URSI - RCRS 2022, IIT (Indore), India, 1 - 4 December, 2022

MuSense: Sensor Data Fusion-inspired Intelligent Music Improvisation Framework in 5G-Internet of Music Things

Samarjit Roy* (1), Anwesha Mukherjee(2), Debashis De(3)
(1) Department of Computer Science and Engineering, D Y Patil International University, Pune-411044, Maharashtra, India
(2) Department of Computer Science, Mahishadal Raj College, Purba Medinipur, West Bengal, 721628, India
(3) Department of Computer Science and Engineering, MAKAUT, West Bengal, (formerly, WBUT), Kolkata-700064, India

Abstract

This article focuses on the emerging music composition, reconstruction, and improvisation in the Internet of Things (IoT). In the current internet era, musical composition and reconstruction schemas are largely dependent on the IoT-based paradigm where human beings can arrange musical compositions through musical things, such as, remotely arranged singing and musical instruments, smart rhythmic behaviors, and intelligent autotuning systems. We have renamed the music composition, sharing, and recommendation phenomenon in the IoT paradigm as the Internet of Music Things (IoMT). We present a fifth-generation (5G) technology-enabled IoMT framework to explore remote music collaboration, multisensory music information fusion, and socially-desirable music composition schema due to its capability of providing ultra-low latency, prominent bandwidth, assured Quality of Services, and lively orchestrated network environment. We incorporate the system performance metrics with the proposed 5G-IoMT in terms of the data transmission latency and potential energy dissipation. We moreover illustrate a Generative Adversarial Network-inspired intelligent music improvisation strategy as a case study for audience interest.

1. Introduction

The IoMT is an emerging research paradigm illustrated as an extension of conventional IoT application scenarios that explore new collective interfaces for the dynamic musical expression, pervasive music analytics, human-machine interaction, intelligent computing, and participatory virtuosity [1]. In the context of computational sciences, IoMT signifies the distributed networks of embedded computational devices, such as physical musical objects offered to quality music composition, significant and suggestive improvisation on humanistic and audience care. IoMT facilitates the association of heterogeneous physical and digital contents through a dedicated network in IoT for interpersonal communications [2]. Enabling remotely scattered musical performances into a single interface is challenging through conventional network technologies due to the high latency and low bandwidth [1], [2]. Evolving 5G technology has the capabilities to afford high bandwidth, ultra-low latency, promised service quality, and responsive network adaptation [3], [4]. The proposed 5G-enabled IoMT schema can enhance geo-spatially distributed and real-time remote music performance collaboration, cost-effective music composition, remote accessibility of rare musical instruments, knowledge-based and remote talent participation in qualitative music-making. Although in the proposed 5G-IoMT framework, systematized personalization and professionalization in music composition are challenging concerns. The musicians and composers compose and/or generate music for the audiences and recommend them to listen for enriching mental health or to lead them for music learning. Intense integration of intelligent systems [5], such as artificial intelligence and machine learning, with music composition and reconstruction strategies, significantly heightens the personalized music composition, recommendation [6], and humanistic care-inspired music computing. Generative Adversarial Network (GAN) is a conceptual sub-domain of an intelligent system that can generate new music instances and help to the music personalization in terms of reconstruction and improvisation [7], [8]. We incorporate a case study in this article on GAN-based music reconstruction schema in the 5G-IoMT paradigm. The proposed framework is named as MuSense. Table 1 represents the summary of the state-of-the-art for illustrating our proposed MuSense paradigm based on architecture, system analytics, intelligence, efficiency, and music improvisation.

Table 1. State-of-the-art summary for MuSense illustration.

	IoT architecture	System analytics	Intelligent system	Efficiency metrics	Music improvisation
IoMT [1-2]	✓	✓	×	✓	✓
5G-IoT [3-4]	✓	×	×	✓	×
Intelligent IoT [5]	✓	✓	✓	✓	×
Music analysis [6]	✓	×	✓	×	✓
GAN-Music [7-8]	×	×	✓	×	✓
GAN in IoT [9]	✓	×	✓	✓	×
MuSense (Proposed)	✓	✓	✓	✓	✓

1.1 Contributions

The key contributions of the MuSense framework are: (a) We illustrate 5G-enabled IoMT architecture; (b) Incorporation of GAN-inspired intelligent computing strategy with IoMT for personalized music reconstruction and quality improvisation; and (c) Depiction of an efficient 5G-enabled IoMT framework.

1.2 Paper Organization

We organize the remaining article as follows: Section 2 presents the projected sensor data fusion-based IoMT system architecture. Section 3 illustrates the system performance metrics in terms of system latency and energy consumptions. Section 4 demonstrates a systematic case study on the GAN-based intelligent and 5G-enabled IoMT schema. We accomplish our paper with the conclusion and the several scopes for future researches in the Section 5.

2. Proposed System Architecture

Figure 1 depicts the five-layered proposed IoMT system architecture. The bottom-most layer (L1) represents the sensor devices that are capable of sensing the remote real-time music information and transmits pre-processed and unprocessed information to the next layer (L2) through the network [1]. This layer contains the user devices. We perform the pre-processing and local analytics in L2. After preprocessing, L2 transmits the scattered remote information to the immediate next layer (L3) that consists of several edge servers/ cloudlets for efficient information processing.

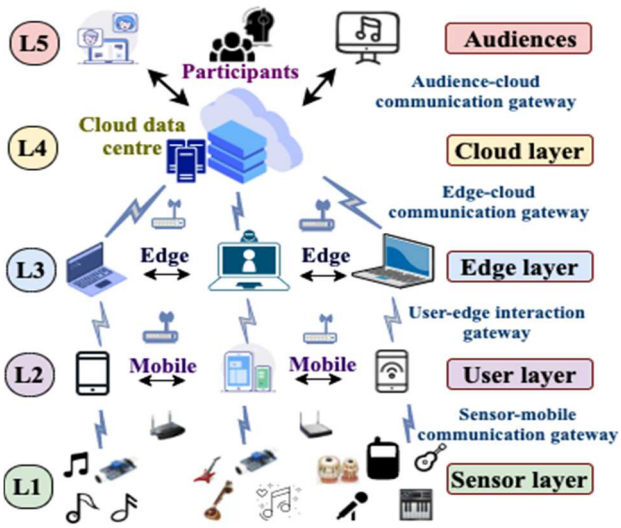

Figure 1. Representation of IoMT architecture.

Based on the music composers' requirements, the edge servers perform distributed tasks, such as composition, generation, and reconstruction. The edge servers in L3 are also capable to provide the infrastructure for intelligent analytics, viz., representation learning, AI applications, and federated intelligence over the musical tasks for minimizing the data traffic and system latency in the immediate next cloud layer (L4). The Cloud layer consists of the data center that stores the aggregated and composed musical performances from the L3. The L4 primarily provides data center-specific digital music to the audiences and task-specific virtual services to the end-users. Based on the recommendation mechanism, the audience can listen to the remotely-arranged music, participate in the personalized music generation, and explore the information and communication technologies (ICT)-inspired music teaching-learning pedagogies.

3. System Performance Metrics

Let the data collection latency at L1 is L_{sen}. Energy consumption (E_{sen}) of user device during this period is calculated as,

$$E_{sen} = E_i \times L_{sen} \tag{1}$$

where, E_i is power consumption of user device in idle mode. Data transmission latency (L_{su}) from L1 to L2 can be calculated as,

$$L_{su} = (1 + f_{su}) \frac{Data_{su}}{R_{su}} \tag{2}$$

where, f_{su}, $Data_{su}$, and R_{su} are link failure rate, transmitted data amount, and data transmission rate from sensor to user respectively. The energy consumption in user device (E_{su}) is evaluated as,

$$E_{su} = E_r \times L_{su} \tag{3}$$

where, E_r is energy consumption of user device in reception mode. Data transmission latency from L2 to L3 (L_{ue}) is,

$$L_{ue} = (1 + f_{ue}) \frac{Data_{ue}}{R_{ue}} \tag{4}$$

where, f_{ue}, $Data_{ue}$, and R_{ue} are link failure rate, transmitted data amount, and data transmission rate. Energy dissipation of the user device (E_{ue}) during this period is,

$$E_{ue} = E_t \times L_{ue} \tag{5}$$

where, E_t denotes energy consumption of user device in transmission mode. Data transmission latency from edge to cloud (L_{ec}) is,

$$L_{ec} = (1 + f_{ec}) \frac{Data_{ec}}{R_{ec}} \tag{6}$$

where, f_{ec}, $Data_{ec}$, and R_{ec} are the link failure rate, transmitted data amount, and the rate of data transmission from L3 to L4 respectively. Energy consumption of user device (E_{ec}) during this period is,

$$E_{ec} = E_i \times L_{ec} \tag{7}$$

where, E_i symbolizes as the consumption of energy of the sensor device in idle state. Information processing latency in L4 (L_p) can be denoted as,

$$L_p = \frac{Data_p}{S_p} \tag{8}$$

where, $Data_p$ and S_p are processed data amount and data processing speed respectively. Energy consumption of user device (E_p) can be evaluated as,

$$E_p = E_i \times L_p \tag{9}$$

Similarly, in the recursive back-tracking mechanism and interpersonal communications within the system, we evaluated the (a) transmission latency from cloud to edge (L_{ce}) and corresponding energy dissipation (E_{ce}), and (b) transmission latency from edge to user device (L_{eu}) and corresponding energy dissipation (E_{eu}).

Therefore, total latency in receiving the outcome by enabled user device is denoted by L_{total}, evaluated as,

$$L_{total} = (L_{sen} + L_{su} + L_{ue} + L_{ec} + L_{ce} + L_{eu} + L_p) \quad (10)$$

Similarly, total energy dissipation within the system (E_{total}) during the stipulated period is,

$$E_{total} = (E_{sen} + E_{su} + E_{ue} + E_{ec} + E_{ce} + E_{eu} + E_p) \quad (11)$$

When an audience raises queries or request for a composed item from the system, data transmission latency (L_{recom}) can be denoted as,

$$L_{recom} = \left[(1 + f_{ue})\frac{Data_q}{R_{ue}}\right] + \left[(1 + f_{eu})\frac{Data_r}{R_{eu}}\right] \quad (12)$$

where, $Data_q$ and $Data_r$ denote data transmission amount for the corresponding request and system response. The energy dissipation during this scenario ($E_{audience}$) is,

$$E_{audience} = E_t \times \left((1 + f_{ue})\frac{Data_q}{R_{ue}}\right) + E_r \times \left((1 + f_{eu})\frac{Data_r}{R_{eu}}\right) \quad (13)$$

Figure 2 presents a comparison between the total latency in conventional cloud-based music analytics framework [1-2] and the projected 5G-enabled edge-based IoMT framework. We observe that the proposed framework reduces the latency by ~17% than the conventional cloud-based framework. Figure 3 demonstrates the energy consumption of the user device. We observe that for 5000-20000k data transmission between two consecutive nodes, the latency, and energy consumption of the user device are 5-30 seconds and ≤300 Joule respectively while using the proposed framework.

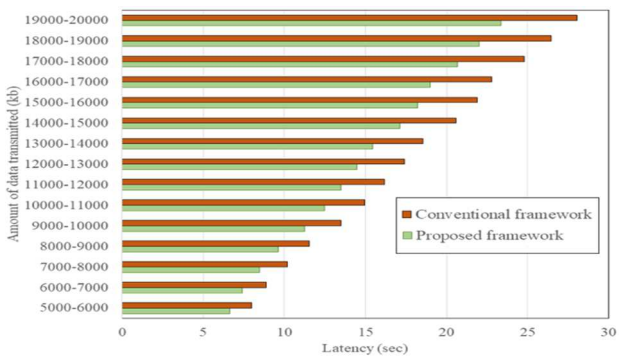

Figure 2. Total latency in proposed and conventional frameworks.

4. GAN in 5G-IoMT: Systematic Case Study

In this section, we demonstrate a systematized case study on GAN-based music reconstruction and improvisation. GAN explores a cloned network-based algorithmic strategy to breed new data-instance typically process as an alternative to real data [7-8]. In our projected scenario, we have illustrated this strategy for (a) musical performance cloning for future composition utilities; (b) high-efficiency evaluations in the performance contexts; (c) missing portion estimation and recovery on interest within a remote music composition, and; (d) exemplification of personalized music improvisation based on music listeners' demands. Table 2 represents the GAN-based system infrastructure that we incorporated for system training. We provide system designing use-case and incorporated dataset in [10]. The known dataset operates as the primary dataset for discriminator for achieving optimum outcome through the sample training. We have seeded the generator with random input but sampled it in the predefined normal distribution. We have applied a conventional back-propagation algorithm [8] for both known and generated networks so that generator generates more accurate samples and the discriminator becomes more efficient. Figure 4 and Figure 5 present the discriminator loss (D-loss) and the generator loss (G-loss) respectively.

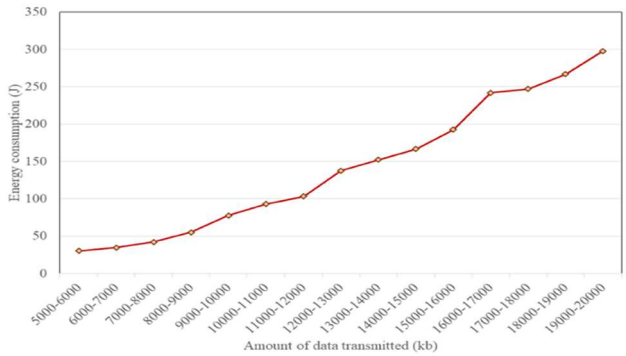

Figure 3. Energy consumption of user device in proposed framework.

Table 2. System specifications for proposed MuSense framework.

System features	Values/specifications
Known sample	27649 nos.
Generated random input	1024 nos.
Generator feature	27649 nos.
Known and generated network	Sequential dense network
Activation function	ReLU
Optimization mechanism	Adam
System learning rate	0.0001
Classification strategy	Binary cross-entropy

Based on the projected motivation, our objective is to observe whether the D-loss may be high and G-loss is gradually decreasing with time and processing of data samples. In Figure 4 and Figure 5, we observe optimum outcome over the time on iterated system training performance through the proposed MuSense schema.

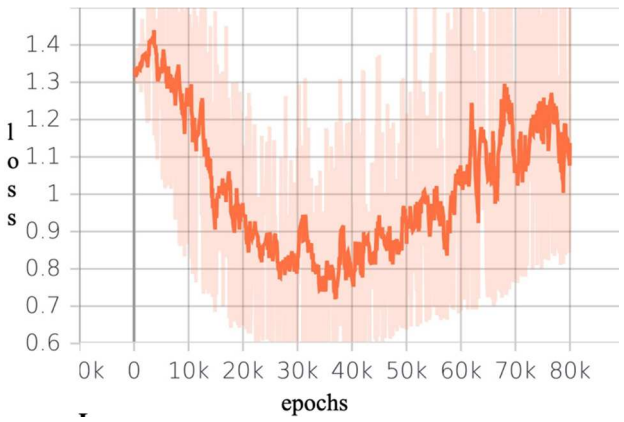

Figure 4. Representation of Discriminator loss (D-loss) on system training over the dataset [10].

We have performed this experiment in the Centre for Mobile Cloud Computing (CMCC) lab of our university in the projected 5G-enabled IoMT schema and illustrated the proposed GAN-based intelligent music construction framework in the distributed edge servers. We have attained 10k cloned data after performing the training over 27649 input data. Our proposed system achieves the Precision ~0.8, Recall ~0.94, F1-score ~0.86, and Accuracy ~0.78 after system training. As far we are aware, this is the first GAN-inspired intelligent music composition, reconstruction, and quality improvisation mechanism in the 5G-enabled IoMT framework.

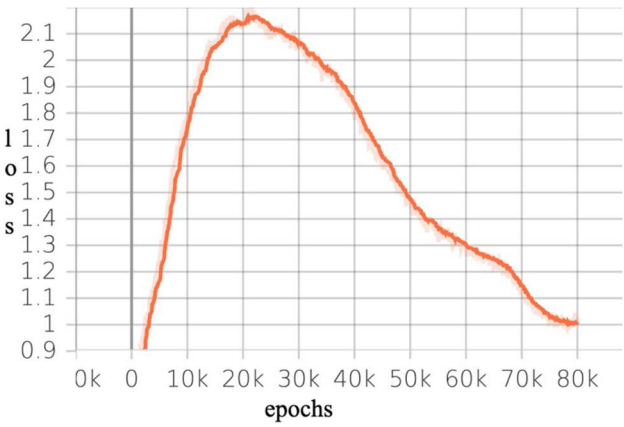

Figure 5. Representation of Generator loss (G-loss) on system training over the dataset [10].

5. Conclusion

We illustrated GAN-inspired intelligent and personalized music improvisation schema in the 5G-enabled Internet of Music Things paradigm. It is a thought-provoking exercise to reconstruct, generate, and recommend remotely-arranged collaborative music performances on audience concern where human sentiments and emotions are correlated. Our proposed MuSense framework (a) depicts the music reconstruction architecture established on sensor information fusion-inspired 5G-IoMT and GAN-based personalized music arrangement schema, (b) evaluate system performance metrics in terms of data transmission latency, and system energy dissipation, and (c) demonstrates a GAN-dependent music reconstruction case-study in the emerging IoMT paradigm. We achieved that the data transmission latency in our proposed MuSense framework is ~17% less than the conventional computational paradigm along with satisfactory energy-efficiency evaluation. Furthermore, the projected GAN-based case study also demonstrates a significant mechanism in the context of music reconstruction with a ~78% accuracy. In our future endeavors, we may undertake to incorporate the (a) channel capacity, bandwidth, jitter, data traffic, and so forth; and (b) multifaceted multiple intelligent computing strategies to achieve more efficient, optimally interoperable multimedia data fusion-induced 5G-enabled Intelligent IoMT system.

6. Acknowledgements

The authors are grateful to the University Grants Commission (UGC), Govt. of India, for sanctioning research fellowship under which this article has been completed.

7. References

1. S. Roy, D. Sarkar D, S. Hati, D. De, "Internet of Music Things: an edge computing paradigm for opportunistic crowdsensing," *The Journal of Supercomputing*, **74**, 11, 2018, pp. 6069–6101, doi: 10.1007/s11227-018-2511-6.
2. S. Roy, D. Sarkar, D. De, "DewMusic: crowdsourcing-based internet of music things in dew computing paradigm," *Journal of Ambient Intelligence and Humanized Computing*, **12**, 2, 2021, pp. 2103–2119, doi: 10.1007/s12652-020-02309-z.
3. S. Li, L. D. Xu, S. Zhao, "5G Internet of Things: A survey," *Journal of Industrial Information Integration*, **10**, 2018, pp. 1–9, doi: 10.1016/j.jii.2018.01.005.
4. L. Chettri, R. Bera, "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems," *IEEE Internet of Things Journal*, **7**, 1, 2020, pp. 16–32, doi: 10.1109/jiot.2019.2948888
5. D. Wang, D. Chen, B. Song, N. Guizani, X. Yu, X. Du, "From IoT to 5G I-IoT: The next generation IoT-based intelligent algorithms and 5G technologies," *IEEE Communications Magazine*, **56**, 10, 2018, pp. 114–120, doi: 10.1109/MCOM.2018.1701310
6. S. Roy, M. Biswas, D. De, "iMusic: a session-sensitive clustered classical music recommender system using contextual representation learning," *Multimedia Tools and Applications*, **79**, 33-34, 2020, pp. 24119–24155, doi: 10.1007/s11042-020-09126-8
7. H. W. Dong, W. Y. Hsiao, L. C. Yang, Y. H. Yang, "Musegan: Multitrack sequential generative adversarial networks for symbolic music generation and accompaniment," *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 2018, pp. 34-41.
8. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE Signal Processing Magazine*, **35**, 1, 2018 pp. 53–65, 2018. doi: 10.1109/msp.2017.2765202
9. L. Nie, Y. Wu, X. Wang, L. Guo, G. Wang, X. Gao, S. Li, "Intrusion Detection for Secure Social Internet of Things Based on Collaborative Edge Computing: A Generative Adversarial Network-Based Approach," *IEEE Transactions on Computational Social Systems*, **9**, 1, 2021. doi: 10.1109/TCSS.2021.3063538
10. S. Roy, "Music Reconstruction using Generative Adversarial Networks," Mendeley data, V2, 2021, Available: 10.17632/t2wp7myg9v.2