

**A COMPARATIVE ANALYSIS OF POLYPHONY RNN AND
PERFORMANCE RNN TRAINED ON FOUR-VOICE FUGUES IN
GENERATING HARMONIES IN THE STYLE OF J.S. BACH**

A Thesis

Presented to the

Department of Information Systems

and Computer Science

Ateneo de Manila University

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Computer Science

by

Carmencita G. Dolina

Ana Patricia E. Fuentes

2023

ABSTRACT

This study compared Polyphony RNN and Performance RNN, two RNN models provided by Magenta, in generating harmonies in the style of J.S. Bach given a user-defined soprano input. The models were trained on 30 of Bach's 4-voice fugues from Art of Fugue and Well-Tempered Clavier I and II. Upon generation, select features were extracted from the MIDIs using jSymbolic. Objective evaluation tests were modeled after Fang et al.'s [3] Bach chorale grader function. Objective evaluation test results were verified through human evaluation tests wherein participants classified themselves based on their musical knowledge and ability then graded select clips from the models based on how much they sounded like Bach. As a result of both objective and human evaluation tests, neither model produced definitively Bach-like generations, but Performance RNN slightly outperformed Polyphony RNN in both rounds of testing.

ACKNOWLEDGEMENTS

We would like to thank Dr. Andrei D. Coronel for his guidance and support as our undergraduate thesis adviser. We also extend our gratitude to our panelists Dr. Jenilyn A. Casano and Mr. Jose Alfredo De Vera III for their insights and suggestions. We are grateful to the entire Department of Information Systems and Computer Science for assisting us throughout the implementation of our thesis.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	vii
CHAPTER	
I	INTRODUCTION
	1.1 Context of the Study
	1.2 Research Objectives.....
	1.3 Research Questions.....
	1.4 Scope and Limitations
	1.5 Significance of Study.....
II	REVIEW OF RELATED LITERATURE.....
	2.1 Discrimination Tests and Human Evaluation
	2.1.1 Objective Evaluation Tests
	2.1.2 Human Discrimination Tests.....
	2.2 Attributes of Bach's Compositions.....
	2.3 Recurrent Neural Networks
III	METHODOLOGY
	3.1 Data Preparation
	3.2 MIDI Conversion and Model Training.....
	3.2.1 MIDI Conversion
	3.2.2 Model Training.....
	3.2.3 MIDI Generation
	3.2.4 Final Feature Extraction
	3.3 Objective Discrimination Tests
	3.4 Human Discrimination Tests
	3.5 Summary
IV	RESULTS
	4.1 Model Training Accuracy Results
	4.2 Objective Test Results
	4.3 Human Discrimination Test Results.....
V	CONCLUSION.....
	5.1 Quantifiable Features of Bach's Compositions
	5.2 Objective Findings.....
	5.3 Human Discrimination Test Findings.....
	5.4 Polyphony RNN vs Performance RNN
	5.5 Recommendations.....
BIBLIOGRAPHY	

APPENDIX

A	TRAINING DATASET	34
B	RESULTS OF THE GRADING FUNCTION	36
C	RESULTS OF THE HUMAN DISCRIMINATION TESTS	39

LIST OF FIGURES

Figure 3.1 Methodology Diagram.....	9
Figure 3.2 Chorale Grader Function by Fang et al. [3].....	13
Figure 3.3 Critical D-value for a Two-sided KS Test with $\alpha = 0.05$ [12]	14
Figure 4.1 Polyphony RNN Metrics/Accuracy	18
Figure 4.2 Performance RNN Metrics/Accuracy	18
Figure 4.3 Bach Fugues Objective Grade Distribution.....	19
Figure 4.4 Polyphony RNN Objective Grade Distribution.....	19
Figure 4.5 Performance RNN Objective Grade Distribution.....	20
Figure 4.6 Polyphony RNN Subjective Grade Distribution	23
Figure 4.7 Performance RNN Subjective Grade Distribution	25

LIST OF TABLES

Table 3.1 Feature List and Equivalent Feature Sources	12
Table 3.2 Methodology Timetable.....	17
Table 4.1 KS-Test Results	21
Table 4.2 Average Grade Results of Polyphony RNN	23
Table 4.3 Average Grade Results of Performance RNN	25
Table 4.4 Preference Count Results	27
Table A.1 Four-voice Fugues	34
Table B.1 Bach Grader Function Results	36
Table B.2 Polyphony RNN Grader Function Results	37
Table B.3 Performance RNN Grader Function Results	37
Table C.1 Clip 1-A Results (Performance RNN)	39
Table C.2 Clip 1-B Results (Polyphony RNN).....	40
Table C.3 Clip 2-A Results (Polyphony RNN)	42
Table C.4 Clip 2-B Results (Performance RNN).....	43
Table C.5 Clip 3-A Results (Performance RNN)	44
Table C.6 Clip 3-B Results (Polyphony RNN).....	46
Table C.7 Clip 4-A Results (Polyphony RNN)	47
Table C.8 Clip 4-B Results (Performance RNN).....	48
Table C.9 Clip 5-A Results (Performance RNN)	49
Table C.10 Clip 5-B Results (Polyphony RNN).....	51

CHAPTER I

INTRODUCTION

1.1 Context of the Study

There have been numerous studies on emulating Johann Sebastian Bach's musical and compositional style through machine learning [5],[8],[9]. In particular, most researchers were interested in his use of counterpoint and polyphonies.

This study aims to do a comparative analysis of two different machine learning methods in order to find out which works best for the music generation of harmonies in the style of J.S. Bach within the domain of computational music.

A fugue [6] is a polyphonic musical form that has a subject which is transposed and repeated across two or more voices. Whenever a voice is singing the subject or answer, another voice will sing a counter subject simultaneously. This subject then develops across multiple episodes with each voice treated as a melodic line. Most studies on generating Bach-style harmonies use chorales, so this study also aims to see if fugues, which are also polyphonic, are also sufficient in generating harmonies in the style of J.S. Bach.

1.2 Research Objectives

This study aims to compare two different machine learning methods in order to generate harmonies in the style of J.S. Bach's polyphonic pieces. That said, the main research objective is: To find the more accurate model between Polyphony RNN and Performance RNN to produce harmonies in the style of J.S. Bach when given a user-defined soprano subject. The sub-research objectives are:

- Determine the quantifiable features of Bach's compositions that are useful for this comparative analysis
- Determine how well generated compositions from the Polyphony RNN model pass for human-made (specifically Bach-made) compositions
- Determine how well generated compositions from the Performance RNN model pass for human-made (specifically Bach-made) compositions
- Determine if fugues are sufficient in generating harmonies in the style of J.S. Bach

1.3 Research Questions

This study aims to answer the following research question: Between Polyphony RNN and Performance RNN, which model is more accurate in producing harmonies in the style of J.S. Bach given a user-defined soprano subject? The following sub-research questions are:

- What are the quantifiable features of Bach's compositions that are useful for this comparative analysis?
- How well do the generated compositions from the Polyphony RNN model pass for human-made (specifically Bach-made) compositions?
- How well do the generated compositions from the Performance RNN model pass for human-made (specifically Bach-made) compositions?
- How sufficient are fugues in generating harmonies in the style of J.S. Bach?

1.4 Scope and Limitations

This study only focused on the four-voice compositions that Bach composed, thus the dataset used is composed of fugues from Well-Tempered Clavier I, Well-Tempered Clavier II, and Art of Fugue.

The group only used two different models, namely Polyphony RNN and Performance RNN, in conducting this study. These models were available through machine learning tools Magenta and Tensorflow and were primarily used to generate harmonies. The group also limited the study to the use of MIDI files in manipulating and generating music.

The MIDI manipulation software used were jSymbolic, music21, mido, and digital audio workstations (DAWs) GarageBand and Logic Pro X. jSymbolic was used to extract features from the MIDI files, such as pitch statistics, rhythm statistics, and parallel fifths among others. The group primarily used music21 and mido to extract features not present on jSymbolic, such as length of pieces in bars and seconds. GarageBand and Logic Pro X were used to manually isolate fugues from preludes for some MIDIs and ensure the number of tracks before training.

The tools used for objective evaluation were pandas, numpy, and scipy in Python. These were used in grading each MIDI and conducting Kolmogorov-Smirnov statistical tests on the resulting grade distributions.

This study mainly targeted machine learning practitioners and computer scientists. Beyond the field of computer science, the general music-listening population would also potentially benefit from this study.

1.5 Significance of Study

In the field of computational music and computer science, this study will contribute to the studies of existing machine learning models in generating harmonies, particularly

in the studies regarding Bach's compositions. Beyond this, the study aims to investigate how human composers can utilize automated processes as brought about by technology and machine learning.

CHAPTER II

REVIEW OF RELATED LITERATURE

2.1 Discrimination Tests and Human Evaluation

Both objective evaluation and human discrimination tests are crucial in comparative analyses in computational music, so much so that evaluating music generation methods remains to be a complex subtopic in the field. Yang et al. [13] assessed multiple subjective and objective methods of evaluation in music generation, highlighting the greater priority given to the former and thus the need to improve the latter. Liu et al. [9] also pointed out the disparity between evaluation metrics and human perception in their study using the Bach chorales dataset, noting that favorable accuracy scores do not necessarily imply musically sound outputs.

2.1.1 Objective Evaluation Tests

There have been many different means of evaluating data objectively in music generation studies. Yang et al. [13] classified these methods based on the features utilized in evaluation. Some methods primarily evaluate generated compositions based on “probabilistic measures” that do not require domain-specific musical information, though results from these methods do not necessarily align with results from their human evaluation tests. Notably, some evaluate compositions based on specific features either their models were meant to highlight or their datasets already showed. Additionally, other researchers choose to evaluate compositions based on features more familiar to those in the field of music, thus being inherently tied to a musical domain and more interpretable to human participants. Due to the study’s goal of evaluating

generated compositions based on their emulation of Bach's style, the group prioritized using features based on these two latter approaches.

Specific to Bach, Fang et al. [3] presented a grading function which is based on certain chorale features that assigns a numerical value to four-part chorales generated to sound like that of Bach's. In their study, they got the Wasserstein distances between each feature distribution of a generated chorale or set of chorales and each feature distribution of the actual Bach chorale dataset, then got the weighted sum of these distances across all features to result in a numerical grade. To test this function, they graded a group of pieces generated by a transformer model. They then conducted a Kolmogorov-Smirnov statistical test on both the grade distribution of true Bach chorales and grade distribution of generated compositions to determine how separated these grade distributions were. This method of evaluation was useful to the group in evaluating the generations produced by the models used.

2.1.2 Human Discrimination Tests

In a notable study, Liang et al. [8] conducted one of the largest human musical discrimination tests on their website for their BachBot model. Participants totaled to 2,336 at the time of publication and has likely increased since then. Participants were classified into 4 groups: novice, intermediate, advanced, and expert. The groups are characterized as casual listeners, individuals who know how to play an instrument, individuals who studied musical composition, and individuals who teach or do research about music respectively. The test was composed of five questions where participants would have to determine which of two audio files featured work truly composed by Bach. Hadjeres et al. [5] conducted a similar discrimination test for their DeepBach

model. Participants were only given one audio file at a time and were made to determine if “Bach” or a “computer” composed the work.

2.2 Attributes of Bach’s Compositions

In determining whether fugues are sufficient embodiments of the features of Bach’s compositions as a whole (stated in sub-research question 4), it was important to include features across multiple types of compositions.

Giraud et al. [4] created an algorithm to be able to detect if a piece was a fugue or not by detecting specific elements of the fugue form, which include the structure of an exposition, episodes, and end. They used MIR techniques to detect repeated patterns, specifically the subject, countersubject, as well as harmonic sequences which can be found in the episode.

There is some overlap between Bach fugue attributes and Bach chorale attributes, such as polyphony, repeated musical patterns, and a typically strict adherence to the rules of counterpoint. As mentioned previously, Fang et al. [3] developed a grading function for generated chorales that grades against Bach chorales given certain features. The grading function is based on 6 features crucial to Bach chorales, namely pitch, rhythm, intervals, harmonic qualities, parallel errors, and repeated sequences.

Notably, Mogren [10] evaluated generations from his proposed continuous recurrent neural network-generative adversarial network (C-RNN-GAN), which was trained on a classical music dataset including works by Bach beyond chorales and fugues, based on measures of polyphony, scale consistency, repetitions, and tone span. To answer the first sub-research question, these aforementioned features were used in the objective evaluation portion of the methodology.

2.3 Recurrent Neural Networks

Recurrent neural networks (RNNs) are commonly used in text generation due to their dependence on sequential data to predict the next output. [9] Many studies on music generation utilize RNNs due to the sequential structure of a musical composition. Musical compositions have long- and short-term dependencies [9] that RNNs are able to accommodate better than typical feedforward neural networks. Feedback loops in the RNN's hidden layers allow the output to be predicted based on both the current and previous inputs. This will be relevant to the group's study due to the repetition of the subject in a fugue.

Many ML studies on the generation and harmonization of or in the style of Bach's compositions have also previously used RNN and its variants.

Notably, Liang et al. [8] developed the BachBot model. It makes use of sequential encoding in order to process note sequences via a RNN with long short term memory (LSTM) cells to account for long-term dependencies. Polyphony RNN, which is one of the models the group will analyze, is the Magenta implementation of the BachBot model.

Like Polyphony RNN, Performance RNN is also LSTM-based, but generates the MIDI's to be more expressive and dynamic compared to other models that tend to quantize into 16th notes [11]. This, combined with all the notes having the same volume, makes it sound less natural as it lacks the varying dynamics of notes human-made music has.

Therefore, both models utilize LSTM-based recurrent neural networks and are both designed to produce polyphonic music, but they differ in how well they can generate music that sounds human-made.

CHAPTER III

METHODOLOGY

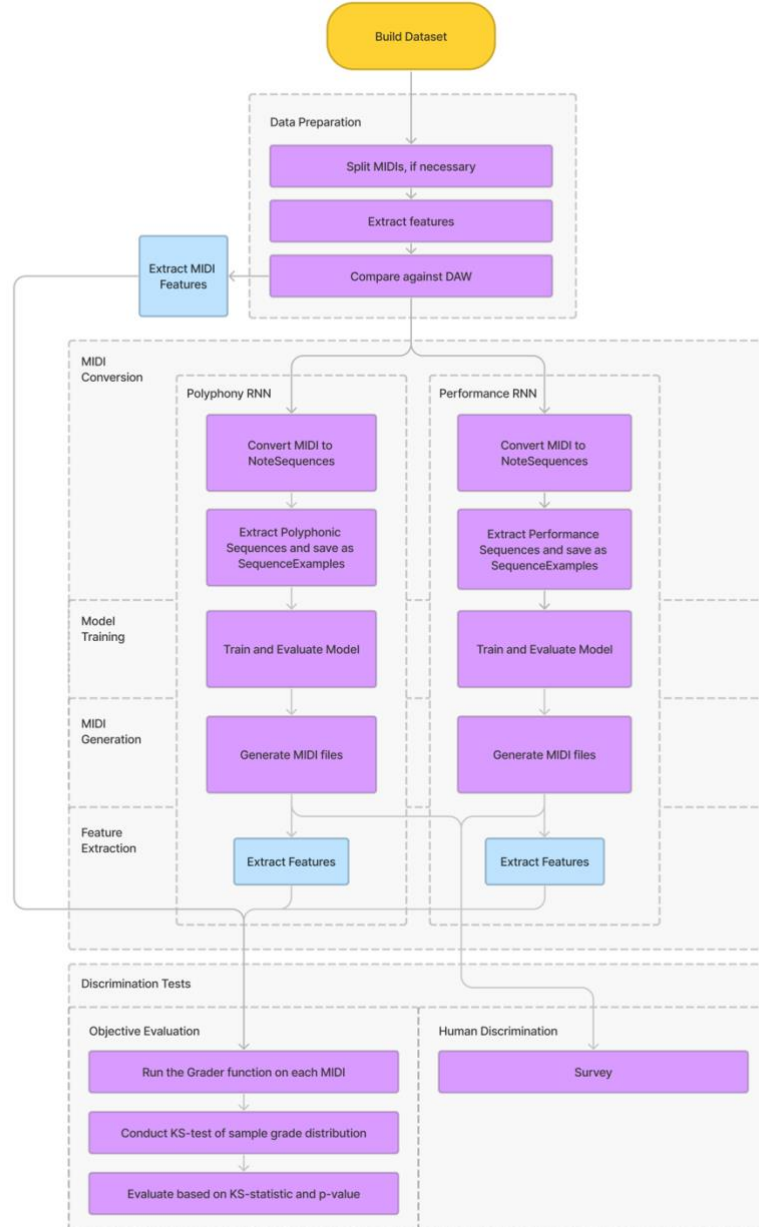


Figure 3.1 Methodology Diagram

The main dataset for this study was composed of Bach four-voice fugues from both Art of Fugue and Well-Tempered Clavier I and II. In total, 30 fugues were used to train each model. Separated MIDI files for each fugue were taken from the kunstderfuge website.

The methodology can be broken down into four stages: Data Preparation, MIDI Conversion and Model Training, Objective Discrimination Tests, and Human Discrimination Tests.

3.1 Data Preparation

Data preparation mostly consisted of cutting down the MIDI files in length. The fugues of Well-Tempered Clavier I were all separated from their preludes to begin with but the Well-Tempered Clavier II MIDI files had to be spliced manually on a DAW, specifically GarageBand or Logic Pro X in this case, in order to isolate their fugues. Additionally, some pieces from Art of Fugue could not be found in standalone MIDI files so these were extracted from a single, compiled version of the entire Art of Fugue using the same method mentioned previously.

Each original MIDI underwent feature extraction on jSymbolic to determine the number of independent voices it had. This number was then compared against the number of tracks shown on DAWs. The MIDI files were assured to be clean when both jSymbolic and the DAWs detected four voices and four tracks respectively.

The group extracted features from the dataset of 30 original Bach four-voice fugues using jSymbolic. The data extracted from the original Bach fugues was compared and evaluated against the data extracted from the generated MIDI files later in the methodology.

This took around two weeks.

3.2 MIDI Conversion and Model Training

3.2.1 MIDI Conversion

This stage required Magenta in order to convert the MIDI of Bach fugues into NoteSequences. Note sequences are a more efficient data format that are easier to work with compared to MIDI files. This step required a directory of MIDI files in order for Magenta to take in as an input and generate an output of NoteSequences. The generated NoteSequences was used as input to generate SequenceExamples for both models, Polyphony RNN and Performance RNN.

To convert NoteSequences to SequenceExamples, it required giving Magenta the input directory in order to produce SequenceExamples. SequenceExamples are the input given to the models to train and evaluate. For this paper, the group used an evaluation ratio of 0.1, meaning 10% of the extracted tracks were saved to be evaluated, while 90% were used in the training collection. Since there are two models, this step was done twice in order to generate the SequenceExamples to two different directories from the same input of NoteSequences, one for Polyphony RNN and one for Performance RNN.

3.2.2 Model Training

The group trained their own instances of the Polyphony RNN and Performance RNN models with the SequenceExamples generated from the Bach fugue dataset.

Magenta required a directory to save the data, the directory of SequenceExamples, number of steps before exiting the loop, and hyperparameters to specify the batch size. This was done for each model.

While training, an evaluate job was run in parallel to produce graphs in TensorBoard that the group used to compare the accuracy of Performance RNN against Polyphony RNN.

Training stopped once the accuracy plateaued.

3.2.3 MIDI Generation

Performance or polyphonic tracks were generated every 5000 steps during training. Magenta required inputting the number of outputs and steps, which indicates the length of each melody. The temperature was set to 0.9 to reduce randomness to make the outputs more repetitive. To replicate user input, the group used the command of primer midi, which is a MIDI of an isolated subject from Art of Fugue 1. The subject was isolated by splitting the MIDI and deleting the alto, tenor, and bass tracks on a DAW. For MIDI generation, the group decided on ten outputs, each 256 steps long (16 bars), or 3200 ms long, from each model, using the same primer melody.

Once the accuracy of the model plateaued on Tensorflow, one last round of MIDI generation was made by using the primer melody command to create 10 original tracks per model.

3.2.4 Final Feature Extraction

Table 3.1 Feature List and Equivalent Feature Sources

Feature in Chorale Grader [3] or C-RNN-GAN [10]	Equivalent jSymbolic Feature
Pitch [3]	Pitch Class Histogram
Scale Consistency [10]	Pitch Class Variability
Repeated Sequences [3] and Repetition [10]	Repeated Notes and Melodic Pitch Variety
Intervals [3] and Tone Span [10]	Average Interval Spanned by Melodic Arcs
Harmonic Qualities [3]	Chord Type Histogram
Rhythm [3]	Rhythmic Value Histogram

Polyphony [10]	Voice Overlap, Parallel Motion, Similar Motion, Contrary Motion, Oblique Motion
Parallel Errors [3]	Parallel Fifths and Parallel Octaves

Once the MIDI files have been generated, the group extracted features from each file using jSymbolic. The chosen features were based on the original features used by Fang et al. [3] in their chorale grader. Because some features used in their study do not map analogously to the features available on jSymbolic, the group approximated them when necessary. This feature list was also supplemented by features used by Mogren [10] in evaluating his proposed C-RNN-GAN. The full list of features used, with their approximated equivalent features from the study by Fang et al. [3], can be found in Table 3.1. Features were extracted in batches, resulting in separate feature distributions for original Bach fugues, Polyphony RNN generations, and Performance RNN generations.

The MIDI Conversion and Model Training section took five weeks.

3.3 Objective Discrimination Tests

$$g(c) = \sum_{f \in \text{features}} w_f \cdot \text{Wass}(P_c^f, P_{\text{Bach}}^f)$$

Figure 3.2 Chorale Grader Function by Fang et al. [3]

The separate feature distributions across the three groups of MIDI files went through the chorale grader function in Figure 3.2 presented by Fang et al. [3] using the features mentioned in the previous section. This was implemented on a jupyter notebook using numpy, pandas, and scipy. For each feature, the group took the Wasserstein distance between a MIDI's score in that feature and the ground-truth distribution of Bach's

scores in that feature. This process was repeated for each feature of the given MIDI. The sum of all Wasserstein distances resulted in a grade for that given MIDI. While the original chorale grader function makes use of a weighted sum instead of a normal summation, specifically to penalize the occurrences of parallel errors in a generation, the group decided to forgo this aspect of the function since jSymbolic's Parallel Fifths and Parallel Octaves features already result in a fractional score. It is important to note that, according to the function, a lower score implies a better generation.

Using the grader function on each MIDI resulted in three grade distributions: one for original Bach fugue grades, Polyphony RNN grades, and Performance RNN grades. Following the method used by Fang et al. [3] in their experiments, the group conducted Kolmogorov-Smirnov tests for each model's grade distributions against the original Bach fugue grade distribution to determine how likely it is that the samples come from the same distribution. This resulted in a KS-test statistic and p-value for each model. With a significance level of 95%, the grade distributions would be deemed significantly different should the KS-test statistic be greater than the critical D-value given by the equation in Figure 3.3.

Objective Discrimination Tests took five weeks.

$$D_{0.05} = 1.36 \sqrt{\frac{m+n}{mn}}$$

Figure 3.3 Critical D-value for a Two-sided KS Test with alpha = 0.05 [12]

3.4 Human Discrimination Tests

A human discrimination test in the form of an online survey was done in order to validate the results from the objective discrimination tests. Prior to this test, five of the ten outputs per model were chosen for use in the survey. The primer melody, which

was automatically placed at the beginning of each generation, was removed using a DAW to both shorten the time allotted for the test and allow the participants to focus less on structure and more on harmonization.

Before the evaluative part of the human discrimination tests, the group gathered data regarding the participants themselves. Much like the discrimination test implemented by Liang et al. [8] among others, the group asked participants for their age in age ranges, i.e. 18–25, 26–45, 46–60, and Over 60. Additionally, participants were asked to categorize themselves based on how well-versed they think they are in classical music. Based on the self-categorization portion of the test implemented by Liang et al. [8], the categorizations are Novice, Intermediate, and Expert. To aid participants in categorizing themselves, further criteria were added based on the music theory placement exams from Iowa State University's Department of Music and Theatre [7].

1. Novice

- I cannot read sheet music.
- I was not previously trained in playing instruments or only play instruments casually.
- I have not listened or do not frequently listen to classical music.

2. Intermediate

- I can read sheet music (treble clef and bass clef, key signatures, time signatures).
- I can identify different voices (SATB) in polyphonic pieces.
- I can identify major, minor, augmented, diminished, perfect intervals.
- I know what relative and parallel majors/minors are.
- I am training/have previously been trained in music theory/performance.

- I listen to classical music and am familiar with different composers' styles.

3. Expert

- I am studying or have previously studied music history and theory at the undergraduate or post-graduate level.
- I am teaching or have previously taught music history and theory at the undergraduate or post-graduate level.
- I can identify the types of motion between voices (contrary, parallel, etc.).
- I can notate a 7th/9th/11th chord.
- I am familiar with the compositional styles of multiple classical composers including J.S. Bach.

For the evaluative part of the human discrimination tests, the group first let participants listen to a snippet of an episode from Art of Fugue 1, isolated through a DAW. This ideally gave participants a better understanding of Bach's harmonization and style without lending too much context to the structure of a fugue. Participants then listened to two unlabeled generated MIDI outputs, one generated using Polyphony RNN and another generated using Performance RNN. After this, they had to rank each audio clip on a scale of 1 to 10 based on how well each output emulated the style of Bach's four-voice compositions. They also had to explain in a short paragraph form why they ranked them a particular way, and choose which clip they preferred in each pair. This process of listening to and ranking two outputs from each evaluated model was done over the five sets of audio clips previously chosen for the discrimination tests. Collection of survey responses ended once there were ten novice, ten intermediate, and five expert respondents.

This took five weeks.

3.5 Summary

Below is a timetable with the allotted time frames for each stage.

Table 3.2 Methodology Timetable

Week	Stages of Development
1-2	Data Preparation (cutting MIDIS, extracting features)
3-7	MIDI Conversion (MIDI -> NoteSequences -> SequenceExamples) Model Training (both mid-, post-training generation and extraction) Feature Extraction
8-12	Objective Discrimination Tests
11-15	Human Discrimination Tests

CHAPTER IV

RESULTS

4.1 Model Training Accuracy Results

The training statistics of both Polyphony RNN and Performance RNN were monitored on Tensorflow's visualization software Tensorboard.

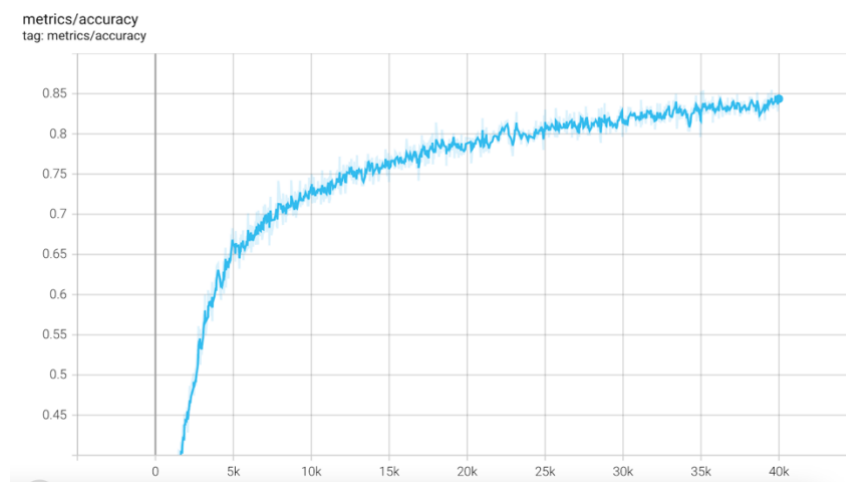


Figure 4.1 Polyphony RNN Metrics/Accuracy

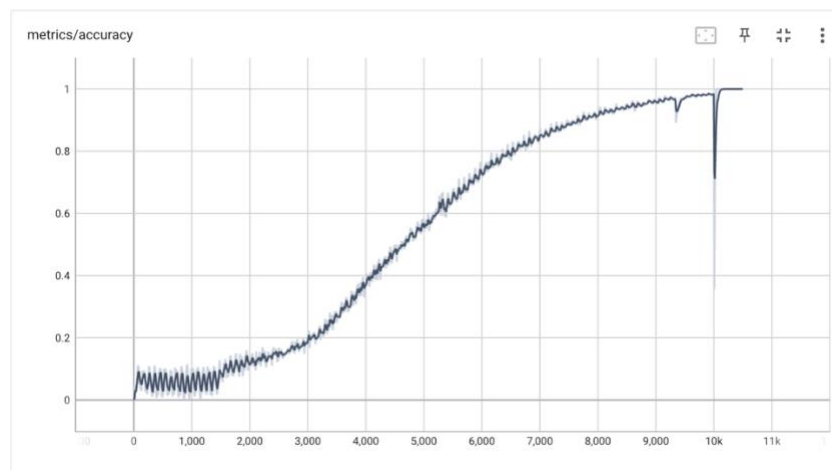


Figure 4.2 Performance RNN Metrics/Accuracy

The results of the training, in figures 4.1 and 4.2, show that Polyphony RNN plateaued at 84.70% after 40,000 steps while Performance RNN plateaued at 99.98% after 10,491 steps.

4.2 Objective Test Results

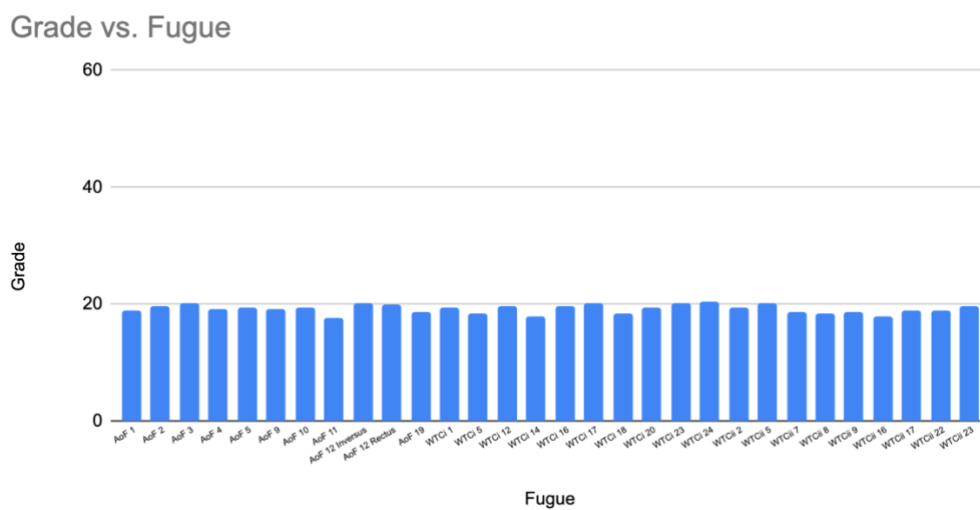


Figure 4.3 Bach Fugues Objective Grade Distribution

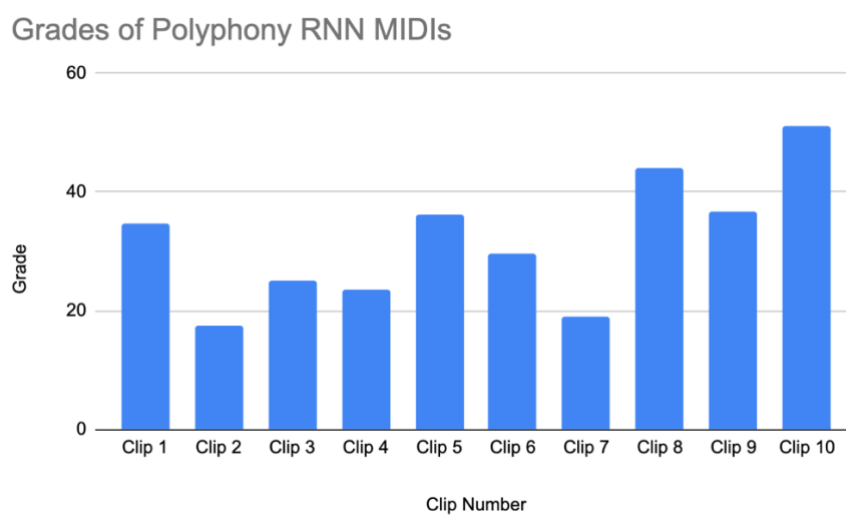


Figure 4.4 Polyphony RNN Objective Grade Distribution

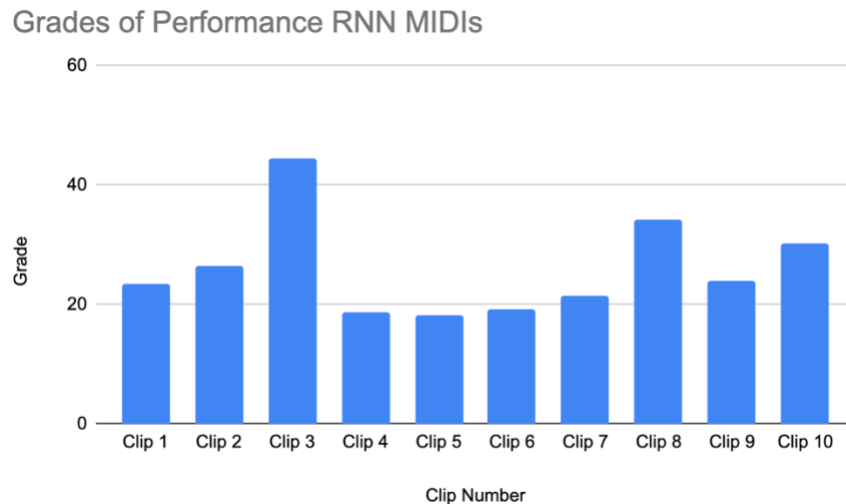


Figure 4.5 Performance RNN Objective Grade Distribution

The grade distributions of the original Bach fugue dataset, Polyphony RNN generations, and Performance RNN generations were evaluated using the chorale grader function presented by Fang et al. [3].

The Bach fugues generally got grades, in Appendix B Table B.1, from 17–20, with the lowest score being Art of Fugue Contrapunctus 11’s grade of 17.5698 and the highest being Well-Tempered Clavier I Fugue No. 24’s grade of 20.4988. This implies that Bach compositions, especially his four-voice fugues, are consistently written across all features extracted for this evaluation. Bach four-voice fugues, compared within and across all of his fugue books, are fairly similar to each other.

The same cannot be said for Polyphony RNN and Performance RNN generations. On the one hand, Polyphony RNN MIDIs’ grades, in Appendix B Table B.2, ranged from 17–51, with the lowest being Clip 2’s grade of 17.5959 and the highest being Clip 10’s grade of 51.1828. On the other hand, Performance RNN MIDIs’ grades, in Appendix B Table B.3, ranged from 18–44, with the lowest being Clip 5’s grade of 18.2103 and the highest being Clip 3’s grade of 44.5431. Generally, both models

produced objectively inconsistent compositions when evaluated against the chosen feature set.

To better understand how large the differences between each model's grade distribution and Bach's grade distribution are, a Kolmogorov-Smirnov test was conducted per pair as seen in Table 4.1. The KS test for the Polyphony RNN grade distribution against the Bach grade distribution resulted in a KS-statistic of 0.8 and a p-value of 0.0000322. Similarly, the KS test for the Performance RNN grade distribution against the Bach grade distribution resulted in a KS-statistic of 0.7 and p-value of 0.0005886. Comparing each KS-statistic to the critical value D given by the formula in Figure 3.2, it can be seen that they are each larger than the critical value. Comparing the p-values against the chosen significance level of 95%, it is noted that both p-values are smaller than 0.05. Due to both these factors, and despite Performance RNN outperforming Polyphony RNN in the KS tests, there is sufficient evidence to claim that both the Polyphony RNN grade distribution and the Performance RNN grade distribution are significantly different when compared with the Bach grade distribution. This means that it is inconceivable for both models' grade distributions to be samples of Bach's grade distribution, and thus highly unlikely for model generations to objectively be emblematic of Bach's style.

Table 4.1 KS-Test Results

Model	KS-Statistic	P-Value
Polyphony RNN	0.8	0.0000322
Performance RNN	0.7	0.0005886

There are numerous possible reasons as to why Performance RNN marginally outperformed Polyphony RNN. First, as mentioned previously, Performance RNN was able to reach a 99.98% accuracy rate in training while Polyphony RNN only reached

an 84.70% accuracy rate in training according to the findings on Tensorboard. Additionally, while both models are relatively similar, Performance RNN takes into consideration expressive dynamics and velocity. It performs significantly better when trained on MIDIIs of live performances, and thus finely quantized MIDIIs, because of this [11]. Since both models were trained on MIDIIs captured by live performance, with dynamics and velocities that Polyphony RNN does not account for, it is possible that this caused Performance RNN to generate outputs marginally more similar to Bach. However, because of the aforementioned objective test results, it still stands that neither model generates statistically conceivable Bach-like outputs. This also implies that there may be more lacking in the dataset rather than the architectures of the models themselves.

4.3 Human Discrimination Test Results

Grades refers to the answer respondents gave to the question “How much does Clip X-Y sound like J.S. Bach?” on a scale of 1–10, meaning “Not at all” to “Exactly like Bach”. Figures 4.6 and 4.7 show the distribution of ratings of all the clips from Novices (Blue), Intermediates (Red), and Experts (Yellow), while Table 4.2 and 4.3 show the average grade per clip. A score of 7/10 and above has been defined as Bach-like.

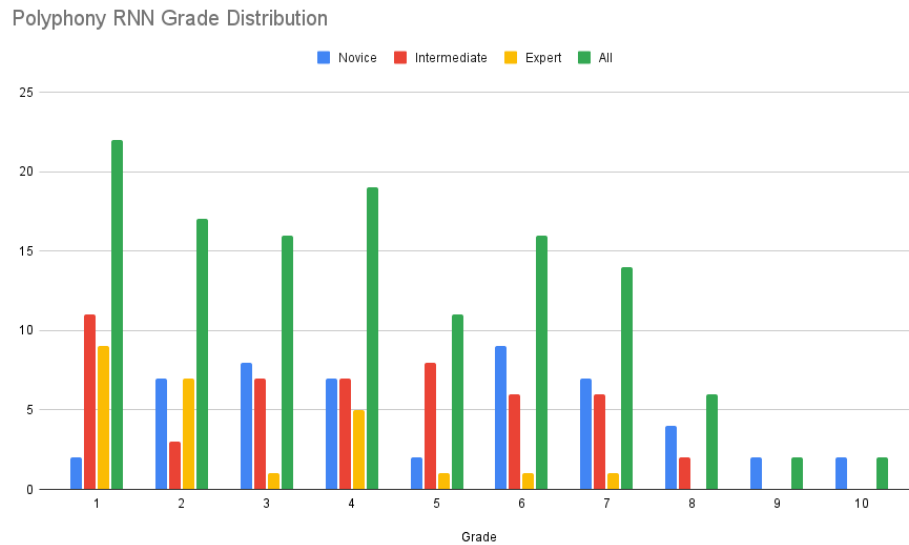


Figure 4.6 Polyphony RNN Subjective Grade Distribution

The overall most common grade was a score of 1. Intermediates and experts contributed to this, where both types of evaluators voted a grade of 1. Intermediates capped the highest grade at an 8, while experts capped at a 7. However, this was not true for novices. A grade of 1 was one of the least common answers, with only 2 participants voting for it. The most common answer among novices was a 6, and the highest grade given was a 10.

Table 4.2 Average Grade Results of Polyphony RNN

Clip No.	Novice (10)	Intermediate (10)	Expert (5)	All (25)
1	5.2500	3.3750	3.3333	3.7200
2	4.1250	5.2500	2.3333	4.2800
3	3.3750	3.5000	1.0000	3.0800
4	7.5000	5.0000	4.0000	5.2800
5	5.3750	4.3750	3.0000	4.2400
Total Average	5.1250	4.3000	2.7333	4.1200

As seen from Figure 4.6, the Average Grades ended up with novices giving the highest average grade among the type of evaluators, followed by intermediates and experts, who graded the harshest. The total average was a 4.12, which means that this model did not consistently produce outputs that sound Bach-Like.

The short responses for novices with regard to Polyphony RNN's clips would often refer to the feeling evoked by the compositions. Those who graded the clips from a 1–3 would say they sounded “chaotic”, “ragged”, “sudden”, and “feeling off” compared to the original snippet. For those who graded the clips from 7–10, they noted that the bpm, notes, and feeling seen in the clips were closer to those of Bach.

The intermediate respondents' reason for a grade of 1–3 would refer to the lack of counter melody, “messy” or “awkward” chord and rhythm choices. For a grade of 7–10, they would note that the melody and chords matched better, noticed the counter melody, and believed the clips had a similar tempo and mood to Bach.

Experts who gave the clips a grade of 1–3 mentioned that it has “no personality” or “no intention” behind it. The rhythms of melodic ideas were very “clunky”, or had a “blocky texture”. The choices such as the “rapid 16th notes without variation in rhythm” were not aligned with “Bach’s principles of composition”. The highest grades given by experts for Polyphony RNN clips ranged from 4–7, as seen in Figure 4.6. The reasons given for these grades were the inconsistent textures and some “jarring” parts such as the “discontinuity” of melodies. Ultimately, the choices were closer to Bach’s ideas, but they did not quite embody them.

A full table of responses can be found in Appendix C. For Polyphony RNN responses, refer to Table C.2, C.3, C.6, C.7, and C.10.

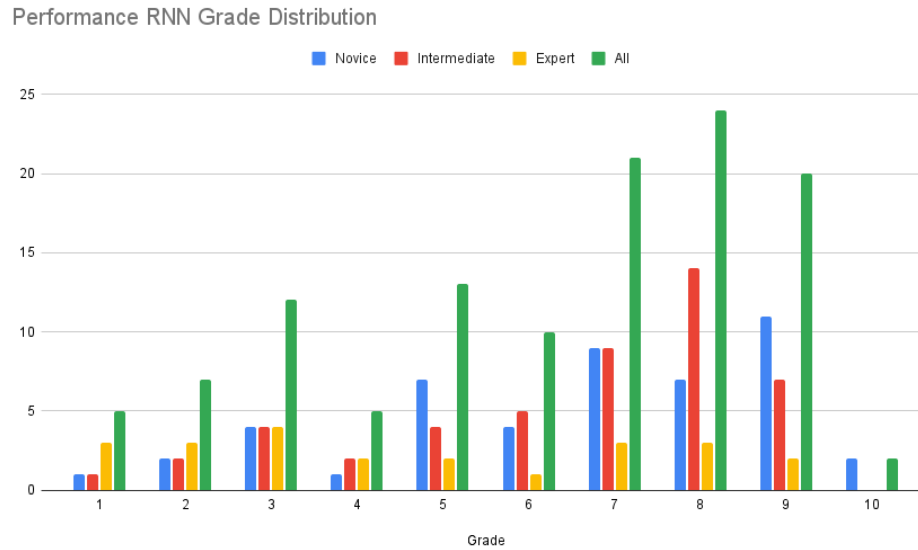


Figure 4.7 Performance RNN Subjective Grade Distribution

The most common answer overall as seen in Figure 4.7 for Performance RNN clips was an 8, however the most common answer for experts was a 3. Intermediates voted 8 the most, while novices voted 9 the most.

Table 4.3 Average Grade Results of Performance RNN

Clip No.	Novice (10)	Intermediate (10)	Expert (5)	All (25)
1	6.5000	6.7500	4.0000	6.0000
2	6.1250	6.6250	5.3333	6.4400
3	6.8750	6.2500	4.3333	5.7600
4	7.5000	7.3750	6.3333	7.1600
5	5.7500	6.7500	4.6667	5.4800
Total Average	6.5500	6.7500	4.9333	6.1680

Table 4.3 shows that intermediates rated Performance RNN clips the highest, followed by novices, then experts. The total average was 6.1680, so Performance RNN also did not consistently produce outputs that sound Bach-like.

The responses evaluating Performance RNN clips were a little different compared to those evaluating Polyphony RNN clips. The short responses from novices who graded it a 1–3 believed that the clips sounded “melancholic” compared to Bach’s “happier tone”, or that they sounded more like “modern day music”. For a higher grade of 7–10, they sounded “pleasing”, “consistent”, “smooth”, and sounded like it could be an “initial draft of Bach”.

Intermediates who rated the clips from 1–3 also noticed that some sounded more contemporary than Bach due to their phrasing and harmony. They also had a difficult time tracing the exchange between voices and thought some clips sounded “slow” and “sad”, which are characteristics they do not associate with Bach. For a grade of 7–10, they mentioned that the rhythm and tonality sounded like a Bach fugue. They were able to detect the counterpoint and notice a similar structure to the fugue in terms of polyphony.

Experts who rated the clips from 1–3 instead thought that they sounded more like “Mozart” or “Chopin” instead of Bach, as some were “closer to romanticism than classical”. They would also mention the discordant harmony and lack of clear melodic ideas. When the grades got higher in the range of 7–10, the experts thought that the melodies were more “baroque” and had “hints of fugal entries”. Additionally, they noted that the use of a pentatonic scale and diatonic notes were reminiscent of Bach. For one clip (Appendix C, Table C.8), the “walking bass in the left hand with the right hand melody” reminded a participant of Bach and overall sounded musical, having “breaths where expected”.

A full table of responses can be found in Appendix C. For Performance RNN responses refer to Table C.1, C.4, C.5, C.8, and C.9.

Table 4.4 Preference Count Results

Preferred	Novice (10)	Intermediate (10)	Expert (5)	All (25)
Polyphony RNN	13	7	5	25
Performance RNN	37	43	20	100
Total Count	50	50	25	125

For each pair of clips, the participants preferred the MIDI generations of Performance RNN over Polyphony RNN's generations 80% of the time, as seen in Table 4.4.

CHAPTER V

CONCLUSION

5.1 Quantifiable Features of Bach's Compositions

Answering the sub-research question “What are the quantifiable features of Bach’s compositions that are useful for this comparative analysis?”, the helpful quantifiable features of Bach’s compositions were Pitch, Scale Consistency, Repeated Sequences and Repetition, Intervals and Tone Span, Harmonic Qualities, Rhythm, Polyphony, and Parallel Errors. The equivalent jSymbolic Features are the Pitch Class Histogram, Pitch Class Variability, Repeated Notes and Melodic Pitch Variety, Average Interval Spanned by Melodic Arcs, Chord Type Histogram, Rhythmic Value Histogram, Voice Overlap, Parallel Motion, Similar Motion, Contrary Motion, Oblique Motion, Parallel Fifths, and Parallel Octaves.

5.2 Objective Findings

In answering “How well do the generated compositions from the Polyphony RNN/Performance RNN model pass for human-made (specifically Bach-made) compositions?”, it is clear that from the results given by the grader function and the KS tests of each grade distribution, both Polyphony RNN and Performance RNN do not produce generations that could convincingly pass for Bach-made compositions. While Performance RNN outperformed Polyphony RNN in terms of grade distribution, with the former’s KS test p-value being 0.0005886 and the latter’s KS test p-value being 0.0000322, it stands that neither model consistently generated MIDI files with harmonies in the style of Bach.

In answering “How sufficient are fugues in generating harmonies in the style of J.S. Bach?”, it is likely that fugues are an insufficient composition type as training data in generating harmonies in the style of Bach. While the objective test results could say something about the models’ architectures themselves, it is important to note that Performance RNN reached 99.98% accuracy in model training but still did not produce MIDIIs with consistently Bach-like grades, thus implying an insufficiency in the training dataset itself.

5.3 Human Discrimination Test Findings

There were notable insights relating to the types of evaluators. In general across both models, grades from novices had a wider spread but seldom reached the extremes. The intermediate grade spread was narrower than that of the novices but was wider than that of the experts, which had the narrowest spread.

For Polyphony RNN, intermediate and expert grades both had a mode of 1, while novice grades had a mode of 6 seen in Figure 4.6. There were a few clips that scored 7 and above, but they did not consistently sound like Bach, especially to experts and intermediate participants. To note a specific example, Clip 4 in Table 4.4 was given an average grade of 7.5 by novices. However, when looking at the total average grade, novices, intermediates, and experts graded it 5.125, 4.300, and 2.730 respectively, which means that the clip did not consistently sound Bach-like.

Performance RNN had more positive results. Intermediate participants and novices were much more generous in grading, with modes of 8 and 9 respectively as seen in Figure 4.7. There were many pieces that did sound Bach-like to intermediate participants and novices, but they did not consistently sound like Bach, thus resulting in an average grade of 6.75 and 6.55 respectively in Table 4.3. Despite both average

grades not reaching a score of 7, clips from this model generally sounded more Bach-like to both casual listeners and those with basic music theory knowledge.

The results of the human evaluation tests validate the findings from objective evaluation tests. Performance RNN outperformed Polyphony RNN, but neither of the models produced generations that were reminiscent of Bach. Thus, in answering the sub-research questions, “How well do the generated compositions from the Polyphony RNN/Performance RNN model pass for human-made (specifically Bach-made) compositions?”, Performance RNN did better than Polyphony RNN. According to Table 4.4, the respondents preferred Performance RNN generations 80% of the time, which may be due to it feeling more human based on the short responses seen in Appendix C. While there are many factors that could have led to this, it is possibly due to the differences between how the two models work. Performance RNN focuses on making generations that sound like a real human performance, so it performs better when trained on input that was not directly quantized (i.e. live performances) and has varying dynamics that make the outputs sound less monotonous. This could have resulted in pieces that sound human-like.

In answering “How sufficient are fugues in generating harmonies in the style of J.S. Bach?”, based on current results, fugues are not very sufficient in generating harmonies in the style of J.S. Bach consistently.

5.4 Polyphony RNN vs Performance RNN

Finally, in answering the main research question “Between Polyphony RNN and Performance RNN, which model is more accurate in producing harmonies in the style of J.S. Bach given a user-defined soprano subject?” Performance RNN is more accurate, but only by a small amount. Performance RNN did slightly outperform

Polyphony RNN in the objective evaluation tests, but it still did not consistently produce Bach-like outputs. It also outperformed Polyphony RNN in the human discrimination tests, especially in terms of preferred clips, but was not consistently graded the decided score of 7 to count as Bach-like.

5.5 Recommendations

As this study primarily focused on training the models on four-voice fugues, and after having found fugues insufficient as a dataset, it would be interesting to conduct similar studies with training datasets of other specific or broader groups of composition types.

If time permits, it would be better to widen the sample size and receive more responses across the different skill levels. It would also help to specifically solicit responses from more musical experts as this would allow for certain grading trends to be more noticeable. There were multiple clips that evoked differing opinions, such as Clip 2-B for Performance RNN (Table C.4). Some experts rated it low, saying it had no clear melodic ideas, while some rated it high and said the opposite. Having more responses from experts might allow for the grades of the survey to be less divisive.

BIBLIOGRAPHY

- [1] ALVIRA, J.R. 2013. Classification of J. S. Bach's 3 and 4 voices fugues according to the order of voices entry in the exposition. URL=<https://www.teoria.com/en/articles/circles/index.php>.
- [2] BOULANGER-LEWANDOWSKI, N., BENGIO, Y., VINCENT, P. 2012. Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. ICML'12: Proceedings of the 29th International Conference on Machine Learning, (2012) 1881-1888.
- [3] FANG, A., LIU, A., SEETHARAMAN, P., PARDO, B. 2020. Bach or Mock? A Grading Function for Chorales in the Style of J.S. Bach. Machine Learning for Media Discovery Workshop at the International Conference on Machine Learning, (2020).
- [4] GIRAUD, M., GROULT, R., LEGUY, E., LEVÉ, F. 2015. Computational Fugue Analysis. Computer Music Journal, 39(2), (2015) 77-96.
- [5] HADJERES, G., PACHET, F., NIELSEN, F. 2017. DeepBach: a steerable model for bach chorales generation. ICML'17: Proceedings of the 34th International Conference on Machine Learning, 70 (2017) 1362-1371.
- [6] HERSCOVITCH, D. 2017. Decoding the music masterpieces: Bach's The Art of Fugue. URL=<https://theconversation.com/decoding-the-music-masterpieces-bachs-the-art-of-fugue-73522> .
- [7] IOWA STATE UNIVERSITY DEPARTMENT OF MUSIC AND THEATRE. Iowa State University Sample Fundamentals Placement Test. URL=https://www.music.iastate.edu/files/page/files/sample_fundamentals_placement_test_1.pdf.

- [8] LIANG, F., GOTHAM, M., JOHNSON, M., SHOTTEN, J. 2017. Automatic Stylistic Composition of Bach Chorales with Deep LSTM. 18th International Society for Music Information Retrieval Conference, (2017), 449-456.
- [9] LIU, I., RAMAKRISHNAN, B. 2014. Bach in 2014: Music Composition with Recurrent Neural Network.
- [10] MOGREN, O. 2016. C-RNN-GAN: Continuous recurrent neural networks with adversarial training. Constructive Machine Learning Workshop (NIPS 2016), Barcelona, Spain.
- [11] OORE, S., SIMON, I., DIELEMAN, S., ECK, D., SIMONYAN, K. 2018. This Time with Feeling: Learning Expressive Musical Performance. Neural Computing and Applications. 32, (2020) 955–967.
- [12] SALAH, A., ROHATGI, V. 2015. An Introduction to Probability and Statistics, 3rd Edition. John Wiley & Sons, Inc.
- [13] YANG, L., LERCH, A. 2018. On the evaluation of generative models in music. Neural Computing and Applications. 32, (2020) 4773–4784.

APPENDIX A

TRAINING DATASET

A.1 List of Four-voice Fugues by J.S. Bach

Table A.1 Four-voice Fugues	
Fugue	Fugue Book
Fugue / fuga I	Well-tempered Clavier I
Fugue / fuga V	Well-tempered Clavier I
Fugue / fuga XII	Well-tempered Clavier I
Fugue / fuga XIV	Well-tempered Clavier I
Fugue / fuga XVI	Well-tempered Clavier I
Fugue / fuga XVII	Well-tempered Clavier I
Fugue / fuga XVIII	Well-tempered Clavier I
Fugue / fuga XX	Well-tempered Clavier I
Fugue / fuga XXIII	Well-tempered Clavier I
Fugue / fuga XXIV	Well-tempered Clavier I
Fugue / fuga II	Well-tempered Clavier II
Fugue / fuga V	Well-tempered Clavier II
Fugue / fuga VII	Well-tempered Clavier II
Fugue / fuga VIII	Well-tempered Clavier II
Fugue / fuga IX	Well-tempered Clavier II
Fugue / fuga XVI	Well-tempered Clavier II
Fugue / fuga XVII	Well-tempered Clavier II
Fugue / fuga XXII	Well-tempered Clavier II
Fugue / fuga XXIII	Well-tempered Clavier II
Contrapunctus I	Art of Fugue
Contrapunctus II	Art of Fugue
Contrapunctus III	Art of Fugue

Contrapunctus IV	Art of Fugue
Contrapunctus V	Art of Fugue
Contrapunctus IX	Art of Fugue
Contrapunctus X	Art of Fugue
Contrapunctus XI	Art of Fugue
Contrapunctus XII (rectus)	Art of Fugue
Contrapunctus XII (inversus)	Art of Fugue
Contrapunctus XIX/Fuga a 3 Soggetti	Art of Fugue

APPENDIX B

RESULTS OF THE GRADING FUNCTION

B.1. Grader Results for All MIDIIs

Table B.1 Bach Grader Function Results

Fugue	Grade
Fugue / fuga I	18.9151
Fugue / fuga V	19.6085
Fugue / fuga XII	20.2229
Fugue / fuga XIV	19.0852
Fugue / fuga XVI	19.4942
Fugue / fuga XVII	19.1992
Fugue / fuga XVIII	19.3388
Fugue / fuga XX	17.5698
Fugue / fuga XXIII	20.1327
Fugue / fuga XXIV	20.0006
Fugue / fuga II	18.7019
Fugue / fuga V	19.3264
Fugue / fuga VII	18.4836
Fugue / fuga VIII	19.6064
Fugue / fuga IX	17.9604
Fugue / fuga XVI	19.7477
Fugue / fuga XVII	20.1616
Fugue / fuga XXII	18.3167
Fugue / fuga XXIII	19.5049
Contrapunctus I	20.2215
Contrapunctus II	20.4988
Contrapunctus III	19.4708

Contrapunctus IV	20.1558
Contrapunctus V	18.6591
Contrapunctus IX	18.4305
Contrapunctus X	18.7091
Contrapunctus XI	17.8692
Contrapunctus XII (rectus)	18.8349
Contrapunctus XII (inversus)	18.9644
Contrapunctus XIX/Fuga a 3 Soggetti	19.7127

Table B.2 Polyphony RNN Grader Function Results

MIDI Generation No.	Grade
1	34.6357
2	17.5959
3	25.1372
4	23.6366
5	36.2740
6	29.6423
7	18.8993
8	43.9096
9	36.8213
10	51.1828

Table B.3 Performance RNN Grader Function Results

MIDI Generation No.	Grade
1	23.3716
2	26.4328
3	44.5431
4	18.7477
5	18.2103

6	19.3104
7	21.3528
8	34.1677
9	24.0835
10	30.3001

APPENDIX C

RESULTS OF THE HUMAN DISCRIMINATION TESTS

C.1 Survey Short Answer Responses

Table C.1 Clip 1-A Results (Performance RNN)

Group	Grade	Responses
Novice	3	Different instrument. While we can argue that Clip 1-A most likely used the same sheet music Bach did, the addition of the extra note as well as the pacing made the piece completely different and unlike Bach's style (which has a faster pace + happier, leaning-to-major-notes style)
Novice	5	Based on the snippet, it "feel" it gives of is different.
Novice	6	it seems similar but not totally the same vibe
Novice	6	It shares a somewhat similar beat but sounds sadder
Novice	7	There was an added rhythm to the clip and the notes are not the same, but the pattern of how the notes go up are the same.
Novice	7	I think clip 1-a possess the same notes as the art of fugue 1 by J.S. Bach. I also think clip 1-a has the same beats per minute and progression.
Novice	7	Comparing it to the example from a while ago it feels like its progressing higher, but the start has a different vibe.
Novice	8	it feels kind of similar in pace and gentle
Novice	9	I noticed that as the chord progresses, a corresponding low pitched chord was being done that complements the high-pitched progression (for ex. in 0:10-0:12 mark, 0:12-0:20 mark). The 0:12 - 0:20 part of the video gave the low pitched chords some harmony and melody with each other and with the high-pitched chords.
Novice	10	it has a similar "pattern" as the Art of Fugue snippet from the previous section
Intermediate	1	there is no exchange between the voices in the treble and bass clef
Intermediate	3	Sounds kinda dissonant. Tempo is also slower than i expected. There doesn't seem to be the counter melody
Intermediate	3	Harmonies employed don't sound very typical of Bach
Intermediate	5	Reminiscent of it, but sounds more like Chopin.

Intermediate	8	slow and the notes seem to be going in a wave sequence wherein it goes back and forth between lower notes to higher notes to lower notes
Intermediate	8	the counterpoint kinda reminds of Bach but with a modern style
Intermediate	8	It emulates the rhythmic choices and phrasing of Bach. Some harmonic choices, however, lacked the dissonance heard in the sample fugue
Intermediate	8	It sounds quite similar to the sample in the beginning both in terms of rhythm and tonality.
Intermediate	9	I believe it is in the same key and tempo as Bach's fugue that was asked to be listened by us, and it utilizes similar harmonies or chords.
Intermediate	9	sounds similar to the previous clip
Expert	2	There is some discordant harmony that is created between the different melodic lines
Expert	3	It sounds more like Chopin than Bach.
Expert	4	too much dissonance in the left hand
Expert	5	While the melodic line is a bit Baroque, the counterpoint and harmony are not. It doesn't have the consistent motorythmic drive (note values too long) and the intervals don't really work..
Expert	6	Cuz

Table C.2 Clip 1-B Results (Polyphony RNN)

Group	Grade	Responses
Novice	1	I couldn't hear the original sound clip in this one anymore.
Novice	2	This flow was more inclined to 8-bit games music. The chords did not incorporate low-pitched chords unlike clip A, and this clip was somewhat monotone since it kept repeating a certain key.
Novice	3	The vibe is completely different, wherein the example clip from earlier had a fancy sort of vibe, this one felt more casual, and didn't seem to have a specific progression of going up or down.
Novice	3	it feels a lot faster and more ragged
Novice	5	This clip has a faster and the way the notes go up are much slower compared to the original one.

Novice	6	I think clip 1-b is a bit more chaotic than clip 1-b which makes it sound less like something J.S. Bach would have composed. Additionally, I think clip 1-b used more or unnecessary notes than clip 1-a.
Novice	6	the vibes don't match, idk how to explain lol i was listening to other pieces also to get more of a grasp (i hope that's okay??) and it's just not the same vibe, but the tempo still feels like one of Bach's faster pieces
Novice	7	seems a bit more similar to bach than the first clip
Novice	8	Same instrument, the fast pacing is there. Majority of the notes were probably played in major.
Novice	8	This clip feels more closer to the initial snippet given.
Intermediate	1	No exchange in the voices. Overused mordents
Intermediate	1	Harmonies chosen do not approximate Bach very well. More repetition of notes than Bach would have employed.
Intermediate	1	Does not sound like Bach, harmonic choices are not reminiscent of his work.
Intermediate	1	Vastly different phrasing and harmonic choices.
Intermediate	3	Too many stops. Doesn't seem like there's a counter melody, just random keys
Intermediate	3	the tempo is faster than the Bach reference clip and seems to be played in a much more different key than the reference clip and 1-A
Intermediate	4	a bit much dissonance and the style is kinda reminds me more scarlatti than bach
Intermediate	4	The melody sounds a bit too random, robotic almost.
Intermediate	6	it does sound like Bach but maybe a different piece as compared to Clip 1-A
Intermediate	7	The clip sounds too complicated to be like Bach's composition, given that the fugue that was asked to be listened by us would be the reference point.
Expert	1	The melodies are not very rhythmically or harmonically sound, they go against conventions of the time
Expert	1	Rapid 16th notes without variation in rhythm, and harmonic choices that are not aligned with Bach's principles of composition.
Expert	2	Melodic fragments are way too short for Bach, and the way they are developed is not very stylistic at all. Texture is not remotely Baroque and the rhythms are very stilted.
Expert	2	Not close
Expert	7	

Table C.3 Clip 2-A Results (Polyphony RNN)

Group	Grade	Responses
Novice	2	This sound clip sound closer to a video game theme song than Bach.
Novice	2	it feels uneven, especially at the end when it starts playing all the keys all at once
Novice	2	I dont think it sounds similar to the snippet.
Novice	3	It was only in the end of the clip that it sounded familiar to Bach.
Novice	3	Same instrument but the tone is melancholic. If I were to make a guess, a lot of the notes were played in minor?
Novice	4	it doesnt seem as similar
Novice	6	The low chords of this clip somewhat complements the high-pitched progression, but it sounded a bit monotone at some of the parts.
Novice	6	This one has the same vibe, and had the same thing with it feeling like it went a direction, but different from the example clip's progression.
Novice	7	The loudness of the notes makes clip 2-a sounds less like a J.S. Bach composition, also, i think the speed of clip 2-a is a bit slower than the previous clips.
Novice	7	the pause then tempo change is a bit strange, feels less "smooth" (if that make sense, it doesn't glide like the pieces i've heard), and there's a distorted bit at the end. i've listened to some of his slower pieces too and again, vibes don't seem to match
Intermediate	1	too many chords/thirds/fifths that made it sound messy
Intermediate	2	Not flow-y enough. Also doesn't seem to have much of a rhythm
Intermediate	3	Rhythm dosent feel like Bach's. Also Im not very good at hearing chords but I think bach wouldnt use that many same consecutive intervals at the end
Intermediate	4	No strong presence of counter melody.
Intermediate	5	Some compositional choices are not characteristic of Bach
Intermediate	5	Phrasing is similar, but, again, harmonic choices are quite different.

Intermediate	6	Bears similarity but it feels a lot more dramatic than the reference clip. Follows a similar back and forth progression between the notes
Intermediate	7	The first half sounded like a Bach composition however I felt that the mood changed after the second half, where there were chord changes per quarter note.
Intermediate	8	The melody was easier to follow and the chords at the end went well with it.
Intermediate	8	sounds like the melody of the sample piece
Expert	1	The harmony is very discordant in some sections, it sounds very randomised and undeliberate, which is atypical of a Bach piece
Expert	2	Nasty
Expert	4	The discontinuity of the melody is very jarring, and the texture is not built up very consistently at all.
Expert	4	just sounded like mashing keys together
Expert	5	Some choices are closer to Bach's ideas, but it's not 100% there yet.

Table C.4 Clip 2-B Results (Performance RNN)

Group	Grade	Responses
Novice	1	This sounds closer to modern day music
Novice	3	It's melancholic, not like Bach's which had a happier tone.
Novice	5	it seems slightly more similar to bach than the first but not totally the same vibe
Novice	6	It feels slightly more similar to the snippet than the last clip.
Novice	7	The build up and the rhythm sounds similar to the original clip.
Novice	8	it sounds pleasing and It doesn't startle me or do something so unexpected like 2-A
Novice	8	I think clip 2-b follows the same speed of clip 1 A and B which makes it sound like a J.S. Bach composition. I do not think clip 2-b should have high notes since I think Bach does not possess those high notes.
Novice	9	The clip from 0:10 onwards was able to harmonize both the low and high-pitched chords, creating a harmony. It felt as if there was direction from both of these two types of chords as the clip progressed.

Novice	9	This one felt a lot simpler than the example clip, but felt like it could be in the same song but further down the line.
Novice	9	much "smoother" compared to the previous clip, a lot more consistent
Intermediate	2	at the final segment, the thirds in the bass clef were not in response to the melody of the treble
Intermediate	4	No counter melody and too slow
Intermediate	5	Phrasing choices are different, but the harmonic choices are a bit more similar
Intermediate	6	Harmonic choices still don't sound like Bach but phrasing seems to be getting closer.
Intermediate	7	Same as before, the melody was easier to follow and the chords at the end went well with it (I just happen to prefer the previous clip).
Intermediate	8	the subject sounds like something Bach would make. But I think the third voice entered before the second voice finished the statement of the answer
Intermediate	8	Sounds more deliberately written. Hews closer to Bach's work.
Intermediate	8	Sounds similar to the reference clip and similar tempo
Intermediate	8	sounds like the melody of the sample piece
Intermediate	9	While it didn't sound exactly like Bach's fugue that was asked to be listened by us, I feel like it incorporates his style (starts slow, monophonic then increases in speed with more harmonies).
Expert	2	There are no clear melodic ideas, it feels very aimless
Expert	5	Decent
Expert	7	Melody is more baroque and hints of fugal entries are more apparent, though it doesn't sound like the second voice is exactly imitating the first. It also doesn't really develop enough.
Expert	8	Yeah, that's getting closer. Pentatonic scale transforming to more extended structures on melody while maintaining classical I-V relationships on the harmony.
Expert	9	actually sounded musical, had 'breaths' where you'd expected them

Table C.5 Clip 3-A Results (Performance RNN)

Group	Grade	Responses
-------	-------	-----------

Novice	3	Melancholic. The pacing is slower. If I were to take a guess, I hear more minor notes than major notes.
Novice	4	Most of the duration of this clip is somewhat monotone with very few changes in melody.
Novice	5	kinda slow but it sounds like something bach would compose
Novice	6	The start felt awkward, but the rest seemed like it could have been taken from some point in the song.
Novice	7	Feels similar
Novice	8	The start confused me a little but it began to sound pleasant again
Novice	8	I would rank clip 3-a the same as clip 2-b since it also has those high notes. Based on the clip of a J.S. Bach composition, it did not have high notes.
Novice	9	It's similar I can see it being an intro to one of his pieces.
Novice	9	It sounds like the original clip that was just tweak a bit.
Novice	9	not quite on the melody i think (not very harmonious, i think?), but the "pattern" is similar again to the Art of Fugue from earlier
Intermediate	2	there was more of a response but it's not that easy to trace the exchange of the melodies
Intermediate	4	Does not sound like a deliberate piece, is more arbitrary in tone.
Intermediate	5	Counter melody sounds ok. Bit dissonant at the start
Intermediate	6	With the exception of 1 or 2 dissonant notes, the phrasing and the counter melody are starting to sound more like Bach.
Intermediate	6	not as lively
Intermediate	6	The melody did not sound like a melody Bach would compose.
Intermediate	7	The melody is easy to follow and builds up nicely
Intermediate	7	Sounds similar, has the back and forth progression
Intermediate	8	The phrasing and harmonic choices come close.
Intermediate	8	apart from the dissonant counterpoint, the flow and rhythm of the voices reminds me a bit of Bach
Expert	1	Again, the discordant harmony and the lack of clear melodic ideas make it very unlike any Bach piece
Expert	1	Shocking, unprepared and unresolved dissonances, unusual textures and harmonies

Expert	3	While I really enjoy these types of compositions, it's closer once again to Chopin than Bach. Certain conventions applied to this clip are closer to the Romanticism period rather than Classical.
Expert	5	Not quite there
Expert	7	had some non-diatonic notes, which worked well

Table C.6 Clip 3-B Results (Polyphony RNN)

Group	Grade	Responses
Novice	1	no sorry its just me not you.
Novice	2	The start felt like it belonged to a like a JRPG boss fight, and the rest felt very random like keysmashes, didn't seem similar to the example clip.
Novice	2	it's ragged and uneven and suddenly goes silent
Novice	3	It doesn't sound happy like Bach's. It sounds like terror / scary music
Novice	4	it doesnt really sound like something bach would make
Novice	4	it doesn't have a consistent flow
Novice	4	It was only in the end of the clip where I think it was similar to Bach
Novice	5	the pauses again are a little weird (like it forgot the next notes and also sounds like is winging it after every pause), much less "smooth", and doesn't match vibes
Novice	6	Although there were low-pitched chords, these low-pitched chords did not harmonize well with the high-pitched progression. At times, some chords felt out of place and odd.
Novice	6	I do not think this sounds similar to a Bach composition because it has stops.
Intermediate	1	melodies are not in the baroque style; no exchange of melodies
Intermediate	1	Sounds heavier and dramatic, notes are more pronounced as compared to other clips wherein notes are...fleeting and light?
Intermediate	1	The rhythm choices were very very different, and harmonies were brighter
Intermediate	2	Again, the composition sounds more arbitrary and does not hew closely to Bach's style (except for the latter section)
Intermediate	3	A lot of stops and kinda dissonant, not as flow-y
Intermediate	3	Phrasing and harmonic choices sound awkward here.

Intermediate	5	sounds like a different genre
Intermediate	5	The second half does not possess the tempo and feeling of Bach's fugue that was asked to be listened by us.
Intermediate	5	The pauses in the melody feel random and not human like.
Intermediate	7	kinda reminds me of his fugue 2 in c minor weird rhythm and counterpoint at the end
Expert	1	It is very randomised and has no clear melodic ideas. The only thing tying it to Bach is the contrapuntal aesthetic, but the melodies have no substance
Expert	1	Bruh
Expert	1	didn't sound like something somebody would actually play
Expert	2	Very blocky texture, the rhythm is not very characteristic, and minimal harmonic logic
Expert	2	Over-reliance on tritones is not characteristic of a time when the church would burn you for intervals like that. However, the latter half sounds pretty Bach.

Table C.7 Clip 4-A Results (Polyphony RNN)

Group	Grade	Responses
Novice	2	sounds ragged again with a lot of sudden gaps and silent parts
Novice	4	The low-pitched chords did not seem to match well with the high-pitched progression. The low-pitched chords feel isolated and odd.
Novice	6	The build up is a bit similar to the original clip.
Novice	7	Although different, the sound and feel of it feels closer to the original snippet.
Novice	7	i think bach would make something like this
Novice	7	Clip 4-a still follows the same speed of a Bach composition but I still think Bach does not have those loud multiple notes at once.
Novice	8	the first few notes sounded like they came out of nowhere haha mildly "smooth", the tempo matches some pieces
Novice	9	It sounds happier. The fast (appropriate for happy music) pace is there. I'm guessing more major notes were used?
Novice	10	Sounds just like the clip but like better audio quality.
Novice	10	it sounds like classical music

Intermediate	1	chord progression is awkward and there is no exchange of melodies
Intermediate	2	The arbitrary nature is not reminiscent of Bach's style.
Intermediate	3	Rhythmic fluidity (phrasing) is quite similar, but harmonic choices, not
Intermediate	4	There are random stops in the music which i don't associate with bach
Intermediate	4	Phrasing is a bit more disjointed than one would have expected. Also some strange harmonic choices.
Intermediate	5	the last 10 seconds did sound like Bach to me but the counterpoint and melodies was kinda all over the place
Intermediate	6	The rhythm in the first half of the clip is a bit weird, but the chords in the latter half are interesting.
Intermediate	7	feels like it has similar tempo to the sample piece
Intermediate	7	It reminded me of Clip 2-A, where the first half is reminiscent of Bach's style but diverges from it in the second half with the chord changes per quarter note.
Intermediate	7	Sound similar to the reference clip minus the slight pauses in between
Expert	1	Bach's melodies always lead to somewhere. This goes nowhere at all.
Expert	2	The rhythm is closer to Bach but the harmony is quite discordant and has no direction
Expert	3	The melodies and harmonies made sound a bit like Bach, but they're not organised very well and the textures are very inconsistent.
Expert	4	Tiny moments of similarity
Expert	6	

Table C.8 Clip 4-B Results (Performance RNN)

Group	Grade	Responses
Novice	5	It feels a bit off from the snippet, but the feel is there.
Novice	5	i think it is kind of similar to what bach composed
Novice	5	It could work. But that first note kind of set the mood so now you can't unhear it as a less happy song.
Novice	7	This clip felt like the low-pitched chords were heading towards a certain direction as they harmonize with the high-pitched progression.
Novice	8	it starts with a few notes and then builds on it

Novice	8	It sounds like another song from Bach based on his the original clip.
Novice	9	Clip 4-b sounds more like Bach composition because of the same beats per minute, progression, and similar notes.
Novice	9	much much "smoother", harmony is still a bit strange, but i can see the vision
Novice	9	Sounds like it could be from the same song or an initial draft of it.
Novice	10	it sounds like classical music
Intermediate	5	theres more of an exchange of melodies in comparison to the previous recordings
Intermediate	6	Sounds good. There is a counter melody
Intermediate	7	Reminds me of a Bach fugue I've heard before
Intermediate	7	Harmonic choices sound more consistent and so does the phrasing.
Intermediate	7	feels like it has similar tempo to the sample piece
Intermediate	7	It sounds too eerie compared to Bach's fugue that was asked to be listened by us.
Intermediate	8	Sounds deliberate, leads somewhere, pacing feels similar to Bach, so does to the melody.
Intermediate	8	The melody and chords work well together.
Intermediate	9	the flow and counterpoint of the voices reminds me of Bach and it kinda does sound like a fugue he would make
Intermediate	9	I can't explain it except that it sounds similar and has the same vibe as the reference clip
Expert	3	This one is closer as it avoids any egregious discordant harmony but the lack of clear melodic ideas and harmonic direction still clearly differentiates it from Bach
Expert	4	Texture is a bit Bachian, but the initial melody (especially the repeated notes) and the counterpoint is unconvincing.
Expert	7	This was solid
Expert	8	Very close. Right hand melody playing the chords while left hand does a walking bass, all close to the diatonic purpose of the chords—very Bach.
Expert	9	sounded musical

Table C.9 Clip 5-A Results (Performance RNN)

Group	Grade	Responses
-------	-------	-----------

Novice	2	More minor notes, from what I'm hearing? So it makes it sound more melancholic.
Novice	2	I sounds too soft compared to the original clip
Novice	3	It was only in the end of the clip where I saw similarity on the original clip.
Novice	5	it seems like something bach would compose
Novice	7	The first 3 seconds feel like a different thing but the rest of it feels correct
Novice	7	I feel like this is more closer to the original snippet in terms of sound.
Novice	7	Similar to clip 4-a, clip 5-a have stops in it but the end part of the clip sounds like a Bach composition.
Novice	7	The start had a random pause again that feels weird out of context but the second half sounds like it could be from the song.
Novice	9	From 0:10 onwards, it REALLY felt like both low-pitched and high-pitched chords and progression were harmonizing with each other. It felt as if both were heading towards a clear direction that is pleasing and reminiscent of the Bach music clip.
Novice	9	the slowness at the start seemed strange, very "smooth" throughout though and the second half felt on point
Intermediate	3	chord progression and melodies seem better
Intermediate	3	Phrasing and harmony sounded more contemporary than Bach
Intermediate	3	Sounds kinda slow and sad which i don't associate with bach
Intermediate	5	A little like Bach, though sounds more like Mozart.
Intermediate	7	Earlier harmonies sound a bit too modern, but towards the end it the phrasing was sounding closer to Bach
Intermediate	7	The second half sounded like a Bach composition, while the first half did not.
Intermediate	8	It sounded very human like.
Intermediate	8	gives off a similar mood
Intermediate	9	I couldn't hear that well so take it with a grain of salt but it did sound like the second voice followed the melody of the first voice and the melody and complementary counterpoint does sound like Bach to me
Intermediate	9	Sounds similar to Art of Fugue, follows the same pattern, notes are fast and light

Expert	1	Block harmonies with melody over them don't sound particularly Bachian. Melody could be baroque but the texture simply isn't.
Expert	2	Beautiful composition but it's more Mozart than Bach.
Expert	3	Again, it's getting closer but some of the incidental harmony caused by the sustain of some of the notes is still quite discordant.
Expert	3	Barely similar
Expert	8	sounded musical. had appropriate spacing between the left and right hand

Table C.10 Clip 5-B Results (Polyphony RNN)

Group	Grade	Responses
Novice	3	It sounds more like something from the Legends of Zelda games
Novice	3	it feels off from the original piece.
Novice	3	This felt very chaotic compared to the example clip.
Novice	4	it's doing the thing again where its pressing all the keys at once
Novice	4	This clip sounded too monotonous with no change on chord progression. The high and low-pitched chords sounded odd and monotone, with no clear direction as to where they're going.
Novice	6	The build up of the clip is similar to the original clip
Novice	6	this sounds more like something bach would compose
Novice	7	the harmony sounds off, but there are attempts at "smoothness"
Novice	8	Similar to the other clips, Clip 5-b have similar beats per minute, progression, and notes.
Novice	9	It sounds more upbeat, much like Bach's music
Intermediate	1	Some rhythmic choices were very distinct (i.e., use of 16th or 32 notes and dotted notes) and the harmony was not similar
Intermediate	1	Does not sound like Bach and once again sounds arbitrary like a cat walking on a keyboard. Too bright and random.
Intermediate	3	chord progression is kinda awkward
Intermediate	4	Phrasing too disjointed.
Intermediate	4	It did not sound very uniformed, unlike Bach's fugue which had melody consistent with the chord changes.

Intermediate	5	There is a counter melody but there are also some stops/longer notes
Intermediate	5	Bears similarities but the style seems different (like the timing of when each note comes in). Although, it can still be possible Bach composed it, since I'm not very familiar with the variety in his compositions.
Intermediate	6	The rhythm sounds random
Intermediate	6	gives off a similar mood
Intermediate	6	The melodic flow dosent really resemble Bach in my mind but the counterpoint kinda reminds me of Bach and I dont really know why because it sounds like it shouldnt but it does?????
Expert	1	Generic drabble. No personality to this one.
Expert	1	The rhythm of the melodic ideas are very clunky, and the same can be said of the harmony
Expert	2	Just doesn't sound like Bach. More similar to video game music.
Expert	4	It's a no from me Dawg
Expert	4	