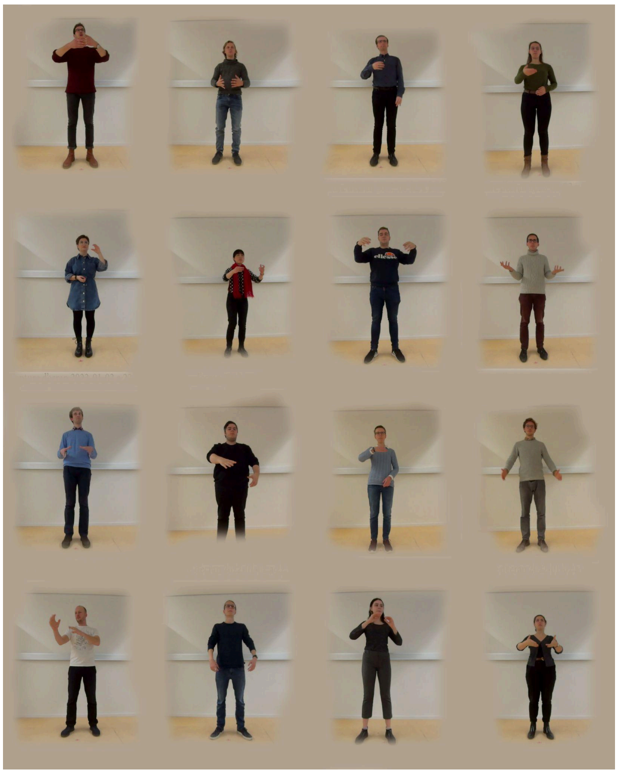

Figure 2. 16 participants producing the *Synekinian pair* /BE/.

This video material enables visual detection of recurring, similar gestures used by participants to express the same syllable. In the strategies employed, few used whole-body movement, and most used only hand gestures. Some used both hands, which then moved in mirror image. Some used only one hand. Some participants didn't quite understand the instructions (deliberately open-ended at this stage of the experiment):

[3] https://youtu.be/Lh8fpqpTlbE

Some did not move, others looked for meaning in the syllables and sometimes made symbolic gestures, and finally two people always used the same gesture, whatever the syllable pronounced. In conclusion, we retained only 8 participants whose gesture amplitude and variability seemed significant.

5. ANALYSIS

In this section, we delve into the pre-processing and analysis steps undertaken to explore the relationship between hand gestures and speech articulation. The study initially aimed to leverage Kinect-recorded data; however, proprietary file limitations led to the adoption of Mediapipe's Pose Landmark Detection for extracting hand motion data from video recordings. By normalizing hand pose markers relative to facial coordinates, we constructed a robust dataset to analyze gesture dynamics. The focus then shifted to quantifying these movements through a comprehensive set of gesture descriptors, enabling both qualitative and quantitative analyses of phonetic classes such as vowels and consonants. Through these descriptors, we sought to uncover patterns linking gesture dynamics to phoneme articulation, shedding light on the interplay between vocal and manual gestures.

5.1 Pre-processing

Our initial intention was to use the data recorded by the kinect. Unfortunately, the file format is proprietary, and the files can only be read in real time in kinect studio. Instead, we decided to extract the skeleton from videos taken with the front camera using mediapipe's Pose Landmark Detection [27]. We use the pose markers of the hand, whose coordinates have been normalized with respect to the center of the face. Unfortunately, hand tracking using mediapipe's Hand Landmark Detection didn't work. So we're using the makers from Pose to describe the hands (wrist, thumb, index and pinky).

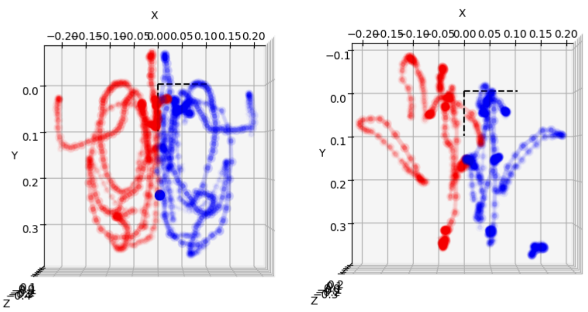

Figure 3. Superimposed left-hand (blue) and right-hand (red) trajectories for 8 participants viewed from the front. Syllables /RO/ on the left and /IT/ on the right.

For each syllable, we began by superimposing the various standardized gestures. Figure 3 shows that it is visually possible to distinguish certain syllables, such as /RO/ and /IT/, from the visualization of the gestures of the two hands of the 8 participants, superimposed and seen from the front. The aforementioned video shows the samples used to generate these figures.

5.2 Gesture descriptors

A comprehensive set of descriptors to quantify the dynamics and geometry of three-dimensional gestures based on spatiotemporal data (x, y, z, t), has been extracted, focusing on hand movements, shapes, and dynamics. Gesture descriptors include the duration of the movement, total path length, average and maximum velocity, average and maximum acceleration, jerk (rate of change of acceleration), and curvature. We also calculate linearity (how straight the trajectory is), distance to a target position, and target directionality. For hand shape, we compute distances between the wrist and key fingers (pinky, index, thumb), as well as angles between these fingers. Dynamic descriptors analyze the variation in distances between the wrist and fingers over time. Additional metrics include symmetry between movement axes, speed variability, acceleration variability, root mean square (RMS) values of velocity and acceleration, kinetic and potential energy (assuming unit mass), and smoothness (inverse of acceleration squared).

5.3 ANOVA in phonetic classes

Then we analyzed phonetic gestures and their associated descriptors, leveraging spatiotemporal data and focusing on phoneme classes (vowels and consonants). We categorized syllables based on their phonetic properties into classes such as vowels, voiced/unvoiced fricatives, voiced/voiceless plosives, nasals, liquids, and glides. We performed an ANOVA to detect significant differences and visualize descriptor distributions. The following significant descriptors (P-values < 0.05) are listed in ascending order of their P-values: "right hand average velocity", "right hand rms velocity", "right hand total distance", "right hand amplitude z", "right hand amplitude y", "right hand acceleration variability", "right hand rms acceleration", "right hand index variation", "left hand average velocity", "right hand symmetry x", "right hand max velocity", "left hand pinky variation", "left hand index variation", "right hand speed variability", "right hand average acceleration", "right hand symmetry y", "right hand max acceleration", "left hand amplitude z", "left hand path length", "right hand kinetic energy".

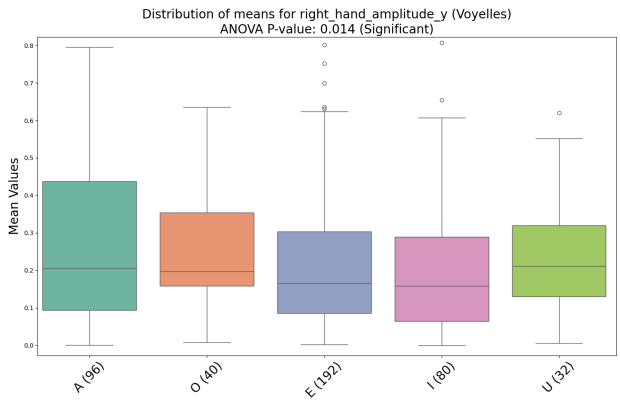

Figure 4. ANOVA of "right hand amplitude y" descriptor means for vowels (with number of samples). The smaller the value, the more upward the gesture (due to normalization with respect to face position).

If we take a closer look at vowel distributions (see figure 4) on one of its descriptors, for example, "right hand amplitude y", we can see a distribution of gesture in the space linked to the vowel, independently of the consonant and its place in the syllable.

5.4 Vowels

Even if a consonant and its position in the syllable have an effect on vowel trajectory, there seems to be a relationship between vowel and hand position in frontal space. Consistent with phonetic schematization, the height of the gesture would reflect the volume of the mouth used to emit the vowel. The smaller the /i/, the higher the gesture. The larger /a/, the more the hands move downwards (see figure 5). This may be due to the fact that all the participants had received singing training, during which they learned how to place their vowels for projection. This may be connected to different factors, such as the opening or volume of the mouth, the placement of resonances in the head or chest, the brightness of the vowel (spectral centroid), or the singer's ability to project sound at different heights. In all cases, Vowels seem distributed within a frontal zone near the body, which can be used in a gesture-based sound design tool or to enable gesture-based control of speech synthesis.

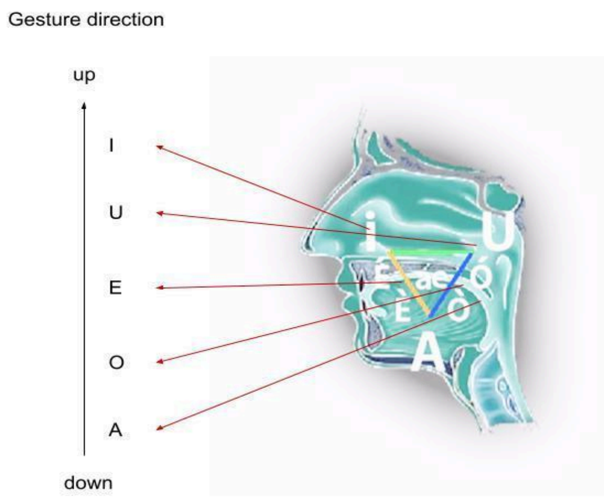

Figure 5. Artwork showing that vowels appear to be distributed on a frontal plane

5.5 Consonants

Depending on the phonological families of consonants, certain features of the gestures employed seem redundant enough to be noticed. Given the number of samples taken per consonant (for some, only 8), a quantitative analysis cannot reveal them. The following characteristics are the result of a qualitative analysis:

- **Trajectory**: Linear/angular vs. round/circular: The gestures accompanying fricatives are distinguished by their sinuosity. While the labiodental /F/ is linear, the interdental /CH/ is arondial.
- **Directionality**: Inward vs outward: Nasals such as /m/ or /n/ are accompanied by gestures from outside to inside, whereas the plosive bi-labials /p/ and /b/ are directed outwards.
- **Dynamics**: Smooth, explosive, subito, rebound: The velar plosive /K/ is expressed by an explosive gesture, while the alveolar fricative /SS/ is expressed by a constant-speed gesture.
- **Hand action**: grasping, opening, pointing: The alveolar plosives /T/ and /D/ are accompanied by a pointing gesture with the index finger, or by a grasping gesture with the index finger and thumb. Liquids /R/ and /L/ are accompanied by a gesture of the hand grasping something.

Gestures accompanying fricatives seem to describe features of airflow in the vocal tract, while some plosives are manifested by hand gestures mimicking labio-dental articulators. Manual and vocal gestures share dynamic motor patterns. Rounded, fluid sounds like /m/ or /l/ require smooth articulatory movements, akin to the circular motions of the hands, while abrupt sounds like /t/ or /k/ involve rapid muscle contractions, paralleling angular hand movements [26]. This suggests that gesture is linked not only to phonation, but also to the supraglottic sources of sound production/modulation (tongue, palate, teeth and lips). It should be noted that ancillary manual gestures appear before and after the sound. In the context of *Synekinian pairs*, whose production is simultaneous, these manual and silent gestures seem to reveal the presence of ancillary gestures of the vocal articulators involved in the planning and termination phases of vocalisation. This hypothesis could be verified by direct measurement of vocal articulators during production, using magnetic resonance imaging (MRI) for example.

6. RELATED FIELDS OF APPLICATION

This preliminary study aims to explore the relationship between manual gestures and speech sounds. Regardless of text, meaning, or language, the findings indicate that manual gestures are closely linked to speech production. This connection holds significant potential for a variety of applications.

First, in the realm of multimodal Human-Machine Interaction involving voice, gesture observation could enhance voice recognition systems. In the field of speech synthesis, learning these correlations automatically could pave the way for a “gesture-to-speech” synthesizer, capable of generating syllables based on gestures and their spatial positioning. Similarly, these correlations can serve as a control framework for articulatory speech synthesizers. Conversely, in the field of gesture choreography, the linguistic symbols of syllables can be used to define an instinctive, legible notation for gestures. Finally, this study is set within the broader framework of an artistic initiative aimed at creating a novlang based on eco-choreographic principles of vocal articulators, maximizing the signifier/signified relationship (Phonesthesia) and whose manual gesture metaphorically constitutes the imaginary part of an expressivity revealed by the real sound and vocal part of language.

7. NEXT STEPS

We would like to make it clear here that we are fully aware of the limitations of our approach from a scientific point of view, but we consider that these results may nevertheless be of interest to artists working on the control of sound through gesture. This experiment and the results observed here can be further validated by the scientific method by considering the following protocol.

In order to improve and refine this study, we need to define a more restricted framework for the acquisition of a new corpus, which can be analyzed quantitatively and exploited directly by the applications mentioned. For this purpose, a set of syllables linguistically varied (by mixing languages) and phonetically transcribed (to avoid the vagaries of pronunciation stemming from the different origins of the participants) must be selected. Some of these syllables need to be duplicated for later data validation (as was the case in this study). The choice of participants' origin, age and expertise in vocal and gestural expression must also be varied to guarantee the universality hypothesis. The number of participants can be increased to achieve statistical significance. The instructions can be made more precise, with explicit reference to the articulatory mimicry strategy. In addition to the video and sound recordings, a detailed recording of body movements, including those of the arms, hands and fingers, is required. The corpus should be segmented by syllable and user, then analyzed using audio and gesture descriptors. A quantitative study should reveal intra-individual and inter-individual correlations between these descriptors. Finally, machine learning must be carried out to enable gesture-based control of a vocal synthesizer, gesture-based control of a sound design tool, gesture notation using phonetic symbols, or the definition of a set of expressive morphemes from which a new language can be built.

8. CONCLUSION

This study highlights the intricate relationship between manual gestures and speech sounds, suggesting that these connections are deeply rooted in shared motor, sensory, and cognitive processes. By exploring the concept of Synekinian pairs, the research uncovers recurring patterns in the way gestures and sounds are naturally linked, offering a promising foundation for both scientific and artistic applications. The qualitative findings, though limited in scope, provide valuable insights into how gestures can intuitively correspond to phonetic and prosodic features, paving the way for innovative tools in human-machine interaction, speech synthesis, and multimodal choreography.

Moving forward, a more robust quantitative framework is necessary to validate and expand these preliminary results. Expanding the participant pool, refining the methodology, and incorporating diverse linguistic and cultural backgrounds will help ensure the universality of the observed phenomena. Additionally, integrating advanced data analysis and machine learning techniques could unlock new possibilities for gesture-based vocal synthesis and expressive language creation.

This research not only bridges the gap between gesture and sound but also opens up a wide array of interdisciplinary opportunities, blending technology, art, and cognitive science in novel and impactful ways.

Acknowledgments

This research was funded under the Program "Innovative Hochschule" (innovative university) by the Federal Ministry for Education and Research (BMBF) of Germany (Grant No. 13IHS232C) and the City of Hamburg. The authors would like to thank Mathilde Vrignaud and the Lab//SEM as well as Prof. Annedore Hacker-Jakobi and the student choir of the HfMT Hamburg for their participation.

9. REFERENCES

[1] Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neuroscience & Biobehavioral Reviews, 30*(7), 949–960.

[2] Corballis, M. C. (2003). *From hand to mouth: The origins of language*. Princeton University Press.

[3] Maurer, D., & Mondloch, C. J. (2006). The infant's cross-modal perception of unity. *Developmental Science, 9*(5), 438–445.

[4] Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995.

[5] Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3-4), 455–479.

[6] Eberhard, D. M. (2008). Simulating speech sounds with gesture: Implications for phonetic learning. *Journal of Phonetics, 36*(3), 431-445.

[7] McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

[8] Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

[9] Lemaitre, G., Scurto, H., Françoise, J., Bevilacqua, F., Houix, O., & Susini, P. (2017). Rising tones and rustling noises: Metaphors in gestural depictions of sounds. *PLoS ONE*.

[10] Wellington, W. (2014). Gesture as a tool for second language pronunciation teaching. *TESOL Quarterly, 48*(4), 753-776.

[11] Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Harvard University Press.

[12] Laban, R., & Lawrence, F. C. (1974). *Effort.* Macdonald & Evans.

[13] Ekman, P., & Friesen, W. V. (1969). *The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding.* Semiotica.

[14] Goodwin, C. (1981). *Conversational Organization: Interaction Between Speakers and Hearers.* Academic Press.

[15] [redacted] «Spatial Sampling in Mixed Reality», enzhen, October 2023

[16] [redacted] The Synekine Project», KiSS Kinetics in oun s an Space, Wolke-Verlag, November 2022

[17] [redacted] «Le Projet Synekine», Alternatives théâtrales, N° 146 – Les scènes du son – perspectives contemporaines, March 2022

[18] [redacted] « Sound Space and Spatial Sampler », MOCO 2015, SFU, Vancouver, August 2015

[19] [redacted] « The Synekine Project», MOCO 2014, IRCAM, Paris, June 2014

[20] [redacted] « Gestural Control Of Real Time Concatenative Synthesis », ICPhS, Hong Kong, 2011

[21] [redacted] « Gestural Control of Real-Time Speech Synthesis in Luna Park », SMC, Padova, 2011

[22] [redacted], « Transformation of xpress v y n peec », The Role of Prosody in the Expression of Emotions in English and in French, ed. Peter Lan . Peter Lang), 2009

[23] [redacted] « Analyse et Modèle n rat e xpress v té : application à la parole et à l'interprétation musicale », PhD Thesis, thèse doctorale de l'Université Paris 6 – IRCAM, 2009

[24] [redacted] G., Obin, N., Rodet, X., « Articulation Degree as a Prosodic Dimension of Expressive Speech », Speech Prosody 2008, campinas, 2008

[25] Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.

[26] Ramachandran, V. S., & Hubbard, E. M. (2001). Synesthesia—a window into perception, thought, and language. *Journal of Consciousness Studies, 8*(12), 3–34.

[27] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C. L., Yong, M., Lee, J., Chang, W. T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A Framework for Perceiving and Processing Reality. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.