

# Sound Space is Threefold

Davide Rocchesso<sup>1</sup>

<sup>1</sup>Dipartimento di Informatica - Università di Milano  
via Celoria, 18 - 20133 Milano - Italy

davide.rocchesso@unimi.it

**Abstract.** *Sound can be described as superposition of sines, noise, and transients, and these components can be embodied as acts of phonation, turbulence, and pulses occurring in the human voice apparatus. The Quantum Vocal Theory of Sound introduced a strict analogy between the threefold description of sound and the three-axes of the Bloch sphere, so that sound time evolution can be described similarly to spin evolution in a magnetic field. A threefold partition of the sound world is at the basis of descriptions of sound objects and of phenomenology of sound, as well as of their embodiment through the concept of musical gesture. The interplay of sounding objects can also be conveniently represented in three-axial space, where trajectories objectify mutual relations.*

## 1 Introducing the Quantum Vocal Theory of Sound

A Quantum Vocal Theory of Sound (QVTS) was first sketched in 2018 [1] as a strict, though rather arbitrary, analogy between a threefold description of the sound world in terms of vocal primitives, and the spin as a three-vector in the Bloch projective representation of a two-dimensional complex Hilbert space. Temporal sound evolution, if represented similarly to how a spin evolves in a magnetic field, could be related to the act of listening, where attention shifts and streaming effects can be described in analogy with quantum measurement [2]. Listening is therefore related to the objecthood of sound, which itself can be embodied through human sound-generating apparati. Repeated cycles of evolution, measurement, and collapse can be applied to the sound material to mimic listening processes, or to transform the sound material itself through such processes. Measurement can follow acts of will, or be driven by context, similarly to coarticulation in phonetics. Each cycle of evolution, measurement, and collapse can be compactly represented through quantum operators and become a segment of quantum computation, thus allowing to represent sonic processes as quantum circuits. The outcome of each measurement stage is inherently probabilistic, thus producing trajectories in a garden of forking paths [3], for creative sound processing and generation [4, 5]. In the restricted context of pitch processing, the same analogy of spin in a magnetic field has been proposed as a model for perception of Shepard-tone melodies, where a superposition of ascending and descending intervals is found for each couple of successive notes, and measurement decides the perceptual direction [6, 7]. In the large context of studies in creativity and cognition, design processes may be represented in the Bloch projection [8].

### 1.1 The phon

The word *phon* was proposed to indicate a spin-like embodied description of sound [2], as seen through the three axes of the Bloch sphere, each made to correspond to one of the major articulatory components of a vocal sound production. The three axes of this “phonetic space” have an embodied vocal meaning, that can be used to describe sound in general:

- $z$  is phonation, giving periodic waveforms, and eliciting a pitch sensation;
- $x$  is turbulence, giving frequency-shaped noises, of variable perceived brightness;
- $y$  is pulsation<sup>1</sup>, giving pulses, of variable perceived rate, or tempo.

A generic phon state  $|\psi\rangle$  can be expressed as the superposition of the basis vectors pitch-up  $|u\rangle$  and pitch-down  $|d\rangle$

$$|\psi\rangle = \alpha_u |u\rangle + \alpha_d |d\rangle, \quad (1)$$

where the coefficients are complex numbers,  $\alpha_u = \langle u|\psi\rangle$ , and  $\alpha_d = \langle d|\psi\rangle$ . The probability to measure pitch-up is  $\alpha_u^* \alpha_u$ , and the probability to measure pitch-down is  $\alpha_d^* \alpha_d$ , with the two probabilities summing up to one. In quantum computation, the vectors  $|u\rangle$  and  $|d\rangle$  give the computational basis of a qubit vector space, and are the eigenvectors of the Pauli-Z operator  $\sigma_z$ , corresponding to a Z gate, also called a phase flip. Applying a measurement along the  $z$  direction to a generic phon state  $|\psi\rangle$  is equivalent to pre-multiplying it by one of the measurement operators

$$M_u = |u\rangle \langle u| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}; M_d = |d\rangle \langle d| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and to normalize the resulting vector to have length one.

The eigenstates of the Pauli-X operator  $\sigma_x$  are  $|r\rangle$  and  $|l\rangle$ , corresponding to turbulent primitive sounds having different spectral distributions, one with the rightmost (or highest-frequency) centroid and the other with the leftmost (or lowest-frequency) centroid. In quantum computation,  $\sigma_x$  corresponds to a X (NOT) gate, and  $|r\rangle$  and  $|l\rangle$  form the Hadamard basis. If the phon is prepared turbulent ( $|r\rangle$  or  $|l\rangle$ ) and then the measurement apparatus is set to measure  $\sigma_z$ , there will be equal probabilities of getting pitch-up or pitch-down phonation as an outcome. This measurement property is satisfied if the turbulent state  $\{|r\rangle, |l\rangle\}$  is defined as

$$\{|r\rangle, |l\rangle\} = \frac{1}{\sqrt{2}} |u\rangle \pm \frac{1}{\sqrt{2}} |d\rangle. \quad (2)$$

<sup>1</sup>This has previously been called the myoelastic component, as low-rate impulse sequences can be generated by supraglottal myoelastic oscillations. However, we here include other kinds of transient sounds, that can be vocally achieved by other means, such as stops and clicks.

The eigenstates of the Pauli-Y operator  $\sigma_y$  are  $|f\rangle$  and  $|s\rangle$ , corresponding to fast-rate and slow-rate<sup>2</sup> pulsation, respectively. If the phon is prepared pulsating ( $|f\rangle$  or  $|s\rangle$ ) and then the measurement apparatus is set to measure  $\sigma_z$ , there will be equal probabilities of getting pitch-up or pitch-down phonation as an outcome. This measurement property is satisfied if the turbulent state  $\{|f\rangle, |s\rangle\}$  is defined as

$$\{|f\rangle, |s\rangle\} = \frac{1}{\sqrt{2}} |u\rangle \pm \frac{i}{\sqrt{2}} |d\rangle. \quad (3)$$

## 2 Threefold representations in audio signal processing

A threefold representation of sound is at the basis of a large part of sound analysis and transformation methods; It can be called Sines+Transients+Noise (STN) [9, 10, 11] or, in a music-oriented approach, Harmonic, Percussive, and Residual (HPR) [12]. The bag of available audio-processing tools, all based on the Short-Time Fourier Transform, makes it possible to play with the QVTS, as the local time evolution, and the decisions on how and when to proceed with the measurement and the subsequent collapse, can be derived from the short-time gist or local sound context – tonal quality, presence of noise, and percussive content – as it unfolds. That is, from a given sound state, or phon, the evolution, measurement, and collapse depend on the sound gist itself, thus making the interaction between system and measurement concrete [13].

To make the connection with audio signal processing clear, consider the twofold Harmonic+Noise representation [14] to describe preparation and measurement in the phonation-turbulence planar section of the QVTS space. An arbitrary direction in such plane is described by the operator

$$\sigma_n = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}, \quad (4)$$

where  $\theta$  is the angular direction, pointing to a superposition of phonation and turbulence. If we prepare the phon in the state corresponding to  $|\lambda_1\rangle$  (eigenvector of  $\sigma_n$  with eigenvalue +1) and we measure along the  $z$  axis, we get a pitch-up with probability  $\cos^2 \theta/2$  and a pitch-down with probability  $\sin^2 \theta/2$  [2]. Similarly, we can use threefold decompositions that extract transient (or percussive) components, and act on the phonation-pulsation plane, on the turbulence-pulsation plane, or on the whole phonation-turbulence-pulsation space. Having a framework of threefold separation of audio, specific audio signal processing techniques can be used to extract the dominant pitch, or tempo, or spectral brightness, often with a measure of confidence of each. This is how quantum measurement can be actually performed and interpreted on the time-frequency plane, which defines the objecthood of sound [15, 16].

## 3 The objecthood of sound

Tensor products of partitions of the “sound universe” are found in the phenomenology of sound as investigated by

Pierre Schaeffer [17, 18]. That was indeed a theory of sound objects, found at “the meeting point of acoustic action and listening intention”, where typo-morphological spaces, that can be reported on the time-frequency plane, support a phenomenological description of sound and its embodiment as human gesture [19]. The partition of sound according to its {impulsive, sustained, iterative} *facture* (ideally deriving from punctual, continuous, or iterative gestures) can be cross-multiplied by the classes of {tonal (pitched), complex (inharmonic or noisy), varied (fluctuating)} *mass*, to give a  $3 \times 3$  mesh of typological categories [20]. A notation for the nine cells of this mesh, using glyphs to represent sound objects and their relations, has been proposed and refined at the morphological level [21]. A partition of a three-dimensional audible space, with axes describing profile/sustain (*facture*), caliber (tonal or complex mass), and variation (varied mass), where to locate sound objects in a typological analysis, has been proposed [22].

Schaeffer’s thesis of a natural perceptual field of the ear, in which sound objects and their relationships emerge and are located in accordance with natural laws, is indeed very close to the QVTS. There might be differences in how the perceptual field is structured and in what are its dimensions, but the idea of sound objects appearing, being identified (typology), and evolving in a space under natural laws parameterized according to a context is indeed part of both theories. And the abstract space, or spaces, where sound objects find their place is not the space we inhabit and experience through vision and touch, but it becomes concrete through the features of the time-frequency plane, that *musique concrète* plays with. That is the playground of Schaeffer’s sound objects [18] as well as Kubovy’s auditory objects [15].

Sound objects are the quantal elements of sound [23], chunks or *gestalts* of length between 0.5 s and 5 s (the temporal meso-scale) that are perceived as wholes and categorized by Schaefferian typology, although they have an internal structure (morphology), and belong to larger-scale structures and context. Chunks occur at discrete moments in time and accumulate a segment of continuous experience. These are now-points in Husserl phenomenology, expressing the thickness of present [24], where temporal coarticulation acts as a contextual smearing [20]. Within this window of present, the perceived order of events may not match their physical order. Considering the phenomena of temporal displacement [25], objects of different types (or different axes of the phon space), or of markedly different quality (or different basis states along the same axis), resist physical simultaneity. Classical experiments in this area of investigation are based on two different low-pitched tones interrupted by either a noise burst or a high-pitched tone [26]. It is found that the middle element is perceptually anticipated or retarded, as compared to the instant of physical occurrence. The probabilistic nature of the QVTS cycle of evolution, measurement, and collapse makes the noise burst, that is physically simultaneous with a note transition, to appear as a collapse

<sup>2</sup>A single impulse is assimilated to very slow pulse rate.

along the  $x$  axis, at different times on different runs of the evolution, within a present as thick as a few hundred milliseconds. When such collapse has occurred, equation 3 shows that the turbulent state can be expressed as a balanced superposition of pitch-up and pitch-down. The representation of noise in the computational basis of tones, or sines, is indeed justified by phenomena of perceptual restoration or auditory tunneling [27, 28], where the phon effectively keeps its  $z$  orientation while traversing a segment of masking noise.

If sound quanta have to be looked for at the meso-scale, it is challenging to understand how a sound object emerges as a gestalt, from the fusion of micro-sonic features [29]. One sensible approach is to treat sound as inextricably linked to movement [20], chunked in perception-action motor gestalts [30]. In QVTS, the space is organized along axes that directly refer to vocal production mechanisms, thus making the link to movement evident and embodied. For example, a train of 30 pulses in a second may be connected with a bilabial trill generated by a pulmonic egressive gesture, and that whole would be a sound object.

A quantum-inspired description of sound should deal with two motion-related phenomena that are central to define the objecthood of sound [20]:

- . Phase transition: a qualitative categorical shift due to incremental change in motion features;
- . Coarticulation: the fusion of otherwise separate elements due to anticipatory (forward) or perseveratory (backward) smearing.

A phase transition is found when a pulse rate becomes high enough to evoke a pitch sensation, or a rotation of the phon from the  $y$  to the  $z$  axis. The categorical shift corresponds, in embodied vocal terms, to a change in the generation mechanism, from the lips to the vocal folds.

Coarticulation is introduced when considering the phon as immersed in an auditory environment, with an evolution alike that of a spin in a magnetic field. Restoring forces act on the phon, coming from the local temporal sound production context. Within a time neighborhood of the present, using a time-dependent yet commutative Hamiltonian

$$H(t) = e^{-kt} \mathbf{S}, \quad (5)$$

we can govern the coarticulation smearing by the damping factor  $k$ , where  $\mathbf{S}$  is a Hermitian matrix representing the local sound gist. The phon is made to have unitary evolution within the time neighborhood:

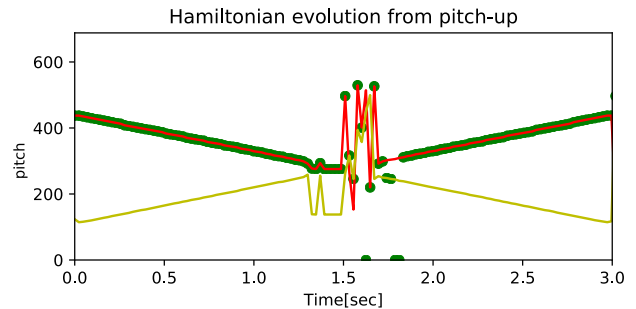
$$|\psi(t)\rangle = e^{-i \int_0^t \mathbf{H}(\tau) d\tau} |u\rangle = \mathbf{U}(0, t) |u\rangle. \quad (6)$$

Coarticulation is what allows motion, including vocal production, to be considered as a connected flow. The QVTS considers sound as a flow in the phon space, where attention and context drive the formation of trajectories, or auditory streams. There is no phon state for which the expectation of the three components  $x$ ,  $y$ , and  $z$  is zero. In other words, “There’s no such a thing as silence” [31].

The trajectories traverse the perceptual field and leave traces, or shape images that represent both the sound-producing body motion and the corresponding sound [32], at the different timescales encompassed by the sound object. At the large timescale that defines the context of habits, training and experiences, we may consider the phenomenal field [33], which may also be made to affect the local Hamiltonian and, in turn, the resulting phon trajectories and gestural interactions.

## 4 Relations

Each of the principal axes of the phon space expresses two polarities that make only relational sense. The  $z$ -axis, that in quantum computing would host the computational basis, in the phon space corresponds to pitch-up and pitch-down, that are states that express the pitch relation between a couple of sound objects. In Schaefferian terms, the calibration of the pitch dimension in QVTS is ordinal rather than cardinal. The noise ( $x$ ) and pulsation ( $y$ ) dimensions are treated in a similar, relational fashion. This framework is convenient because it affords the description of streaming phenomena, as they emerge in auditory scene analysis. The  $z$ -axis is agnostic to the details of harmonic morphology, as long as it supports a relation of order between pitches in a couple. Multiple pairwise relations between pitches can be handled through multiple qubits, or phons.



**Figure 1: Hamiltonian evolution of the phon from pitch-up. The red line indicates pitch-up; The yellow line indicates pitch-down. The green dots represent the state at discrete time instants, being turbulent when on the bottom horizontal boundary. Reproduced from [5].**

Figure 1 shows an example evolution of the phon state from pitch-up, for two gliding and crossing tones, interrupted by a 200ms-band of white noise, intervening at time 1.5s. The two crossing lines are given by the two most salient pitches, as derived from short-time Fourier analysis. Pitch attribution becomes wiggly when the tones get close to each other, and even more when they are interrupted by noise. The phon sticks to phonation until well inside the noise band, with occasional switches to turbulence. The phon evolution is still uncertain after the noise interruption, until the tones are sufficiently far from each other, and the phon takes a stable pitch-up. Such uncertainty on noise location is consistent with the known perceptual fact that bursts of noise overlapped to a tone transition are not

precisely located, with errors that can be up a few hundred milliseconds [34]. The phon evolution represents the listening process, going through the relation between the tones, and the relation between tones and noise, and drawing a trajectory in the perceptual field.

The quantal view of sound objects is applicable to a substrate that is made of discrete elementary events, that we may call onsets, or pulses. In this domain, the coherent gestalts that emerge at the temporal meso-scale can be called rhythmic objects [35, 36, 37]. Humans (and animals) appear to be analyzers of time as a dimension of and between things, in a contextual relationship with a range of differences in sound qualities. The phases of chunking, segmentation (serial integration, ordinal contour), and segregation (parallel representation, streaming) can be distinguished as fan-shaped sectors of a plane, whose abscissa is the contextual inter-onset interval at the sub-sound-object scale, and the ordinate is the considered sound quality, such as pitch, on an interval scale [38, 39].

Similarly to tones along the  $z$  axis, we can consider the relation between objects along the pulsation ( $y$ ) axis. Think of two concurrent metronomes, beating time at generally-different tempi. As long as the metronome pulses are auditorily distant, by timbre or spatial location, the phon can be made to evolve by taking pulsation-fast ( $|f\rangle$ ) or pulsation-slow ( $|s\rangle$ ) states, corresponding to focusing on one source or the other. Or, for specific tempo ratios, the listener may grasp the overall cyclic repetition of pulses, as a rhythm. For example, a three against four polyrhythm may be treated along the  $y$  axis, with a listening process that follows the evolution of a phon, possibly switching between the fast ( $|f\rangle$ ) and slow ( $|s\rangle$ ) orthogonal states. The rhythmic object emerges when it is equally likely, in such evolution, to measure one of the two states, “with concurrent rhythm motion layers becoming a single-layer motion”, and a polyrhythm becoming “a monophonic series of durations” [35]. Listening establishes a relation, of probabilistic nature, between  $|f\rangle$  and  $|s\rangle$ . “Relationships are not ends but means, which allow the event to emerge” [40].

## 5 Trajectories

A relational view is at the basis of the RhyGlyph visualization, that embeds the idea that rhythmic patterns reveal objects in motion, so that the relation between concurrent sequences of inter-onset-intervals becomes a trajectory in spacetime [41, 42]. The pairwise interactions can be represented as outwards-going trajectories within defined sectors of the plane, each corresponding to the relation between two sounding objects, as “polyrhythm does not so much generate acoustic space as show us how one might move through it” [43]. At a glance, a compact glyph can show all three relations between three sounding objects, such as Kick, Snare, and Hi-hat of a drumset. More sounding objects can tessellate the plane, with redundant trajectories (see fig. 2). Within a sector delimited by two distinct metronomes, a bisecting line would represent the trajectory of iso-temporal sequences of pulses, while a three-

against-four polyrhythm would give rise to a trajectory that is tilted towards the faster tempo. This representation can be phase blind (as in a quantum measurement), although the relative phase can be introduced by visualizing the local push-pull dynamics between the two sounding objects. Once rhythms are actively perceived as concrete movements [33], a form of embodied interaction by direct manipulation of spacetime trajectories becomes possible, as perception of motion enables the interaction gestalt of a rhythmic object [44].

The RhyGlyph trajectories can be interpreted as showing the probabilistic interdependence between two variables, as in relational quantum mechanics, where variables gain meaningful values only at discrete moments (i.e., the onsets), and the value of each variable is always relative to the interacting system [13]. The garden of forking paths, that would fill the relational sectors of figure 2, represents quantum rhythmic listening, or in a generative perspective, quantum improvisation as it was envisioned by Pauline Oliveros, as a mean to find “new ways to express and understand the relationship between mind and matter” [45, 46].

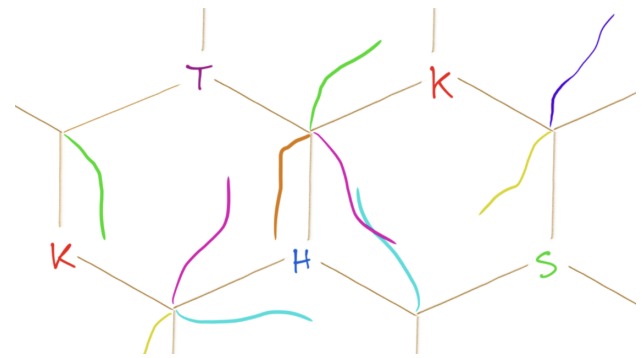


Figure 2: Sketch of five sounding objects in a honeycomb tessellation, with trajectories of pairwise rhythmic relations.

## Acknowledgment

I thank Çağrı Erdem for his work on rhythmic trajectories and for discussing the contents of this paper. Research on trajectories was funded by NextGenerationEU, Mission 4 Component 1 (CUP B53D23013150006).

## References

- [1] Davide Rocchesso and Maria Mannone. Embryo of a quantum vocal theory of sound. In *Proc. XXII Colloquium Mus. Inform.*, Udine, Italy, 2018. AIMI.
- [2] Davide Rocchesso and Maria Mannone. A quantum vocal theory of sound. *Quantum Information Processing*, 19(9):292, 2020.
- [3] Jorge Luis Borges. *El Jardín de senderos que se bifurcan*. Editorial Sur, 1941.
- [4] Maria Mannone and Davide Rocchesso. *Sing and Measure: Sound as Voice as Quanta*, pages 297–312. Springer International Publishing, Cham, 2022.
- [5] Maria Mannone and Davide Rocchesso. *Quanta in Sound, the Sound of Quanta: A Voice-Informed Quantum Theoret-*

- ical Perspective on Sound*, pages 193–226. Springer International Publishing, Cham, 2022.
- [6] Bogusław Fugiel. Quantum-like melody perception. *J. of Mathematics and Music*, 17(2):319–331, 2023.
- [7] Bogusław Fugiel. Quantum approach to jam session. *J. of Mathematics and Music*, 18(3):296–310, 2024.
- [8] Bas Versteegen, Elif Özcan, and Stefano Delle Monache. Quantum mechanics, ambiguity and design: Towards a framework. In *Proc. of the 14th Conference on Creativity and Cognition*, pages 575–582, New York, NY, USA, 2022. ACM.
- [9] Tony Verma, Scott Levine, and Teresa Meng. Transient Modeling Synthesis: a flexible analysis/synthesis tool for transient signals. In *Proc. International Computer Music Conference*, pages 164–167, Thessaloniki, Greece, 1997.
- [10] Scott N. Levine and Julius O. Smith III. *A Compact and Malleable Sines+Transients+Noise Model for Sound*, pages 145–174. Springer New York, New York, NY, 2007.
- [11] Leonardo Fierro and Vesa Välimäki. Enhanced fuzzy decomposition of sound into sines, transients, and noise. *J. of the Audio Engineering Society*, 71:468–480, July 2023.
- [12] Jonathan Driedger, Meinard Müller, and Sascha Disch. Extending harmonic-percussive separation of audio signals. In *15th International Society for Music Information Retrieval Conference, 2014. ISMIR, 2014*.
- [13] Carlo Rovelli. The relational interpretation. In *The Oxford Handbook of the History of Quantum Interpretations*. Oxford University Press, 03 2022.
- [14] Jordi Bonada, X Serra, X Amatriain, and A Loscos. Spectral processing. In Udo Zölzer, editor, *DAFX: Digital Audio Effects*, pages 393–445. John Wiley & Sons, Ltd, 2011.
- [15] Michael Kubovy and David Van Valkenburg. Auditory and visual objects. *Cognition*, 80(1):97–126, 2001.
- [16] David Van Valkenburg and Michael Kubovy. In defense of the theory of indispensable attributes. *Cognition*, 87(3):225–233, 2003.
- [17] Pierre Schaeffer. *Traité des objets musicaux*. Editions du Seuil, Paris, 1977 edition, 1966.
- [18] Michel Chion. *Guide des objets sonores: Pierre Schaeffer et la recherche musicale*. Recherches Musicales. INA-GRM/Buchet-Chastel, Paris, 1983.
- [19] Rolf Inge Godøy. Gestural-Sonorous Objects: embodied extensions of Schaeffer’s conceptual apparatus. *Organised Sound*, 11(2):149–157, 2006.
- [20] Rolf Inge Godøy. Generic motion components for sonic design. In Alexander Refsum Jensenius, editor, *Sonic Design*, pages 3–24, Cham, 2024. Springer Nature Switzerland.
- [21] Lasse Thoresen and Andreas Hedman. Spectromorphological analysis of sound objects: an adaptation of Pierre Schaeffer’s typomorphology. *Organised Sound*, 12(2):129–141, 2007.
- [22] Andrea Valle. Towards a semiotics of the audible. *Signata. Annales des sémiotiques/Annals of Semiotics*, (6):65–89, 2015.
- [23] Rolf Inge Godøy. Quantal elements in musical experience. In Rolf Bader, editor, *Sound - Perception - Performance*, pages 113–128. Springer International Publishing, Heidelberg, 2013.
- [24] Pierre Schaeffer. *À la recherche d’une musique concrète*. Éditions du Seuil, 1952.
- [25] Giovanni Bruno Vicario. Temporal displacement. In Rosolino Buccheri, Metod Saniga, and William Mark Stuckey, editors, *The Nature of Time: Geometry, Physics and Perception*, pages 53–66. Springer Netherlands, Dordrecht, 2003.
- [26] Davide Rocchesso, Guillaume Lemaitre, and Massimo Grassi. Evidence for a spatial bias in the perception of sequences of brief tones. *The J. of the Acoustical Society of America*, 133(5):EL346–EL350, 04 2013.
- [27] Giovanni B. Vicario. La scissione degli eventi sonori. *Rivista di Psicologia*, 54(1):1–28, 1960.
- [28] Albert S. Bregman. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA, 1990.
- [29] Curtis Roads. The perception of microsound and its musical implications. *Annals of the NY Ac. of Sciences*, 999(1):272–281, 2003.
- [30] Stuart Klapp and Richard Jagacinski. Gestalt principles in the control of motor action. *Psychological Bull.*, 137(3):443–462, 2011.
- [31] John Cage. *Silence: Lectures and Writings*. Wesleyan University Press, Middletown, CT, 1961.
- [32] Rolf Inge Godøy. Motion shapes for sound shaping. *Frontiers in Psychology*, Volume 15, 2024.
- [33] Dag Svanæs. Interaction design for and with the lived body: Some implications of Merleau-Ponty’s phenomenology. *ACM Trans. Computer-Human Interaction*, 20(1), April 2013.
- [34] Giovanni Vicario. La “dislocazione temporale” nella percezione di successioni di stimoli discreti. *Rivista di Psicologia*, 57(1):17–87, 1963.
- [35] Rolf Inge Godøy. Thinking rhythm objects. *Frontiers in Psychology*, Volume 13, 2022.
- [36] Çağrı Erdem. *Controlling or Being Controlled? Embodiment, Agency and Artificial Intelligence in Interactive Music Performance*. PhD thesis, University of Oslo, 2022.
- [37] Çağrı Erdem. Exploring musical agents with embodied perspectives. In Alexander Refsum Jensenius, editor, *Sonic Design*, pages 321–341, Cham, 2024. Springer Nature Switzerland.
- [38] Mari Riess Jones. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological rev.*, 83(5):323, 1976.
- [39] Massimiliano Viel. *Listening Patterns: From Music to Perception and Cognition*. PhD thesis, University of Plymouth, 2018.
- [40] Daniel Charles. *Gloses sur John Cage*. 10/18: Série Esthétique. Union générale d’éditions, Paris, 1978.
- [41] Davide Rocchesso, Alessio Bellino, Gabriele Ferrara, and Antonino Perez. Spacetime trajectories as overlapping rhythms. *Int. J. of Human-Computer Studies*, 192:103358, 2024.
- [42] Çağrı Erdem and Rocchesso Davide. RhyGlyph: Radial glyph visualization of rhythm interactions. In *Proceedings of the International Conference on Audio Mostly*. 2025.
- [43] Erik Davis. “Roots and Wires” remix: Polyrhythmic tricks and the black electronic. In *Sound Unbound: Sampling Digital Music and Culture*. The MIT Press, 03 2008.
- [44] Lasse Scherffig. *From Action Capture to Interaction Gestalt*, pages 79–92. De Gruyter, Berlin, Boston, 2018.
- [45] Pauline Oliveros. Quantum listening: From practice to theory (to practice practice). In Sin wai Chan and Siu Tong Kwok, editors, *Culture and Humanity in the New Millennium: The Future of Human Values*, pages 27–41. Chinese University Press NT, Sha Tin, Hong Kong, 2002.
- [46] Pauline Oliveros. Quantum improvisation: The cybernetic presence. In Paul Miller, editor, *Sound Unbound: Sampling Digital Music and Culture*. The MIT Press, 2008.