CONDITIONAL SEMANTIC MUSIC GENERATION IN A CONTEXT OF VR PROJECT "GRAPHS IN HARMONY LEARNING"

Anna Shvets
Fablab by Inetum
Paris, France
anna.shvets@inetum.com

Samer Darkazanli
iMSA
Montauban, France
darkazanli.samer@imsa.msa.fr

ABSTRACT

The article proposes a perspective on the use of generative artificial models in a context of the VR project "Graphs in harmony learning". The usage of LSTM, ConvLSTM and conditional GAN with convolutional 1D layers for semantic music generation is discussed. The efficiency of the novel data encoding scheme, along with the design patterns based on the system of graphs, are shown.

1. INTRODUCTION

1.1 Context

The article presents a new system of representation, based on a graph theory [1]. The representation methodology has been proven efficient in a multi-step pedagogical experiment in a context of hybrid learning. The experiment demonstrated a substantial increase of the quality of knowledge in a group of students, benefiting from the system application in a learning process [2, 3]. The method has also been applied in building graphic interface of the award-winning mobile and VR applications [4, 5].

1.2 Motivation

The global pandemic crisis had shown the need for research of convenient forms of distance learning to compensate the reduced interactions, detachment and isolation of individuals, which brings harmful consequences. As student surveys indicates, more than one in two students had thought of dropping the classes during the pandemic and 71% of surveyed confessed being worried about their mental health[1]. The same resource points out the lack of immersion for the videoconference format of courses, resulting in great difficulties in following the course content, which led 70% of the surveyed students to express their pessimism about their academic success. Another study[2], listing the reasons of dropping the university, mentions a lack of practice, especially for individuals who are not equipped for theoretical courses of the university's curriculum.

With the application of teaching method based on the effective representation methodology and via the interaction with the tangible elements of virtual reality, it is expected that the degree of immersion will closely approach the level of practical face-to-face lessons.

2. STATE OF THE ART

2.1 AI in Education

A meta-analysis studying the application of AI in teaching, which included 146 recent articles in the field [6], proposes a categorization according to the target group, such as learner-oriented AI, teacher-oriented AI and system-oriented AI. Following this taxonomy, the learner-oriented AI tools comprise software solutions to learn a subject, such as adaptive or personalized learning systems and intelligent tutoring systems; teacher-oriented AI tools are used to help the teacher in reducing a workload by automating tasks such as administration, assessment, feedback and plagiarism detection; system-oriented AI tools provide information to administrators and managers at the institutional level to monitor the acceptance rates in faculties or the employability trends.

The development of intelligent tutoring systems is mainly focused on visual recognition, aiming to provide a system feedback coherent with the implicit reactions of the learner [7, 8]. Another trend comprises creation of intelligent campus ecosystems, with a strong reliance on chatbot solutions [9]. The above mentioned meta-analysis reveals a lack of educational theory to support a technological choice. Current research in the field is oriented towards analysis of data patterns to build AI models or to support administrative decisions using known statistical and machine learning methods. The authors point out that there is very little evidence for the advancement of pedagogical theories related to AI-based educational technology.

The VR project specified in a present article makes a big difference compared to the state of the art by proposing the application of an innovative pedagogical strategy. This strategy is based on a new knowledge representation methodology, it exploits the possibilities of AI for the forms of

Copyright: © Anna Shvets, Samer Darkazanli. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[1] https://etudiant.lefigaro.fr/article/covid-19-depuis-le-debut-de-l-epidemie-plus-d-un-jeune-sur-deux-a-envisage-d-arreter-ses-etudes_65135afa-622e-11eb-8fde-d92bf2ba0bfe/

[2] https://fr.cursus.edu/10328/les-enjeux-du-decrochage-universitaire

activity inherent to the artistic practice (analysis of the structural properties of chords in harmonic sequences), and therefore respond to the needs of educational process.

2.2 XR in Music

Another recent meta-analysis [10] brings together research on virtual reality (VR), augmented reality (AR), augmented virtuality (AV), mixed reality (MR) and extended reality (XR), applied to music. The meta-analysis covers 260 publications appeared from 1990 to 2020 which present musical research in XR, encompassing technical, artistic, perceptual and methodological fields and reveals an exponential growth of publications starting from 2015, which is explained by the increased accessibility of XR hardware and software tools.

The study proposes a definition of the term *musical XR system* by classifying the existing systems according to the ways the sound and music are used (diegetic mode of usage versus non-diegetic use), the cinematics of sound events (fixed audio playback versus sound events triggered in an interactive manner or procedurally generated), and sound spatialization (fixed stereo sound presentations versus dynamic spatial audio implementations based on user location tracking). Finally, the grouping of articles is made according to the main objective of the musical XR system (performance, education, composition, sound engineering, entertainment, perception study, development), end-user (performer, student, composer, member of the audience, studio engineer, developer), social experience (individual or multi-user experience) among others.

This grouping allows the assumption that the existing educational applications of virtual reality mainly focus on improving the practice of performance, using interfaces of traditional instruments modelled in 3D (piano, guitar, drums), neither of applications presented a system of abstraction of musical knowledge. Most of applications focus on the beginner level and a very few are designed for the expert or intermediate level musicians; almost all systems were made for students and dedicated to self-training.

In view of the state of the art in the field, the VR project in a present article is not only in accordance with the best practices in the construction of a musical XR system (spatialization of the sound source, triggering of the sound events in an interactive way, the unfolding of musical content in a diegetic way), it also fills the gap for a graphical interface independent from attachment to a specific musical instrument. It also proposes the systematization of musical knowledge in a non-verbal representation, compatible with immersive worlds.

2.3. Conditional music generation

The approaches to music generation process falls into three groups: conditional, controllable and constraint generation. Conditional generation takes one element as input to generate another element as target [11, 12], while controllable generation uses the change in input features to manipulate different aspects of the output generation [13, 14]. The third group, containing constraint generation, makes use of the template-based approach to influence a shape of the output result [15]. The research in controllability mainly explores features disentanglement, proposing systematic studies [16] and datasets [17], designed to foster further experiments in the field. The existing resources, however, are mostly gathering monophonic music examples and therefore are not suitable for harmonic sequences generation.

The research in conditional music generation presents a spectrum of generative architectures, such as LSTM, Transformer [18], GAN [19, 20], hybrid versions, such as LSTM-GAN [13] or GAN with an inception model [21]. The latter architecture exploits convolutional layers, followed by the time distribution layer that captures sequential data, which enforces the convolutional layers considering the time relationship in a similar manner as RNN layers. As a comparison to this type of architecture, a hybrid ConvLSTM architecture processes the sequential data, where each element of the sequence passes through a convolutional layer followed by LSTM layer [22]. The previous experiment has shown that this type of architecture captures well two-dimensional sequential data [23] and will be exploited in the experiment on novel encoding scheme application, presented further.

3. METHODOLOGY

The methodology consists of the system of graphs and an augmented score representation of harmonic sequences. The system of graphs is organized in horizontal and vertical triads, reflecting logical relations between chords, whereas the augmented score contains a meaningful color scheme for a visual distinction of the chord functions, along with color shades, revealing the chord structure.

3.1 The system of graphs

The system of graphs embraces functional correlations between chords, consisting of the Roman numerals, representing degrees of the scale, placed in a specific order [24]. The placement order is defined by variation potential within a multitude of diatonic progressions. The slots containing the information about chords are mapped to a specific type of information. In consequence, the slots reserved for the seventh chords cannot contain the information about triads (see in Figure 1). This limit comes from the theory of graphs [1], defining a graph as a graphical representation of frames, corresponding to a specific knowledge representation.

Figure 1. Unfilled graph structure (on the left) along with the graph filled with the information about three passing progressions, between the seventh chord of the tonic and its first inversion (on the right). The tonic seventh chords

are inside the tops of the graph and three triads are inside the edges of the graph.

The graphs are organized in horizontal triads (see Figure 2) and vertical triads (see Figure 3).

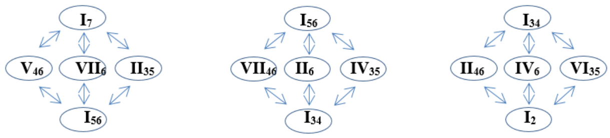

Figure 2. Horizontal triad regroups the graphs containing tops sharing the same degree. In this example, a horizontal triad embraces all possible passing progressions between the tonic seventh chord and its inversions.

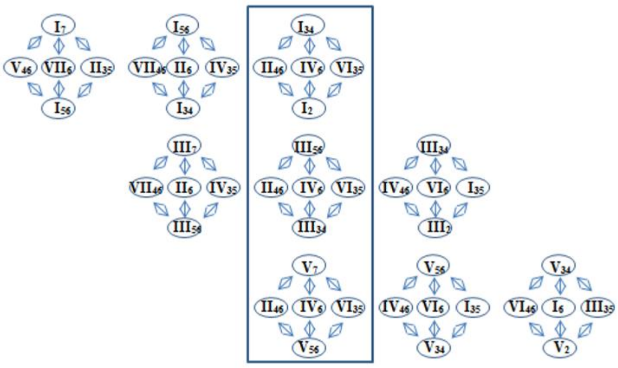

Figure 3. A vertical triad groups the graphs, sharing the same information inside edges.

A harmonic sequence is therefore represented as a path within the system, connecting the nearest instances of the chord in the system of graphs.

3.2 The color mapping

The second element of the visual representation methodology within a framework of this project consists of an augmented score, which explicitly identifies the structure of chords with their adherence to musical functions (tonic, dominant, subdominant).

The color gradient aims to visually distinguish the tones of a chord in relation to their importance in the structure of a chord (an example for the tonic seventh chord is shown in Figure 4). Such an explicit representation becomes particularly useful for an open position of chords – a range of open positions of the tonic triad and seventh chord with their inversions is shown in Figure 5.

Another visual contribution is based on a palette of colors, gathering the degrees in groups of functions – tonic (I), dominant (V, III, and VII degrees) and subdominant (IV, II and IV degrees), as shown in Figure 6.

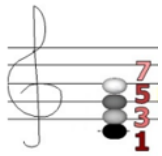

Figure 4. Color shades, that show the importance of a tone inside a chord structure.

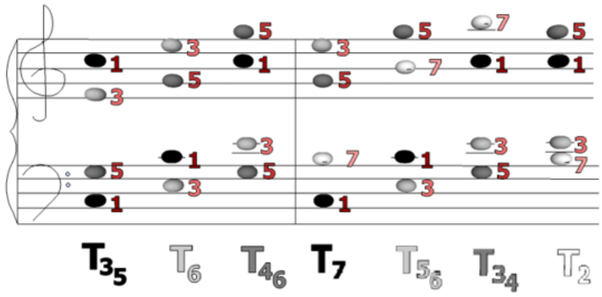

Figure 5. The chord inversions in an open position might create difficulties for students in finding the fundamental tone of the chord. The explicit shading of tones helps to lift the complexity of the graphical representation, allowing to focus the attention on the chord phonism.

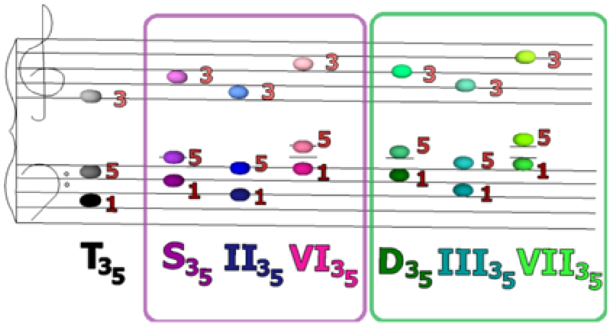

Figure 6. The color palette for the function groups: tonic (T), subdominant (S, II, VI) and dominant (D, III, VII).

The application of this representation methodology in a pedagogical process in a blended learning format was very successful [2, 3], which can be explained by the conformity of the graph structure with the possibilities of the short memory (episodic buffer) functioning, able to retain four to seven elements at a time (depends on a person). Indeed, a graph in a constructed system contains five elements, where two elements are grouped together, because they represent the same function and contain the information about 4-notes chords (the top and the bottom vertices of the graph); the remaining three elements are grouped by their structure, as they represent triads, being in a relation of the interval of a third between them.

4. VR APPLICATION

The VR project "Graphs in harmony learning" is based on the original methodology of tonal harmony representation described previously. The methodology elements receive a tangible 3D embodiment in a context of VR, enabling multimodal interaction. The VR context aims to increase a degree of immersion and therefore facilitate understanding and practice of the learning material.

The content of the application is divided into several VR rooms – the *Entry* (see Figure 7), explaining the purpose of the application, the *Temple of Knowledge* (Figure 8), which prepares the user to understand the content of the main activity, the *Study Room* (Figure 9), where the interaction with the graph system occurs. The *Test Room* and the *Practice Room* (Figures 10 and 11), propose the activities to test the acquired knowledge and put it in practice. Both latter rooms are based on AI models.

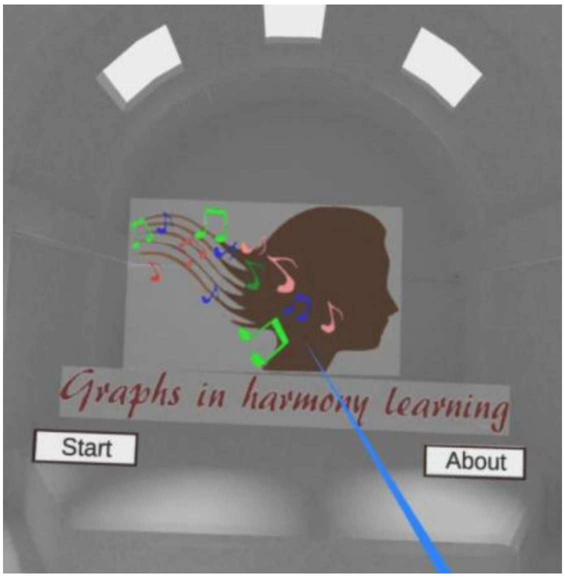

Figure 7. *Entry.*

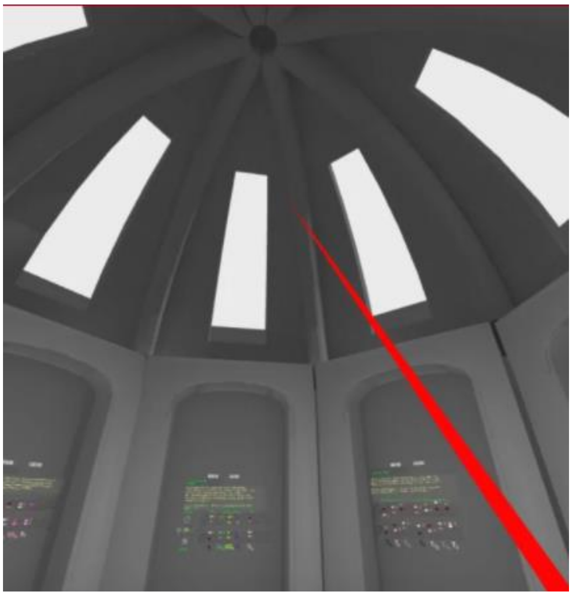

Figure 8. *Temple of Knowledge.*

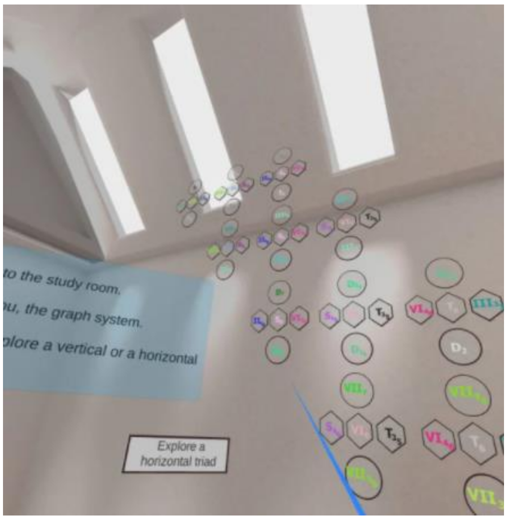

Figure 9. *Study Room.*

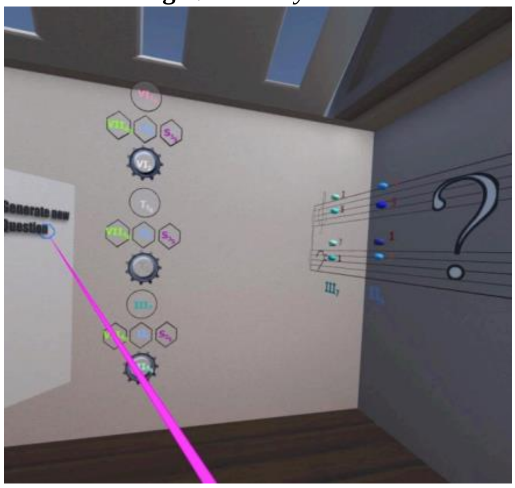

Figure 10. *Test Room.*

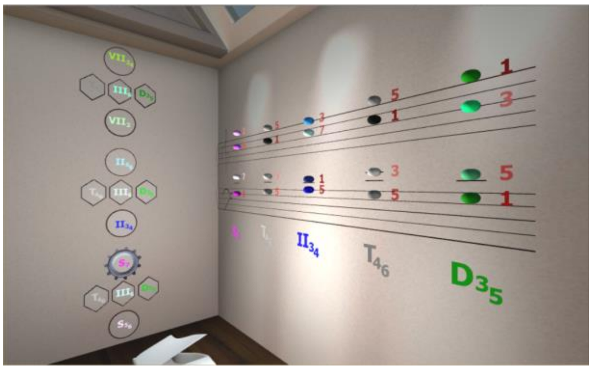

Figure 11. *Practice Room.*

The user experience begins in the *Entry* room, where the additional information about application can be seen using the *About* button; it also allows entering the *Temple of Knowledge* space using the *Start* button.

The *Temple of Knowledge* contains the explanations to provide the basics of music theory necessary for a better learning experience. This VR space contains 8 walls, 7 of which are occupied by a singular explanatory card.

The *Study Room* presents an interactive system of graphs with two modes of content exploration: *Horizontal Triad* and *Vertical Triad*. Both modes allow the user to choose either between guided and free learning experiences.

During guided learning experience, the graph element for interaction is highlighted by an animation (shown in Figure 12). If the user chooses this graph element, it triggers the augmented score appearance, the sequence audio playback and two animations: chords appearance on the staff and inside the system of graphs (both synchronized with the audio). In the guided learning experience, the interaction elements of the graph are put in a defined order and promote the discovery of all harmonic sequences from the chosen graph triad.

In case of a free learning experience, user chooses the interaction elements of the graph independently. The same mode allows to repeat a sequence as many times as needed. A free learning experience is recommended after completing guided learning experience.

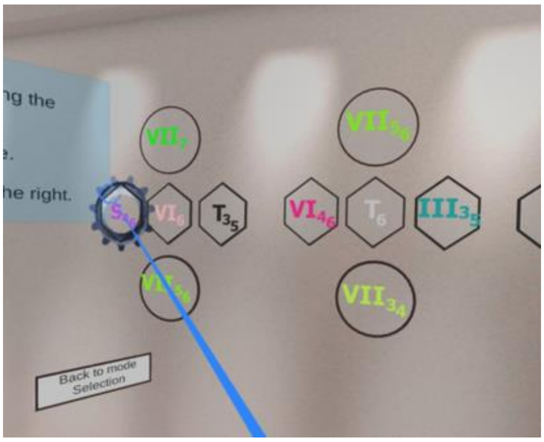

Figure 12. Highlighting edges to interact with in the *Horizontal mode*.

Horizontal Triad mode allows a horizontal graph exploration, allowing the discovery of potential for variability in passing chords (the edges in the structure of the graph). User can choose to return to a main view in order to choose another horizontal triad or switch modes at any time during the learning experience.

Vertical Triad mode gives the ability to explore the entire vertical triad, revealing the potential for harmonic variability for the seventh chords and their progressions (tops in the graph structure). The user still interacts with the graph edges to trigger the score appearance, but the choice of the edge must be made in a graph of interest inside the vertical triad. As with the horizontal triads, the user can choose to return to the main menu at any time.

The spatialization of the sound is integrated to reinforce memorization of the sequences of chords, since each chord finds its distinct spatial placement in a graph structure.

5. NOVEL DATA ENCODING

The value of the graph representation consists not only in the systematization of knowledge representation, useful in a pedagogical context, but also for the encoding of semantic music data. The main interest of such encoding lies in a possibility of training and evaluation methods application, inherent to the visual domain. This way, harmonic sequences features learning may be done with the use of 2D convolutional LSTM layers, discussed below. Another advantage of such encoding is the possibility of considerable data augmentation with the application of a small variance term during the normalization process.

To obtain the feature maps out of the arrays of harmonic sequences, the following transformation steps must be performed:

1. Chords dictionary creation.
2. Two-dimensional matrix creation using the graph system frame and replacing the chords with the values from previously created dictionary.
3. Creation of 28x28 matrices of harmonic paths in a 2D space using harmonic sequences mapped to the chords dictionary in a progressive way: one sequence of 5 chords results into 5 matrices, gradually filled in with the chord values.
4. Normalization of the feature maps with a changing normalization term (in a range between 0.01 and 0.1).

Using this data conversion strategy, it was possible to obtain 2160 data entries out of the initial handcrafted 216 harmonic sequences.

6. PREVIOUS MODELS

The main restriction of this project consists in the obligation of generating diatonic harmonic sequences in a C key only, since the mapping of 3D objects is made exclusively for this tonality. Therefore, the usage of pretrained models on multi-tonal music examples with possible alterations was not an option. Hence the need for developing and training own AI models to support the activities in a *Test Room* and *Practice Room*.

For the *Test Room* a simple long short-term memory (LSTM) deep neural network architecture with one embedding layer, two LSTM layers and two dense (linear) layers was build, using *Keras* framework. The model had 36,196 trainable parameters. The activation function of the LSTM layers was hyperbolic tangents (default *Keras* settings), the activation function of the first dense layer was ReLU (Rectified Linear Unit) and of the second dense layer – softmax. Adam optimizer with learning rate of 0.001 was applied as well. The training was done with 500 epochs with the batch size of 9 sequences. The autoregressive nature of the model allowed predicting the 3rd chord of the sequence, given two previous chords as an input. The architecture of the model is given in Figure 13.

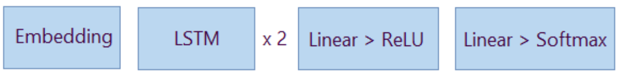

Figure 13. LSTM.architecture

For the *Practice Room*, a conditional generative adversarial (GAN) networks was developed using *Pytorch* deep learning framework. Conditional GAN model consisted of convolutional layers (transposed 1D convolution for the generator and 1D convolution for the discriminator) and was generating 4 chords given the first chord as an input. The number of input classes was equal to 7, representing 7 degrees of the diatonic scale. The architecture of the model is presented in Figure 14.

The training data for both models were partially hand-crafted, partially taken from a Kaggle dataset *Classical Music*. The harmony information extraction was done with the *Music21* Python library application [25], and further tokenization and features extraction – using *Pytorch* framework tools.

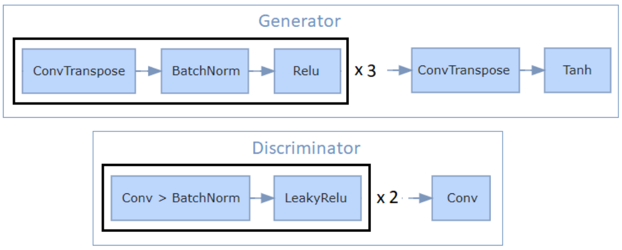

Figure 14. Conditional GAN architecture.

Trained models were stored in a cloud using *Flask* framework for performing inference. *JSON* protocol was used to exchange data between models and the VR application.

7. EXPERIMENT WITH THE NOVEL DATA ENCODING SCHEME

A new data encoding scheme have been tested in unsupervised setting, for which a ConvLSTM model was developed. The model's architecture comprised a doubled stack of convolutional 2D LSTM layers, followed by 3D batch normalization (third dimensionality was necessary to account for time steps information). The tangent activation function was used for gated mechanism inside a ConvLSTM cell. The sigmoid activation function was applied to the output of the final convolutional 2D layer, generating prediction on the input batch. The entire model architecture is shown in Figure 15.

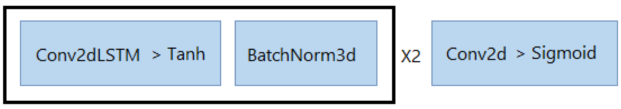

Figure 15. ConvLSTM model architecture.

Measurements of the 2D encoding scheme efficiency is therefore made by comparing the ConvLSTM model, trained on 2D data and the LSTM model, trained on the equivalent 1D data. For the sake of consistency, the previous LSTM model was rewritten in *Pytorch* framework.

The similar hyperparameters were applied to both models, such as mini-batch size (15 sequences per batch), optimizer (Adam), learning rate (1e-4), sequence length (5 chords in a sequence), number of epochs (100 epochs). The loss function for LSTM and ConvLSTM layers were cross entropy and binary cross entropy respectively.

The validation strategy for the received result was k-folds with the k value equal to 10, meaning that for each network architecture, 10 models were trained on different parts of the training and validation splits.

One-dimensional data consisted of 216 sequences of 5 chords each, where 4 chords served as an input and the 5th chord was treated as a target. In two-dimensional data split, each chord of the sequence was converted into a 28x28 matrix form, where the first 4 matrices represented the input features and the last matrix represented a target. This way, four input chords conditioned the prediction of the final chord of the sequence.

7.1 LSTM vs ConvLSTM

The result of the comparison between two models (presented in Table 1) shows that although LSTM model had smaller loss values, comparing to the ConvLSTM model, the difference in the end of training between train and validation loss for the LSTM model augments, compared to the beginning of training. Moreover, the loss of the LSTM model for the validation split becomes bigger in the end of training, which points to overfitting problem and a feeble generalization capacity. On the contrary, ConvLSTM model, having started with bigger loss values in the beginning of training, ends up with times smaller loss values. The data used for training and validation were not augmented at this stage.

Model	Train start	Val start	Train end	Val end
LSTM	0.26	0.29	0.10	0. 37
ConvLSTM	10.10	53.60	0.37	0.58

Table 1. Comparison of data dimensionality augmentation tested with LSTM and ConvLSTM models.

7.2 Data augmentation

The second stage of the experiment intended to measure the efficiency of data augmentation, made possible with the varying normalization term, using the novel encoding method. ConvLSTM model was therefore trained on datasets with and without data augmentation (2160 vs 216 data entries). The result of this stage of the experiment is shown in Table 2.

Data type	Train start	Val start	Train end	Val end
Non-augmented	10.10	53.60	0.37	0.58
Augmented	0.06	1.56	0.01	0.38

Table 2. Comparison of the data augmentation tested with ConvLSTM model.

The application of data augmentation technique has shown much better results in all four columns of the table – the first feedforward hidden representation is considerably better for the model trained on augmented data, which ameliorates the validation result at the beginning of training. In the end of training, the loss for train and validation splits substantially diminishes, meaning that the model learned well the input data shape, especially compared to the loss for the model trained on non-augmented data.

8. CONCLUSIONS

The article presented the use of a graph representation application in both ways: as knowledge representation method in a context VR application for ear training and as a novel data encoding technique applied to the generative models training. The novel encoding scheme was tested with ConvLSTM model, designed to process two-dimensional sequential information, being compared to the LSTM model with the similar hyperparameters, trained on the equivalent one-dimensional data. The results of the experiment have shown an important training result amelioration, especially when the data augmentation technique was applied. Finally, a high relevance of the design patterns used in a current VR project has been demonstrated via comparison with the recent development in the field.

7. REFERENCES

[1] M. Minsky, "A framework for representing knowledge," in P. Winston (Ed.), *The psychology of computer vision*, New York: Winston, 1974, pp. 211-277.

[2] A. Shvets and J. Jemielnik, "Visualization Strategies in the E-learning Course on Music Harmony at University – Effects and Possibilities," in Andrzej Michalski (Ed.), *Music Pedagogy. Ideals. Values. Pragmatics*, Gdansk: Athenae Gedanenses, 2016, pp. 218-229.

[3] A.S hvets, "Contemporary methods of functional harmony teaching in a high school context," in *Proceeding of Electronic Imaging & the Visual Arts (EVA Florence)*, Florence, 2019, pp. 142-150.

[4] A.S hvets, "The system of graphs in music harmony: a user interface for mobile learning game development," in *Proceeding of Electronic Visualisations and the Art (EVA London)*, London, 2016, pp. 193-194.

[5] A.S hvets and S. Darkazanli, "Graphs in harmony learning: AI assisted VR application," in *Proceeding of Electronic Visualisations and the Art (EVA London*, London, 2020, pp. 104-105.

[6] O. Zawacki-Richter, V. I. Marin, M. Bond and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education–where are the educators?" in *International Journal of Educational Technology in Higher Education, 16(1)*, pp. 1-27.

[7] A. Lino, A. Rocha and A. Sizo, "Virtual teaching and learning environments: automatic evaluation with artificial neural networks," *Cluster Computing*, 22(3), 2019, pp. 7217-7227.

[8] B. Zhang, H. Xia and H. O. Lim, "Development of an A.I. based teaching assisting system," in *International Conference on Artificial Intelligence, Information Processing and Cloud Computing*, Sanya, China, 2019, pp. 1-5.

[9] W. Villegas-Ch, A. Arias-Navarrete and X. Palacios Pacheco, "Proposal of an Architecture for the Integration of a Chatbot with Artificial Intelligence in a Smart Campus for the Improvement of Learning," *Sustainability 2020, 12(4):1500*, https://doi.org/10.3390/su12041500

[10] L. Turchet, R. Hamilton and A. Camci, "Music in Extended Realities," in *IEEE Access*, *9*, 2021, pp. 15810-15832.

[11] H. Liu and Y. Yang, "Lead sheet generation and arrangement by conditional generative adversarial network," in *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018, pp. 722-727.

[12] Y. Yu and S. Canales, "Conditional LSTM-GAN for melody generation from lyrics," in *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *17*, 2021, pp. 1-20.

[13] Z. Wang, D. Wang, Y Zhang, Y. and G. Xia, "Learning interpretable representation for controllable polyphonic music generation," in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR 2020)*, Montréal, 2020, pp. 662-669.

[14] H. H. Tan and D. Herremans, "Music FaderNets: Controllable music generation based on high-level features via low-level feature modelling," in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR 2020)*, Montréal, 2020, pp. 109-116.

[15] S. Lattner, M. Grachten and G. Widmer, "Imposing higher-level structure in polyphonic music generation using convolutional restricted boltzmann machines and constraints," in *ArXiv*, *abs/1612.04742*, 2016.

[16] A. Pati and A. Lerch, "Is disentanglement enough? On latent representations for controllable music generation," in *ArXiv, abs/2108.01450*, 2021.

[17] A. Pati, S. Gururani and A. Lerch, "dMelodies: a music dataset for disentanglement learning" in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR 2020)*, Montréal, 2020, pp. 125-133.

[18] D. Makris, K.R. Agresand and D. Herremans, "Generating lead sheets with affect: a novel conditional seq2seq framework," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, 2021, pp. 1-8.

[19] H. Liu and Y. Yang, "Lead sheet generation and arrangement by conditional generative adversarial

network," in *Proceedings of the 7th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Florida, 2018, pp. 722-727.

[20] A. Shvets and S. Darkazanli, "Conditional GAN for Diatonic Harmonic Sequences Generation in a VR Context" in *Proceeding of EVA London 2021*. British Computer Society (BCS): London, 5th–9th July 2021, pp. 97-100.

[21] S. Li and Y. Sung, "INCO-GAN: variable-length music generation method based on inception model-based conditional GAN," in *Mathematics 2021*, 9*(4):387*, https://doi.org/10.3390/math9040387

[22] S. H. I. Xingjian, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting" in *Advances in neural information processing systems*, 2015, pp. 802-810.

[23] A. Shvets, "Structural harmony method in the context of deep learning on example of music by Valentyn Sylvestrov and Philipp Glass," in *Proceeding of EVA London 2020*, London, 2019, pp. 318-320.

[24] A. Shvets and M. Desainte-Catherine, "Schemographe: Application for a new representation technique and methodology of analysis in tonal harmony" in *Proceedings of International Conference on Evolutionary and Biologically Inspired Music and Art*, Copenhagen, 2016, pp. 212-223.

[25] M. Cuthbert and C. Ariza, "music21: A Toolkit for Computer-Aided Musicology and Symbolic Music Data," in *Proceedings of International Conference on Music Information Retrieval*, Utrecht, 2010, pp. 637-642