

# Intelligent Networked Music Performance Experiences



Luca Comanducci 

**Abstract** A Networked Music Performance (NMP) is defined as what happens when geographically displaced musicians interact together while connected via network. The first NMP experiments begun in the 1970s, however, only recently the development of network communication technologies has created the necessary infrastructure needed to successfully create an NMP. Moreover, the widespread adoption of network-based interactions during the COVID-19 pandemic has generated a renewed interest towards distant music-based interaction. In this chapter we present the Intelligent networked Music PERforMANce experiENCEs (IMPERMANENCE) as a comprehensive NMP framework that aims at creating a compelling performance experience for the musicians. In order to do this, we first develop the neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE) framework in order to understand which are the main needs of the participants in a NMP. Informed by these results we then develop IMPERMANENCE accordingly.

**Keywords** Networked music performance · Deep learning · Spatial audio · Microphone array

## 1 Introduction

During the last decades, the increasing speed of network communication technologies has grown steadily, creating a fertile environment for the development and consumption of online media-content. Moreover, the recent COVID-19 pandemic made online communications, such as live-streaming or online meetings, a part of

---

L. Comanducci (✉)  
Politecnico di Milano, Milan, Italy

Dipartimento di Elettronica, Informazione e Bioingegneria, Via Ponzio 34/5, 20133 Milano, Italy  
e-mail: [luca.comanducci@polimi.it](mailto:luca.comanducci@polimi.it)

© The Author(s) 2023  
C. G. Riva (ed.), *Special Topics in Information Technology*,  
PoliMI SpringerBriefs, [https://doi.org/10.1007/978-3-031-15374-7\\_10](https://doi.org/10.1007/978-3-031-15374-7_10)

daily lives for a considerable amount of people. In this context, also the online fruition of music content has grown, both for what concerns listening, either streaming or remote-concert attendance, or performing, e.g. rehearsals or remote live-performances. We define a performance where musicians are connected via network as a Networked Music Performance (NMP). The research in NMPs has begun as early as the 1970s [7], but only recently it has grown enough to be actually a viable possibility for geographically displaced musicians. In order to create a NMP experience that feels compelling to the musicians, several aspects need to be considered, which can be broadly separated into temporal factors, related to the synchronicity between the musicians, and spatial factors, related to the audiovisual perception. Several NMP-dedicated software solutions have been created over the years [25, 31], such as LOLA [24], UltraGrid [27] or JackTrip [9]. Basically, all these software frameworks are low-latency audio/video streaming protocols. While the minimization of the latency is extremely important in creating a satisfying NMP, it does not suffice in creating an experience that feels compelling from the point of view of the musicians.

In this chapter, we present a general-purpose framework for NMP denoted Intelligent networked Music PERforMANce experiENCEs (IMPERMANENCE). In modeling this framework, we do not aim at creating a NMP that hopelessly tries to mimic the hypothetical real environment, instead we try to model the NMP as a medium on its own. In order to do this we build the components of IMPERMANENCE by first analyzing which are the needs of the musicians, by developing a research framework, denoted neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE). Following the guidelines established through TEMPERANCE, we organize a series of experiments aimed at understanding the impact of temporal and spatial factors on the perceived Quality of Experience (QoE) of the musicians. Informed by the results obtained through these experiments we develop IMPERMANENCE accordingly. We take advantage of both signal processing, well suited to treat audio signals, and deep learning techniques. The latter allow us to overcome limitations posed by classical signal processing techniques, related to the number of sensors needed and to the decrease in the performance quality in adverse scenarios, i.e. where reverberation and noise are present.

The research described in this chapter summarizes the results obtained in my Ph.D. thesis [12] and the scientific publications in [3, 4, 11, 13–22, 30], obtained during my Ph.D. The rest of this chapter is organized as follows. In Sect. 2 we present the preliminary analysis aimed at understanding which are the main aspects of a NMP that need to be tackled by the IMPERMANENCE framework. Specifically, we first present the research framework TEMPERANCE and two experimental studies performed with musicians. In Sect. 3 we present the general structure of the IMPERMANENCE framework, while in Sects. 4 and 5, we present how we plan to treat the temporal and spatial factors, respectively. In Sect. 6 we present preliminary results related to a technique that could provide audio compression. Finally, in Sect. 7 we draw some conclusions.

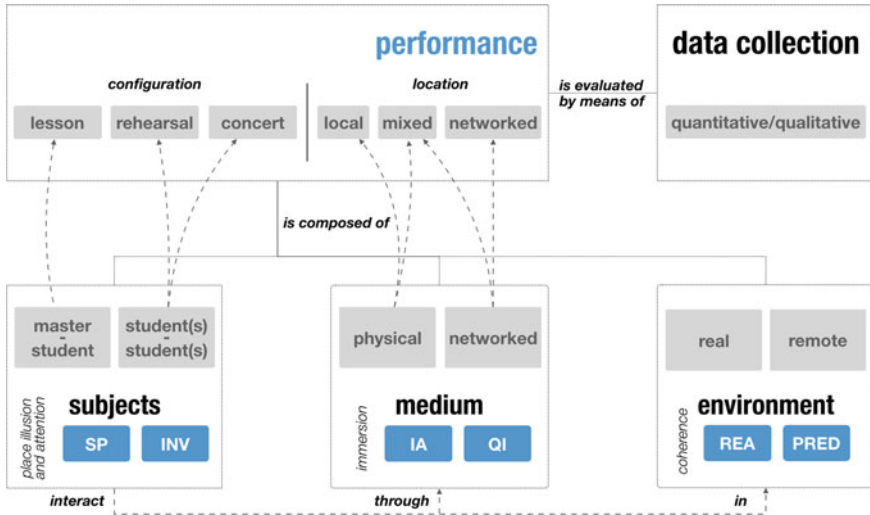


Fig. 1 The TEMPERANCE framework for NMP research [21]

## 2 The neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE) Framework

The TEMPERANCE framework [21] is devoted to the design and realization of NMP experiments and scenarios. While the framework is general enough to be applicable to a wide variety of music genres and types of remote performances, it was developed focusing on the remote teaching of chamber music, during the INTERactive environment for MUSIC learning and practising (INTERMUSIC)<sup>1</sup> project.

The general structure of the TEMPERANCE framework is depicted in Fig. 1. A **performance** may be defined as what occurs when two or more **subjects** interact together through a **medium** in a certain **environment**. The performance is the entity that stands at the highest conceptual level and can assume two main configurations, namely a *performed music composition*, or a *taught lesson*. It can either be a *rehearsal* or a *concert*, both involving the inclusion of at least two musicians and possibly additional people. Depending on the location, when the subjects are in the same room we have a *local performance*, when they are geographically displaced we have a *networked performance*, that is the focus of this chapter; finally in the case where more than one subject is located in one of the remote rooms we talk of a *mixed performance*. The *medium* is distinguished as being either *physical*, i.e. air propagation, or *networked*, i.e. internet connection. Given a musician, we

<sup>1</sup> <http://intermusicproject.eu>.

define the environment where she/he is playing as *real environment*, while the representation of the remote one, where the other musician/s are playing as the *remote environment*. We define a series of presence-based [23] constructs characterizing the NMP experience, general enough to be able to be applied in scenarios different from the chamber music ones. For a more thorough explanation, we refer the interested reader to [21]. In order to analyze/evaluate the performance it necessary to conduct a data collection step, which then allows to evaluate in terms of both objective and subjective measures. The former serve to the purpose of numerically evaluating the performance [31], while the latter, mainly consist of questionnaires investigating the proposed presence constructs and focus on the general QoE of the NMP.

We now present two studies, both performed at the Conservatorio di Musica “Giuseppe Verdi” di Milano using conservatory students aged from 14 to 29 years. The purpose of the studies it to both verify the hypotheses behind the creation of the TEMPERANCE framework and to understand which are the main aspects that need to be treated in order to create a satisfying NMP through IMPERMANENCE. More specifically, the two studies aim at separately exploring the impact of latency and of the audiovisual setup on the performance. In both cases the musicians were placed into two separated rooms connected via LOLA [24]. A server was placed between the two rooms in order to act as a network emulator simulating real-world-like conditions.

## 2.1 Study I: Latency Perception in NMP

The first study [18, 20] concerns the impact that the network latency has on the performance of the musicians. Ten volunteers participated, divided into five duets. Each of them performed under 6 different latency conditions, ranging from 28 ms to 134 ms (2-way). The order under which the conditions were performed was random for each couple. The stimuli (i.e. the music) were based on Béla Bartók *Mikrokosmos* [1] piano pieces, based on exercises exploring *rhythm-melody-expression* relationships [21]. The performance of the musicians were analyzed both in terms of objective measurements such as tempo trend, i.e. how the musicians vary the tempo during the performance, and asymmetry, i.e. the disalignment between the performance of the musicians, and in terms of subjective measurements, based on a questionnaire exploring the presence-based constructs. The results show that no single general trend is understandable from the way that musicians cope with different levels of latency, in the sense that each musician behaves differently. It is interesting to analyze how, even if confronted with high delay times, musicians are able to somehow adapt themselves or at least to adopt different type of strategies. These findings motivated us to avoid a latency-minimization approach, already present in the aforementioned NMP software solutions, and instead to focus on providing tools that are able to help the musicians in coping with the latency, that is, the adaptive metronome approach presented in the IMPERMANENCE framework.

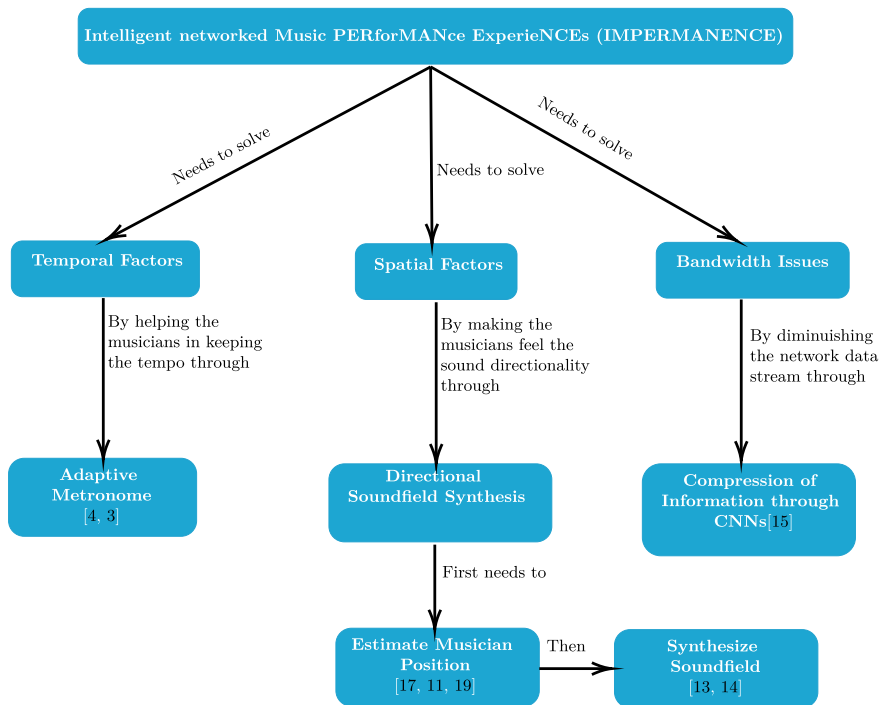
## 2.2 Study II: Audiovisual Immersion in NMP

The second study concerns the audio-visual immersion in NMP environments. We considered 8 musicians, corresponding to 4 duets, playing Béla Bartók's "44 Duos for 2 Violins" [2], that are pedagogical pieces addressed to train motor responses to aural problems and rhythmic and structural features. During this study, the duets had to play using two different setups corresponding to different degrees of immersion. The first setup consisted of a 24-inch screen, positioned in front of the performer at a distance of approximately 1.76 m from the back of the chair, while sound was rendered through a pair of loudspeakers positioned at the sides of the monitor. The second setup was created in order to be more resemblant of a face-to-face situation. A 50-inch screen was positioned at the side in order to project the remote performer as if it was laterally positioned with respect to the local musicians. Sound was rendered through a pair of open headphones (Sennheiser HD-650), augmented with a custom-made head-tracker [22] whose data were fed to a Pure Data (PD) patch that provided binaural audio in real-time through a series of Head Related Transfer Functions (HRTFs). The impact of the two setups on the musicians' performances were analyzed through a questionnaire [22]. The obtained results provided important design implications. The simple frontal screen setup was perceived as plausible by the musicians, enabling us to avoid to specifically treat the visual perception in the IMPERMANENCE framework. On the other hand, the auditory perception was felt as more puzzling. While the 3D audio was perceived as useful by the musicians, the presence of the headphones was sometimes felt as obtrusive, motivating us to pursue research in 3D audio rendering based on loudspeaker setups.

## 3 The Intelligent Networked Music PERforMANce experiENCes (IMPERMANENCE) Framework

By exploiting the information extracted from the two experiments presented in Sects. 2.1 and 2.2 through the TEMPERANCE framework, we are able to propose a unified framework for NMPs whose characteristics are focused on satisfying the needs of the musicians, by tackling both temporal and spatial factors. In Fig. 2 we present the conceptual map summarizing IMPERMANENCE.

In order to mitigate the impact that latency has on the musicians, we propose to use a technique based on the adoption of an adaptive metronome [3, 4]. A metronome is a device that produces a tick at regular intervals, commonly used by musicians in order to keep the correct tempo while practicing. The metronome becomes adaptive, by combining it with a beat tracker, that is a device able to retrieve the tempo of a generic musical audio signal, in order to modify its tempo depending on the musicians' performance. Differently from most NMP frameworks proposed in the literature, we do not try to minimize the latency, but instead help the musicians in coping with it. This is due to the findings presented in Sect. 2.1 that demonstrate that the musicians



**Fig. 2** The IMPERMANENCE framework for the implementation of Networked Music Performances

are able to devise strategies to contrast the effect of the latency. Treating spatial factors means to tackle both the auditory and visual perception. Following the findings obtained through Study II, presented in Sect. 2.2, where it was shown that a screen could suffice for what concerns the visual perception, we decided to focus on creating an interesting environment from the auditory point of view. For this reason, we present a technique for the rendering of the soundfield based on irregular loudspeaker setups. Since the IMPERMANENCE framework would imply that additional information needs to be transmitted via network (e.g. multichannel signals, metronome signal), we also preliminary explore the possibility of compressing the audio via Convolutional Neural Networks (CNNs).

## 4 Latency Compensation Through Adaptive Metronomes

The latency present in all network-based communications is of one of the main challenges that musicians face when performing a NMP. As already pointed out, several low-latency solution exists, however their major drawback is that they often need

high-end hardware, often not available to the musicians. In IMPERMANENCE, we propose to treat the latency, not by trying to minimize it, but instead, by trying to help the musicians to cope with it. A viable solution consists in using metronomes, light-weight devices, commonly used by musicians. The use of metronomes in NMP contexts was already explored in the literature [28]. In the IMPERMANENCE framework, we want to modify the metronome paradigm in order to make it akin to a Virtual Conductor (VC) [8], that is a software that gives tempo indications to the musicians, much like a conductor in a real orchestra does. In order to do this, we propose to use *adaptive metronomes*, that is, metronomes that are able to track the tempo of the musicians during the performance, through an off-the-shelf beat tracking algorithm [26], and to modify their tempo accordingly. We propose [3, 4] three adaptive metronomes techniques, differing by the complexity with which the tempo information is processed, namely: Single Beat Tracking with Master/Slave Approach (SBT), Crossed Beat Tracking (CBT) and Unique Metronome with Virtual Conductor (UMVC).

The SBT technique was proposed in order to get a first exploration of the viability of the adaptive metronome solution. It works by making the assumption that musicians behave following the *master-slave* latency compensation technique [10]. The musicians take two distinct roles: the *leader* determines the tempo of the performance and the *follower* tries to be synchronized with the leader's tempo. The SBT technique uses a single adaptive metronome that tracks the tempo of the leader and provides it to the follower, hopefully helping him/her in keeping a more steady tempo. SBT was tested in the Conservatorio Giuseppe Verdi di Milano [4] with real world musicians, showing promising results.

We then explored techniques that track the tempo of the two musicians at the same time. In the CBT technique, both musicians are tracked by two adaptive metronomes and each of them listens to a metronome whose tempo is based on the performance of the other musician. The UMVC technique instead uses two beat trackers to track both musicians, but provides only one common reference metronome signal to both of them, acting more similarly to a conductor. The details of the proposed algorithms are contained in [3], CBT and UMVC were tested only with amateur musicians and while results are still preliminary they encourage us to further developments of the techniques.

## 5 Spatial Audio Reproduction

The spatial perception of the sound emitted by the musical instruments, plays an important role in defining the quality of the NMP experience. However, as found out through study II, presented in Sect. 2.2, the physical devices used to perform reproduction should not be experienced as obtrusive by the musicians. For these reasons, in IMPERMANENCE we propose to reproduce the audio through an irregular array loudspeaker setup, that is an array where the spacing between the loudspeakers is not constant. This is motivated by the fact that while classic soundfield synthesis techniques are based on extensive loudspeaker arrays, NMP rooms are often cramped

with instruments and are hardly able to accommodate such setups. An irregular loudspeaker setup is more easily deployable in such scenario. However, irregular setups are challenging from the signal processing point of view, since the non-uniform sampling of the soundfield causes distortions during reproduction. A function modeling the correct driving signals for such loudspeaker configuration would be highly non linear and complex to model analytically. To overcome these limitations we propose to use techniques derived from deep learning, which enable the automatic extraction of highly non linear functions. In order to be able to reproduce the desired soundfield, we must know beforehand the position of the musicians in the room. This allows us both to reproduce correctly the perceived location (i.e. directionality) or to move them inside the remote room. Similarly to soundfield synthesis techniques, source localization ones are often based on extensive microphone array setups; we propose two techniques for the localization of the musicians based on minimal setups.

### ***5.1 Source Localization Using Distributed Microphones Based on Ray Space Transform and Deep Learning***

The first proposed localization technique is able to perform 2D localization of the musician, through the deployment of a small number of, arbitrarily deployed, microphones. If the microphones are synchronized, it is possible to compute the Generalized Cross Correlation (GCC) [29] between a reference microphone and the other ones. The highest peak of the GCC corresponds to the time-delay of the sound emitted by the source and measured at the microphones, which can then be exploited to perform localization. Unfortunately, noise and reverberation create several spurious peaks in the GCC, making the localization task harder. In the technique proposed in [17], we take advantage of CNNs and of the Ray Space Transform (RST) [6], a compact representation of the soundfield acquired by multiple Uniform Linear Arrays (ULAs) of microphones in terms of acoustic rays. Through a CNN we map the noisy GCC obtained at the real sparsely deployed microphones and map it to the simulated RST computed in anechoic conditions using ULAs surrounding the whole room. As demonstrated in [17] this procedure allows an accurate source localization even when dealing with highly challenging environments.

### ***5.2 Source Localization Through Frequency-Sliding Generalized Cross-Correlation***

The second proposed localization technique is based on an extension of the GCC framework, that entails the computation of the Frequency-sliding Generalized Cross-correlation (FS-GCC) following a sliding window approach. This enables us to obtain a set of sub-bands GCCs which can be stacked together into a single matrix, where



each row of the matrix corresponds to the GCC relative to the corresponding frequency band. The usefulness of this approach is given by the fact that in the anechoic (i.e. without noise and reverberation) scenario, the FS-GCC matrix is rank one. This allows us to exploit low-rank approximations and to separate the noisy components from the desired ones. Results presented in [11] show that the proposed approach is able to obtain better results than the GCC over real measurements. The proposed technique does not require any deep learning technique, and so it is more suitable to scenarios where extensive computational power may not be available. However, we experimented with the possibility of applying deep learning to the FS-GCC method in [19] by using a CNN to denoise the noisy input FS-GCC, showing better performances.

### ***5.3 Soundfield Synthesis Through Irregular Loudspeaker Arrays***

In order to reproduce the soundfield, we implement a technique that leverages on deep learning and on the Plane Wave Decomposition of the soundfield. Operatively, we consider the method already proposed in [5], where a Model-based acoustic Rendering (MR) technique is proposed. The MR technique reproduces the soundfield accurately when the considered setup is regular, but suffers from reproduction errors when dealing with irregular setups. The proposed technique, detailed in [13] considers a CNN that takes as input the driving signals obtained through the MR technique and outputs a compensated version of them. Since no ground truth is available for the compensated driving signals, we incorporate into the network architecture a soundfield estimation part, that simply convolves the obtained driving signals with the corresponding point-to-point Green's function, obtaining the estimated reproduced soundfield. The difference between estimated and ground-truth soundfield in terms of modulus and phase is used to compute the loss and train the CNN. We recently proposed an extension of this technique [14] following the same approach, with the difference that the driving signals are not obtained by compensating the ones obtained through a pre-existing method, but by extracting them through a CNN directly from discrete points of the environment where the soundfield is computed.

## **6 Speech Reconstruction from CNN Embeddings**

Transmitting all the information needed in a NMP may create bottlenecks for what concerns the data transmission, especially when dealing with multidimensional signals such as the ones needed in a spatial audio framework. In [15] we explore a technique that could enable the compression of audio data. We do this by first studying a preliminary problem that has also relevant implications from an Explainable Arti-

ficial Intelligence (XAI) perspective. CNN models dealing with audio, usually take as input time-frequency representations of the audio signal, which is consecutively compressed by the layers of the model in order to extract high-level features. In [15] we consider pre-trained CNNs as feature extractors and build specular decoder networks in order to reconstruct the input time-frequency representation from the output of the intermediate layers. Results show that it is easy to reconstruct the input from convolutional layers, while it is considerably more complex when dealing with fully connected ones. This motivates us to further investigations, since by building proper networks, it would be possible to use their intermediate layers in order to perform compression on the multidimensional input audio signals.

## 7 Conclusion

In this chapter we presented the main results concerning the development of IMPERMANENCE as a comprehensive framework for the realization of NMPs. The biggest innovation of IMPERMANENCE is that it does not simply aim at reducing the issues already treated in the literature, e.g. low latency audio/video streaming, but it aims at creating an all around environment that feels coherent and whole to the musicians. In this sense, we treat the NMP as a Functional eXtended Reality (FXR) experience, where the objective is not the simplistic realism, but the creation of an environment that it is based on the needs of the musicians. In order to do this we base the design choices of IMPERMANENCE on the results obtained through experiments performed using the research framework TEMPERANCE, considering real musicians. We believe that the proposed method will enable NMP to be treated not as a mere approximation of the physical performance, but as a type of performance on its own.

## References

1. Bartók, B.: *Mikrokosmos, 153 Progressive Piano Pieces. New Definitive Edition.* Boosey & Hawkes Music Publishers Limited, London (1987)
2. Bartók, B.: *44 duos for 2 violins, vol. 1–2.* Universal Edition (1992)
3. Battello, R., Comanducci, L., Antonacci, F., Cospito, G., Sarti, A.: Experimenting with adaptive metronomes in networked music performances. *J. Audio Eng. Soc.* **69**(10), 737–747 (2021)
4. Battello, R., Comanducci, L., Antonacci, F., Sarti, A., Delle Monache, S., Cospito, G., Pietrocola, E., Berbenni, F.: An adaptive metronome technique for mitigating the impact of latency in networked music performances. In: *2020 27th Conference of Open Innovations Association (FRUCT)*, pp. 10–17. IEEE (2020)
5. Bianchi, L., Antonacci, F., Sarti, A., Tubaro, S.: Model-based acoustic rendering based on plane wave decomposition. *Appl. Acoust.* **104**, 127–134 (2016)
6. Bianchi, L., Antonacci, F., Sarti, A., Tubaro, S.: The ray space transform: a new framework for wave field processing. *IEEE Trans. Signal Process.* **64**(21), 5696–5706 (2016)

7. Bischoff, J., Gold, R., Horton, J.: Music for an interactive network of microcomputers. *Comput. Music J.* 24–29 (1978)
8. Bos, P., Reidsma, D., Ruttkay, Z., Nijholt, A.: Interacting with a virtual conductor. In: *International Conference on Entertainment Computing*, pp. 25–30. Springer (2006)
9. Cáceres, J.P., Chafe, C.: Jacktrip: under the hood of an engine for network audio. *J. New Music Res.* **39**(3), 183–187 (2010)
10. Carôt, A., Werner, C.: Network music performance—problems, approaches and perspectives. In: *Proceedings of the “Music in the Global Village”-Conference, Budapest, Hungary, vol. 162*, pp. 23–10 (2007)
11. Cobos, M., Antonacci, F., Comanducci, L., Sarti, A.: Frequency-sliding generalized cross-correlation: a sub-band time delay estimation approach. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 1270–1281 (2020)
12. Comanducci, L.: *Intelligent networked music PERforMANce Experiences (IMPERMANENCE)*. Ph.D. thesis, Politecnico di Milano (2022)
13. Comanducci, L., Antonacci, F., Sarti, A.: Synthesis of soundfields through irregular loudspeaker arrays based on convolutional neural networks (2022). [arXiv:2205.12872](https://arxiv.org/abs/2205.12872)
14. Comanducci, L., Antonacci, F., Sarti, A.: A deep learning-based pressure matching approach to soundfield synthesis. In: *17th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE (2022)
15. Comanducci, L., Bestagini, P., Tagliasacchi, M., Sarti, A., Tubaro, S.: Reconstructing speech from cnn embeddings. *IEEE Signal Process. Lett.* **28**, 952–956 (2021)
16. Comanducci, L., Borra, F., Bestagini, P., Antonacci, F., Sarti, A., Tubaro, S.: Ray space transform interpolation with convolutional autoencoder. In: *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 261–265. IEEE (2018)
17. Comanducci, L., Borra, F., Bestagini, P., Antonacci, F., Tubaro, S., Sarti, A.: Source localization using distributed microphones in reverberant environments based on deep learning and ray space transform. *IEEE/ACM Trans Audio Speech Lang. Process.* **28**, 2238–2251 (2020)
18. Comanducci, L., Buccoli, M., Zanoni, M., Sarti, A., Delle Monache, S., Cospito, G., Pietrocola, E., Berbenni, F.: Investigating networked music performances in pedagogical scenarios for the intermusic project. In: *Proceedings of the 23rd Conference of Open Innovations Association (FRUCT)*, pp. 119–127. IEEE (2018)
19. Comanducci, L., Cobos, M., Antonacci, F., Sarti, A.: Time difference of arrival estimation from frequency-sliding generalized cross-correlations using convolutional neural networks. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Process (ICASSP)*, pp. 4945–4949. IEEE (2020)
20. Delle Monache, S., Buccoli, M., Comanducci, L., Sarti, A., Cospito, G., Pietrocola, E., Berbenni, F.: Time is not on my side: network latency, presence and performance in remote music interaction. In: *Machine Sounds, Sound machines-Proceedings of the XXII CIM Colloquium on Music Informatics*, pp. 152–158 (2018)
21. Delle Monache, S., Comanducci, L., Buccoli, M., Zanoni, M., Sarti, A., Pietrocola, E., Berbenni, F., Cospito, G.: A presence-and performance-driven framework to investigate interactive networked music learning scenarios. *Wirel. Commun. Mobile Comput.* **2019** (2019)
22. Delle Monache, S., Comanducci, L., Cospito, G., Sarti, A., Avanzini, F.: Synchronous online ensemble performance. In: Johnson, C., King, A. (eds.) *Music, Technology, Innovation: Industry and Educational Perspectives*. Routledge (2022)
23. Draper, J.V., Kaber, D.B., Usher, J.M.: Telepresence. *Hum. Factors* **40**(3), 354–375 (1998)
24. Drioli, C., Allocchio, C., Buso, N.: Networked performances and natural interaction via lola: Low latency high quality a/v streaming system. In: *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250. Springer (2013)
25. Gabrielli, L., Squartini, S.: *Wireless Networked Music Performance*. Springer (2016)
26. Giorgi, B.D., Zanoni, M., Böck, S., Sarti, A.: Multipath beat tracking. *J. Audio Eng. Soc.* **64**(7/8), 493–502 (2016)

27. Holub, P., Matela, J., Pulec, M., Šrom, M.: Ultragrid: Low-latency high-quality video transmissions on commodity hardware. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 1457–1460. ACM (2012)
28. Hupke, R., Beyer, L., Nophut, M., Preihs, S., Peissig, J.: Effect of a global metronome on ensemble accuracy in networked music performance. In: Audio Engineering Society Convention 147. Audio Engineering Society (2019)
29. Knapp, C., Carter, G.: The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process. **24**(4), 320–327 (1976). <https://doi.org/10.1109/TASSP.1976.1162830>
30. Pezzoli, M., Comanducci, L., Waltz, J., Agnello, A., Bondi, L., Canclini, A., Sarti, A.: A dante powered modular microphone array system. In: Audio Engineering Society Convention 145. Audio Engineering Society (2018)
31. Rottondi, C., Chafe, C., Allocchio, C., Sarti, A.: An overview on networked music performance technologies. IEEE Access **4**, 8823–8843 (2016)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

