An Analysis of Musical Instruments Directivity Based on Spatial Complexity

Federico Miotello*†, Juliano G. C. Ribeiro‡, Ryo Matsuda‡, Jorge Trevino‡, Mirco Pezzoli†

†Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy

{federico.miotello, mirco.pezzoli}@polimi.it

‡Yamaha Corporation, Hamamatsu, Shizuoka, Japan

{juliano.ribeiro, ryo.matsuda, jorge.trevino}@music.yamaha.com

Abstract—**Characterizing and modeling the spatial radiation of musical instruments is a challenging task in acoustics due to the complex and high-dimensional nature of their directivities. In this paper, we address this challenge by presenting a comprehensive analysis of musical instrument directivities using Spatial Complexity, a novel metric that quantitatively describes how energy is distributed across spherical harmonics coefficients. Lower Spatial Complexity values indicate a concentration of energy in lower-order spherical harmonic components, corresponding to nearly isotropic radiation patterns, while higher values reflect energy distributed across higher-order harmonics, denoting highly directional and intricate beam patterns. To demonstrate the utility of this metric, we analyze an extensive dataset of measured instrument directivities, uncovering systematic correlations between complexity and physical radiation properties. Further, we train a rotation-equivariant neural network designed to preserve the geometric symmetries of spherical harmonics representations. The model's latent space organizes directivities along gradients tightly aligned with complexity, showing the metric's ability to capture fundamental attributes of spatial radiation. These findings highlight the effectiveness of complexity as a concise and interpretable analytical tool for comparing and categorizing instrument directivities, with potential applications in acoustic research, instrument design, and spatial audio technologies.**

Index Terms—**directivity, spatial complexity, spherical harmonics.**

I. INTRODUCTION

Directivity describes how an acoustic source emits sound in space, outlining the variation in sound intensity with respect to different directions. In the context of musical instruments, the directivity pattern significantly influences both the perceived sound quality and the effectiveness of spatial audio reproduction [1]–[3]. Traditionally, researchers have focused on capturing high-fidelity directivity measurements in controlled settings using anechoic chambers [4], [5] or near-field acoustic holography with scanning microphone arrays [6]–[8]. Alternative procedures suitable for low-reverberant settings have also been introduced in [9]. Perceptual studies [10] have shown that listeners can discern differences between omnidirectional and directional sources. Moreover, fluctuations in directivity due to musician movements have been found to affect listener perception in both anechoic and reverberant conditions in [11], while in [12] the significance of capturing frequency-

dependent directivities over simplified average representations is demonstrated.

More recently, the increasing demand of spatial audio and immersive technologies [13], [14] has renewed the interest in modeling and characterizing the directivity. In fact, the modeling of directional sound sources is required for enhanced sound field reconstruction for immersive scene navigation [15], [16], and different simulation frameworks incorporate directivity through boundary and finite element methods (FEM) [17], numerical simulations [18], and geometrical acoustics [19].

Nonetheless, systematic and objective characterization of directivities of musical instruments remains a challenging task. Typical methods of analysis are limited to visual comparison or simple metrics as in [20], which presents a major dataset comprising directivities of forty-one orchestral instruments. In [21], the instruments have been measured during live performances by musicians, showing that the presence of the performer's body tends to smooth directivity patterns. However, while [21] offers valuable visual insights into these patterns, its evaluation remains largely confined to graphical analysis without a systematic characterizations. Thus, in order to compare the directivities several metrics based on spherical harmonics correlations [22], [23] and more recent composite measures [24], [25] have been proposed to encapsulate the multifaceted nature of instrument directivity.

In addition, deep learning has recently driven significant advances in audio processing [14], [26]–[29], particularly through feature-learning architectures that extract low-dimensional latent spaces representations from raw signals, enabling a broad class of applications [26], [27]. When applied to musical instrument directivities, data-driven methods could reveal latent features that are not evident from traditional directivity analyses. However, the development of succinct, robust metrics to compare directivities within these learned representations remains an open challenge. Establishing such metrics is pivotal for evaluating how faithfully latent spaces capture complex directivity patterns, especially when comparing multiple instruments, playing conditions, or performance nuances.

In this paper, we propose a novel metric called Spatial Complexity, which captures how acoustic energy is distributed among spherical harmonic coefficients. Through an extensive analysis on a dataset of measured instrument directivities

[20], we show that Spatial Complexity correlates closely with fundamental radiation properties, thus providing a concise and interpretable descriptor. To further demonstrate its utility, we train a rotation-equivariant neural network [30] designed to handle spherical harmonics representations. The model's latent space naturally arranges instrument directivities along gradients that align tightly with Spatial Complexity, showing the metric's capacity to encapsulate essential features of spatial radiation. Overall, our findings reveal Spatial Complexity as a powerful analytical tool for comparing and categorizing instrument directivities, with broad applications in acoustic research, instrument design, and emerging spatial audio technologies.

II. SIGNAL MODEL

As defined in [9], let us consider an acoustic source, whose acoustic center (i.e., the effective origin of radiation) is placed at the center of a three-dimensional coordinate system. Consider a field point far from the source and expressed in spherical coordinates by the triplet $\mathbf{r} = (\phi, \theta, r)$, where r is the radial distance from the acoustic center, ϕ is the azimuth and θ is the co-elevation. Under a far-field approximation (i.e., r is sufficiently large) the source's free-field pressure $P_f(\mathbf{r}, \omega)$ at angular frequency ω can be written as [31]

$$P_f(\mathbf{r}, \omega) = \frac{e^{-j\frac{\omega}{c}r}}{r} S(\omega) D(\phi, \theta, \omega), \qquad (1)$$

where $S(\omega)$ is the Fourier transform of the source signal at the acoustic center, and $D(\phi, \theta, \omega) \in \mathbb{L}^2(\mathbb{S}^2)$ a complex square-integrable function describing the source directivity [32] depending on direction (ϕ, θ) and frequency ω. The magnitude $|D(\phi, \theta, \omega)|$ of the directivity function, which describes the energy distribution of the directional sound radiation, is commonly referred to as *radiation pattern* or *polar pattern*. Provided that the source signal $S(\omega)$ and the measurement position $\mathbf{r}$ are known, the polar pattern can be obtained by inverting Eq. (1) as

$$|D(\phi, \theta, \omega)| = r \frac{|P_f(\mathbf{r}, \omega)|}{|S(\omega)|}. \qquad (2)$$

It is worth noting that the signal model relies on the assumption that the source can be approximated as a point-like emitter and holds if the acoustic source exhibits a linear behavior with respect to the energy of the excitation signal $S(\omega)$ [9].

Given its definition, the polar pattern in Eq. (2) can be conveniently represented in an orthonormal basis on the unit sphere. A commonly employed choice is the set of spherical harmonics SHs $\{Y_n^m(\phi, \theta)\}$, which can represent any arbitrary function on a sphere [31]. In particular, one may expand the polar pattern, by means of the SH expansion [31], as

$$|D(\phi, \theta, \omega)| = \sum_{n=0}^{N} \sum_{m=-n}^{n} C_n^m(\omega) Y_n^m(\phi, \theta), \qquad (3)$$

where $Y_n^m(\phi, \theta)$ are the real (or complex) SH functions of degree n and order m, $C_n^m(\omega)$ are frequency-dependent expansion coefficients, and N is the maximum expansion order. Without loss of generality, we employ real SHs [33] in this work. From Eq. (3), it follows that $(N+1)^2$ coefficients are required to fully characterize a directivity of order N.

III. SPATIAL COMPLEXITY OF SOURCE DIRECTIVITY

Building on the signal model and directivity framework presented in Section II, and similarly to [34], we now introduce a concise, quantitative metric that characterizes the spatial distribution of acoustic energy in directionality patterns. We call this measure the *Spatial Complexity* (SC). Conceptually, SC condenses the level of directional detail required to characterize a radiation pattern into a single scalar. In doing so, it facilitates a compact comparison between instruments or sources that would otherwise require detailed inspection of the complete directivity function. Formally, we define SC at angular frequency ω as

$$\text{SC}(\omega) = \frac{\sum_{n=0}^{N} \lambda_n \|\mathbf{C}_n(\omega)\|_2}{\sum_{n=0}^{N} \|\mathbf{C}_n(\omega)\|_2}, \qquad (4)$$

where $\mathbf{C}_n(\omega) \in \mathbb{R}^{2n+1}$ is the vector of real SH coefficients of order n, $\| \cdot \|_2$ denotes the Euclidean (ℓ_2) norm, and λ_n is a scalar weight that can emphasize (or de-emphasize) energy in each SH order. Specifically, $\|\mathbf{C}_n(\omega)\|_2$ measures the total energy at SH order n, while multiplying by λ_n in the numerator can penalize or reward higher orders more strongly, depending on the desired notion of *complexity*. Hence, $\text{SC}(\omega)$ is a dimensionless ratio that remains invariant under any global scaling or rotation of the radiation pattern. Larger values of $\text{SC}(\omega)$ indicate that a significant portion of the total radiation energy is carried by higher-order terms (i.e., more finely structured directivity), whereas smaller values suggest that the pattern is dominated by lower orders and varies more slowly with angle, indicative of a more uniform energy distribution. In the limiting case, the lower bound of $\text{SC}(\omega)$ corresponds to an ideal omnidirectional source that radiates energy isotropically.

A. Choice of weighting factors λ_n

Building on Eq. (4), it is clear that the choice of weights λ_n is crucial in defining the SC metric. In our formulation, these weights determine the relative contribution of each SH order to the overall measure of complexity. To ensure that energy concentrated in higher SH orders (which correspond to finer angular variations) yields a higher complexity value, we impose a strictly increasing sequence of weights

$$\lambda_0 < \lambda_1 < \cdots < \lambda_N. \qquad (5)$$

This condition guarantees that directivity patterns with significant energy in higher orders receive larger SC values, while patterns dominated by lower orders yield lower SC scores. Furthermore, we enforce a coherence condition

$$\lambda_M > M \lambda_{M-1} - \sum_{n=0}^{M-2} \lambda_n, \qquad (6)$$

which, while being quite strict, ensures that even the *least complex* pattern of order M is still more complex than the *most complex* pattern of order $M - 1$. By selecting all λ_n

to satisfy both the monotonicity and coherence conditions, the SC metric provides a well-defined, monotonically increasing and comparable measure of complexity that aligns with the intuitive notion that directivity patterns exhibiting finer angular detail (i.e., energy at higher SH orders) should be quantitatively distinguished from those with simpler, lower-order distributions.

Fig. 1: Histogram of SC values of directivities in [20].

(a)

(b)

Fig. 2: Directivity pattern of (a) a historical basset horn playing note B_2 (118.94 Hz) with SC = 2.75, and (b) a modern viola playing note G#$_6$ (1711.61 Hz) with SC = 8.17.

IV. ANALYSIS OF DIRECTIVITY DATA

We now focus on practical analyses of measured directivity patterns to demonstrate the utility of the SC metric. For the tests, we exploit the dataset presented in [20], which offers an extensive collection of directivity measurements from 41 modern and historical musical instruments, each recorded while playing various single notes. Captured with a 32-channel microphone array in an anechoic environment, the dataset provides high-resolution spatial radiation patterns that capture the frequency-dependent behavior of the instruments. For simplicity, our analyses focus on the polar patterns at the fundamental frequency f_0 of each note, a component that typically dominates the acoustic energy and encapsulates the instrument's core tonal characteristics. This focus offers a robust and consistent basis for comparing directivity across instruments while reducing the complexity associated with overtones harmonic content. Given the 32 measurement points available and by means of the SH expansion detailed in Eq. (3), we can fully characterize the directivities using SH coefficients up to order $N = 4$, thereby enabling the application of the SC metric to quantify the angular variability in the radiation patterns. In the following experiments, we set λ_n in order to respect both constraints defined in Eq. (5) and Eq. (6) Starting from the initial value $\lambda_0 = 1$, the other values are set to $\lambda_1 = 2$, $\lambda_2 = 4$, $\lambda_3 = 10$, $\lambda_4 = 34$.

A. Dataset analysis

As an initial analysis, we examine the distribution of SC across all the directivity patterns in the considered dataset [20]. The histogram in Fig. 1 illustrates that the majority of the SC values are clustered in the lower range. This indicates that, for most instruments, the energy distribution is predominantly confined to the lower SH orders, leading to simpler, more isotropic radiation patterns. Such behavior is typically observed in instruments that either inherently exhibit an isotropic directivity or tend to radiate more uniformly when playing lower notes [2]. For instance, Fig. 2a shows the directivity pattern of a historical basset horn playing note B_2 (118.94 Hz), yielding a SC value of 2.75. This example demonstrates how a lower SC value aligns with a relatively uniform energy distribution. In contrast, a small subset of the dataset exhibits higher SC values, which reflect directivity patterns with significant contributions from higher SH orders. This implies more pronounced angular variations and a greater degree of spatial intricacy, possibly arising from unique structural features or specific excitation conditions [2]. As an example, Fig. 2b shows the directivity pattern of a modern viola playing note G#$_6$ (1711.61 Hz), with an associated SC
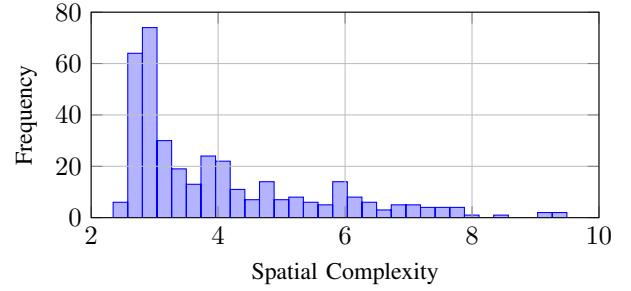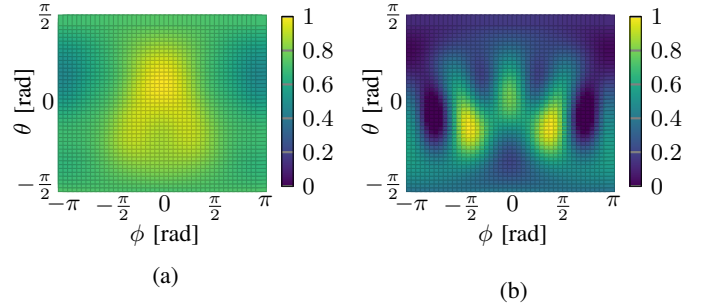
value of 8.17. This case illustrates how a higher SC value corresponds to a more complex and detailed angular distribution of radiated energy. These observations underscore the utility of the SC metric in distinguishing between instruments with fundamentally different radiation characteristics, providing a coherent and interpretable measure of directional detail across the dataset.

B. Latent space analysis

As a further study, we employed the SC metric within a feature learning framework to investigate its effectiveness in analyzing and creating a latent representation of directivity data. Inspired by the work in [35], we have chosen to employ an *equivariant autoencoder* [30], [36]. Equivariant neural networks have found widespread application in fields where the data or underlying phenomena exhibit symmetry, such as computer vision or molecular modeling [30]. An equivariant neural network is built so that transformations (such as rotations) in the input space lead to predictable, corresponding transformations in the latent feature space. This property is especially critical for directivity data, as it is naturally defined on the surface of a sphere, exhibiting inherent rotational symmetries, meaning that the physical characteristics of an acoustic source's radiation pattern are invariant to its orientation. By using an equivariant architecture, the model is better able to capture the intrinsic spatial structure of the data, ensuring that the learned features are robust to rotations while preserving physical interpretability. In our work, we exploit the architecture proposed in [30], namely an SO(3)-equivariant autoencoder. This model is designed to preserve equivariance with respect to the group of three-dimensional
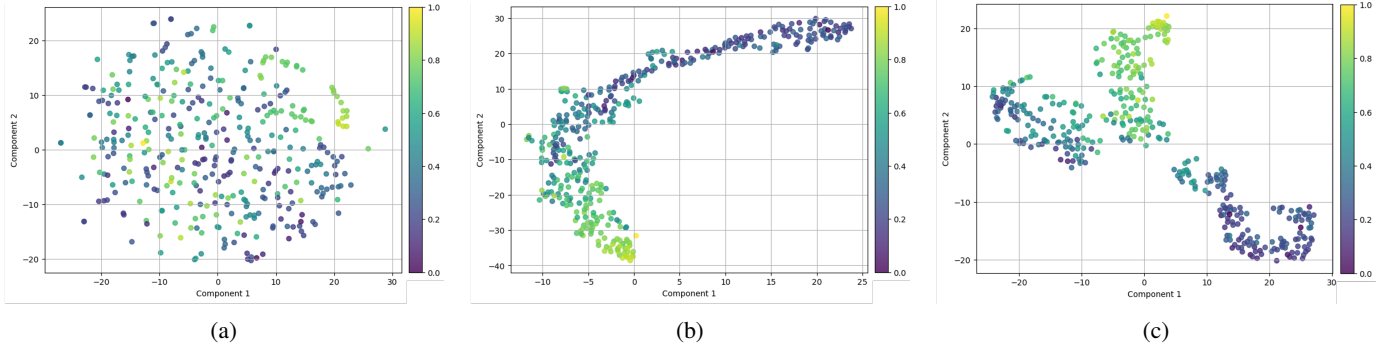
Fig. 3: 2D t-SNE representations of the latent space (a) before training, (b) after training with $\mathcal{L}_{data}$, and (c) after training with $\mathcal{L}_{SC}$. Colors indicate the normalized SC value of each data sample in the test set.

rotations, denoted by SO(3), consisting of all rotations about the origin in the three-dimensional Euclidean space.

1) Data loss: As a first experiment, we evaluated the learned representations of the employed SO(3)-equivariant autoencoder architecture [30]. The model was first trained on a designated portion of the dataset, where the input consisted of the SH coefficients representing the directivity patterns. At each layer, the data is characterized by $(N + 1)^2$ coefficients and M multiplicities (similar to the channel dimension in convolutional neural networks), and the encoding layers iteratively and equivariantly transfer information from higher orders to lower ones, with the decoding layers performing the reverse. As a result, the network learns a so-called disentangled latent space, consisting of an invariant embedding which captures the intrinsic properties of the directivity pattern independent of orientation, and an equivariant frame that describes the data orientation. The loss function used for training is

$$\mathcal{L}_{data} = \mathrm{MSE}(\mathbf{C}, \hat{\mathbf{C}}), \qquad (7)$$

where MSE is the mean squared error operator, $\mathbf{C}$ are the coefficients of the SH representation of the directivity pattern in input to the network and $\hat{\mathbf{C}}$ are the coefficients of the SH representation of the reconstructed directivity pattern in ouput to the network. After training, we extracted the latent features for samples in the validation set. To visualize the distribution of these representations, we applied the well-known t-SNE [37] dimensionality reduction technique to the latent space. The resulting 2D t-SNE projection illustrates how the validation samples are positioned relative to each other in the learned feature space, providing insight into whether the model has successfully captured the intrinsic structure of the directivity data. Fig. 3a displays the distribution of test data samples in the latent space before any training is applied. The colormap, normalized for clarity, represents the SC value for each sample. At this stage, the latent space does not exhibit any discernible organization related to SC, indicating that the raw features do not yet capture the intrinsic structure of the directivity patterns. In contrast, Fig. 3b shows the latent space after 2000 training epochs. The same normalized colormap is used to denote the SC values, but now the data points are clearly organized according to their SC. This

structured arrangement demonstrates that the training process has successfully embedded the SC information into the latent representation, making it easier to distinguish between patterns based on their SC. This result shows that naturally the training process effectively organizes the latent space according to the SC metric, making it easier to distinguish between different directivity patterns.

2) Spatial Complexity loss: As a second experiment, we tried to further embed spatial information into the latent representation by adding a SC loss term in the loss function used to train the model. In particular, the loss function is

$$\mathcal{L}_{SC} = \alpha\,\mathrm{MSE}(\mathbf{C}, \hat{\mathbf{C}}) + \beta\,\mathrm{MSE}(\mathrm{SC}(\mathbf{C}), \mathrm{SC}(\hat{\mathbf{C}})), \quad (8)$$

where $\alpha = 100$ and $\beta = 10$ are two weighting factors. By directly penalizing discrepancies in SC, the network is encouraged to learn latent representations that not only capture overall acoustic features but also faithfully reflect the detailed angular variations that are critical for distinguishing between different instruments. The obtained latent space after training is represented in Fig. 3c. It is clear that the data samples are clustering with respect to SC in the latent space, indicating that incorporating $\mathcal{L}_{SC}$ in the loss function, effectively guides the model to capture the intrinsic variations in directivity patterns. This results in similar SC values grouping together, which not only makes the latent representation more interpretable but also enhances the model's ability to differentiate and analyze the acoustic characteristics of different instruments.

V. CONCLUSION

In this work, we introduced a novel metric, namely SC, to quantify the directional variability of acoustic radiation patterns. By decomposing the source directivity into SH components and appropriately weighting them, the SC metric condenses the degree of directional detail required to characterize a radiation pattern into a single, interpretable scalar. We conducted a first analysis on directivity data from a diverse set of musical instruments, computing the distribution of SC inside the data set and showing that most instruments exhibit predominantly lower-order radiation characteristics, while a select few display marked complexity with significant contributions from higher-order terms. Next, we assessed SC

in a deep-learning context by exploiting an SO(3)-equivariant autoencoder and examining how its latent representations capture intrinsic spatial features. In our first analysis, where the model was trained solely on a data loss, t-SNE visualizations revealed that the latent space naturally organized itself according to SC values. In our second analysis, embedding the SC metric directly into the loss function resulted in even stronger clustering, indicating that explicitly encouraging the network to preserve spatial complexity significantly enhances the separation of directivity patterns in the latent space. These complementary analyses confirm that the SC metric is not only inherently captured during training but also serves as a powerful guide for learning more structured and interpretable representations. Future work will extend these findings by incorporating multi-frequency analyses and a rigorous study of the impact of weighting factors λ_n in SC.

REFERENCES

[1] H. F. Olson, *Music, physics and engineering*, vol. 1769. Courier Corporation, 1967.

[2] J. Meyer, "Directivity of the bowed stringed instruments and its effect on orchestral sound in concert halls," *The J. Acoust. Soc. Am.*, vol. 51, no. 6B, pp. 1994–2009, 1972.

[3] J. Meyer, "The influence of the directivity of musical instruments on the efficiency of reflecting or absorbing areas in proximity to the orchestra," *Acta Acustica united with Acustica*, vol. 36, no. 3, pp. 147–161, 1976.

[4] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 856–865, 2008.

[5] J. Curtin, "Measuring violin sound radiation using an impact hammer," *J. Violin. Soc. Am. VSA Pap. XXII*, no. 1, pp. 186–209, 2009.

[6] J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. theory of generalized holography and the development of NAH," *The J. Acoust. Soc. Am.*, vol. 78, no. 4, pp. 1395–1413, 1985.

[7] D. Fernandez Comesana, T. Takeuchi, S. Morales Cervera, and K. R. Holland, "Measuring musical instruments directivity patterns with scanning techniques," in *25th International Congress on Sound and Vibration, ICSV 2019*, 2019.

[8] X. Luan, M. Olivieri, M. Pezzoli, F. Antonacci, and A. Sarti, "Complex-valued physics-informed neural network for near-field acoustic holography," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, pp. 126–130, IEEE, 2024.

[9] A. Canclini, F. Antonacci, S. Tubaro, and A. Sarti, "A methodology for the robust estimation of the radiation pattern of acoustic sources," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 28, pp. 211–224, 2020.

[10] L. M. Wang and M. C. Vigeant, "Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities," *Appl. Acoust.*, vol. 69, no. 12, pp. 1281–1293, 2008.

[11] D. Ackermann, C. Böhm, F. Brinkmann, and S. Weinzierl, "The acoustical effect of musicians' movements during musical performances," *Acta Acustica united with Acustica*, vol. 105, no. 2, pp. 356–367, 2019.

[12] A. Corcuera Marruffo and V. Chatziioannou, "A pilot study on tone-dependent directivity patterns of musical instruments," in *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*, Audio Engineering Society, 2022.

[13] J. G. Tylka and E. Y. Choueiri, "Fundamentals of a parametric method for virtual navigation within an array of ambisonics microphones," *J. Audio Eng. Soc*, vol. 68, no. 3, pp. 120–137, 2020.

[14] M. Cobos, J. Ahrens, K. Kowalczyk, and A. Politis, "An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction," *EURASIP J. on Audio, Speech, Music. Process.*, vol. 2022, no. 1, p. 10, 2022.

[15] M. Pezzoli, F. Borra, F. Antonacci, A. Sarti, and S. Tubaro, "Reconstruction of the virtual microphone signal based on the distributed ray space transform," in *26th European Signal Processing Conference (EUSIPCO)*, pp. 1537–1541, IEEE, 2018.

[16] M. Pezzoli, F. Borra, F. Antonacci, S. Tubaro, and A. Sarti, "A parametric approach to virtual miking for sources of arbitrary directivity," *IEEE Trans. on audio, speech, language Process.*, vol. 28, pp. 2333–2348, 2020.

[17] R. Mehra, L. Antani, S. Kim, and D. Manocha, "Source and listener directivity for interactive wave-based sound propagation," *IEEE transactions on visualization computer graphics*, vol. 20, no. 4, pp. 495–503, 2014.

[18] J. Ahrens and S. Bilbao, "Computation of spherical harmonic representations of source directivity based on the finite-distance signature," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 29, pp. 83–92, 2021.

[19] J. Klein and M. Vorländer, "Simulative investigation of required spatial source resolution in directional room impulse response measurements," in *EAA Spatial Audio Signal Processing Symposium*, pp. 37–42, 2019.

[20] S. Weinzierl, M. Vorländer, G. Behler, F. Brinkmann, H. von Coler, E. Detzner, J. Krämer, A. Lindau, M. Pollow, F. Schulz, and N. R. Shabtai, "A database of anechoic microphone array measurements of musical instruments," 2017.

[21] J. Pätynen and T. Lokki, "Directivities of symphony orchestra instruments," *Acta Acustica united with Acustica*, vol. 96, no. 1, pp. 138–167, 2010.

[22] F. Hohl and F. Zotter, "Similarity of musical instrument radiation-patterns in pitch and partial," *Fortschritte der Akustik, DAGA, Berlin*, 2010.

[23] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, pp. 20–23, 2006.

[24] T. Carpentier and A. Einbond, "Spherical correlation as a similarity measure for 3d radiation patterns of musical instruments," in *16ème Congrès Français d'Acoustique*, HAL Open Science, 2022.

[25] M. Pezzoli, A. Canclini, F. Antonacci, and A. Sarti, "A comparative analysis of the directional sound radiation of historical violins," *The J. Acoust. Soc. Am.*, vol. 152, no. 1, pp. 354–367, 2022.

[26] A. Défossez, J. Copet, G. Synnaeve, and Y. Adi, "High fidelity neural audio compression," *Trans. on Mach. Learn. Res.*, 2023.

[27] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 30, pp. 495–507, 2021.

[28] F. Miotello, M. Pezzoli, L. Comanducci, F. Antonacci, and A. Sarti, "Deep prior-based audio inpainting using multi-resolution harmonic convolutional neural networks," *IEEE/ACM Trans. on Audio, Speech, Lang. Process.*, vol. 32, pp. 113–123, 2023.

[29] F. Miotello, L. Comanducci, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, "Reconstruction of sound field through diffusion models," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1476–1480, IEEE, 2024.

[30] G. M. Visani, M. N. Pun, A. Angaji, and A. Nourmohammad, "Holographic-(v)ae: An end-to-end so(3)-equivariant (variational) autoencoder in fourier space," *Phys. review research*, vol. 6, no. 2, p. 023006, 2024.

[31] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. Elsevier, 1999.

[32] M. Pezzoli, R. Malvermi, F. Antonacci, and A. Sarti, "Similarity evaluation of violin directivity patterns for musical instrument retrieval," in *Proceedings of the 24th International Society for Music Information Retrieval Conference ISMIR 2023*, pp. 138–145, 2023.

[33] J. Ahrens and S. Bilbao, "Interpolation and range extrapolation of sound source directivity based on a spherical wave propagation model," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4662–4666, IEEE, 2020.

[34] T. Tanaka and M. Otani, "An isotropic sound field model composed of a finite number of plane waves," *Acoust. Sci. Technol.*, vol. 44, no. 4, pp. 317–327, 2023.

[35] X. Chen, F. Ma, Y. Zhang, A. Bastine, and P. N. Samarasinghe, "Head-related transfer function interpolation with a spherical cnn," *arXiv preprint arXiv:2309.08290*, 2023.

[36] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *International conference on machine learning*, pp. 2990–2999, PMLR, 2016.

[37] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne.," *J. machine learning research*, vol. 9, no. 11, 2008.