DeepBach: a Steerable Model for Bach Chorales Generation

Corey Ford
Eleanor Row
JingJing Tang

Background

Deep learning (dependency network)

Music generation + Bach

Problems with previous Bach models [3]

Deep bach aims to...

- **Enforce** unary user-defined constraints (rhythm, notes, parts, chords)
- **Generate** music that is identifiable as ‘Bachlike’ by people from various musical backgrounds (They did a comparison test)
- **Accessible** (interacting with DeepBach using a notational software such as Musescore)
- **Increase** speed of generation (parallel gibs)

Research Aims + Rationale

The DeepBach [3] paper was limited in its evaluation, only using three categories to distinguish between participants expertise.

We aimed to extend this, exploring more deeply:

- The relationship between musical expertise and people's ability to guess the ML models.
- Analyse the musical qualities of the model outputs, discovering their weaknesses.

Ethics

Submitted 27th October; Approved on the 28th October!

Consent forms based on QMUL's templates

Hearing damage warnings

Participants over 18 and in-house (C4DM mailing list)

Completely anonymous

Aimed for 15 participants (~10% of C4DM members).

No incomplete responses logged; stop at any time

REF NO. QMERC20.015

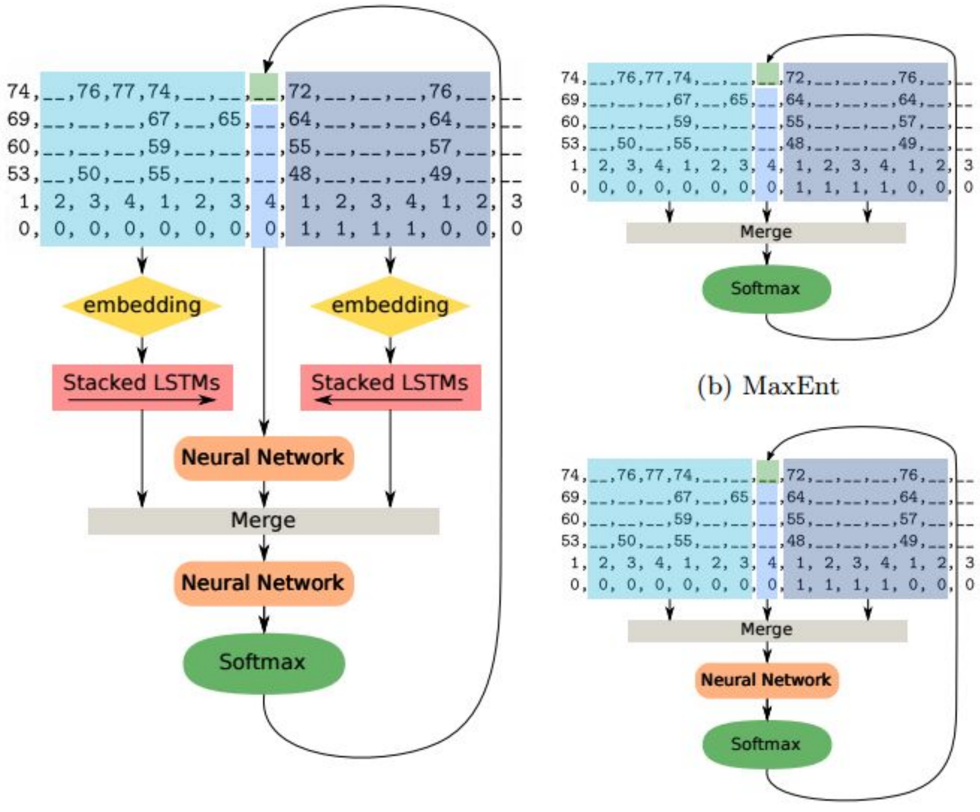
Study Design: Data and Model Architectures

Data Preparation

- Four voice lists
    - represented by pitch numbers
- Two metadata lists
    - subdivision list
    - fermata list

Model Architectures

- DeepBach
- MaxEnt (Maximum Entropy)
- MLP(Multilayer Perceptron)



(a) DeepBach (b) MaxEnt (c) MLP

[2]

Study Design: Generation

Pseudo-Gibbs Sampling Procedure

- non-reversible MCMC algorithms

Flexibility of The Sampling Process

- **fix the soprano** and sample voice 2, 3, 4 in step (3)
  - allow MIDI-input of the soprano
- **fix fermata list** to impose the end of music phrases
- **fix specific subsets of notes** within the range of voice *i* to restrict specific chorales

---

Algorithm 1 Pseudo-Gibbs sampling

1: **Input:** Chorale length L, metadata $\mathcal{M}$ containing lists of length L, probability distributions (p_1, p_2, p_3, p_4), maximum number of iterations M
2: Create four lists $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4)$ of length L
3: {The lists are initialized with random notes drawn from the ranges of the corresponding voices (sampled uniformly or from the marginal distributions of the notes)}
4: **for** m from 1 to M **do**
5:     Choose voice i uniformly between 1 and 4
6:     Choose time t uniformly between 1 and L
7:     Re-sample $\mathcal{V}_i^t$ from $p_i(\mathcal{V}_i^t | \mathcal{V}_{\setminus i,t}, \mathcal{M}, \theta_i)$
8: **end for**
9: **Output:** $\mathcal{V} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4)$

Study Design: Implementation Details

- **Train** the DeepBach, MaxEnt, MLP using Keras with the Tensorflow backend (CPU version).
- Use the **dataset** of chorale harmonizations by J.S. Bach included in music21 toolkit (obtaining 352 pieces)
- **Augment** the dataset by adding all chorale transpositions which fit within the vocal ranges (obtaining 2503 chorales, 80% for training, and 20% for testing)
- Model only **local interactions** between a note and its context (only elements with time index *t* betweent (*t-16, t+16)*).

Study Design: Implementation Details

- For each model architecture, we obtained four models for each voice list and integrated them for generation.
- We selected **four Bach pieces of four-bar length** from two chorales, and wrote the **soprano part** of those pieces as MIDI files to input into the models.
- For each Bach piece, we had the corresponding generation from each of the three model architectures using the fixed soprano of that piece.
- We obtained **twelve generations;** only nine generations were used for the human listening test.

MLP 153 p2	
Bach 153 p2	
MLP 29 p1	
DeepBach 153 p2	
Bach 153 p1	
ME 29 p1	
DeepBach 29 p1	
ME 153 p1	
DeepBach 153 p1	
MLP 153 p1	
Bach 29 p1	
ME 153 p2	

Study Design: Questionnaire

Each generation was synthesised using the Leeds Town Hall Organ soundfont (like in DeepBach [3]).

Presented next to the binary choice “Bach” or “Computer”.

Also, 3 background measures were collected:

- Choices used in the original paper: I seldom listen to classical music / music lover or musician / Student in music composition or professional musician.
- Goldsmiths Musical Sophistication Index Measure [4] (Short)
- Measure of Bach Familiarity: “I am familiar with bach chorales”

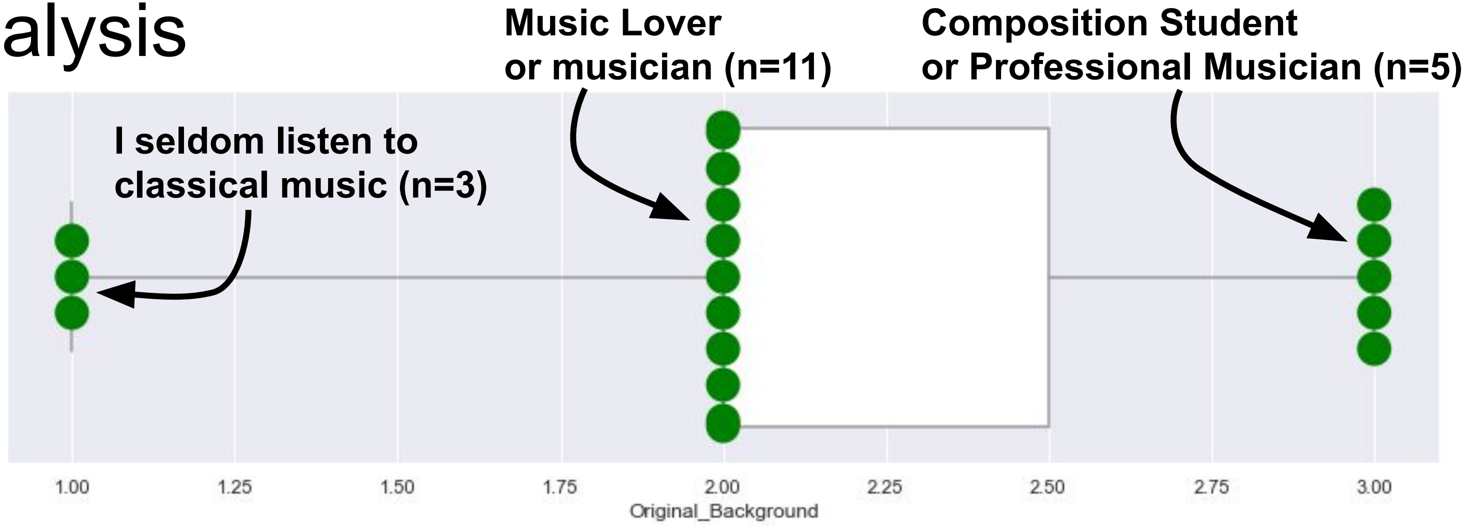
Results + Analysis

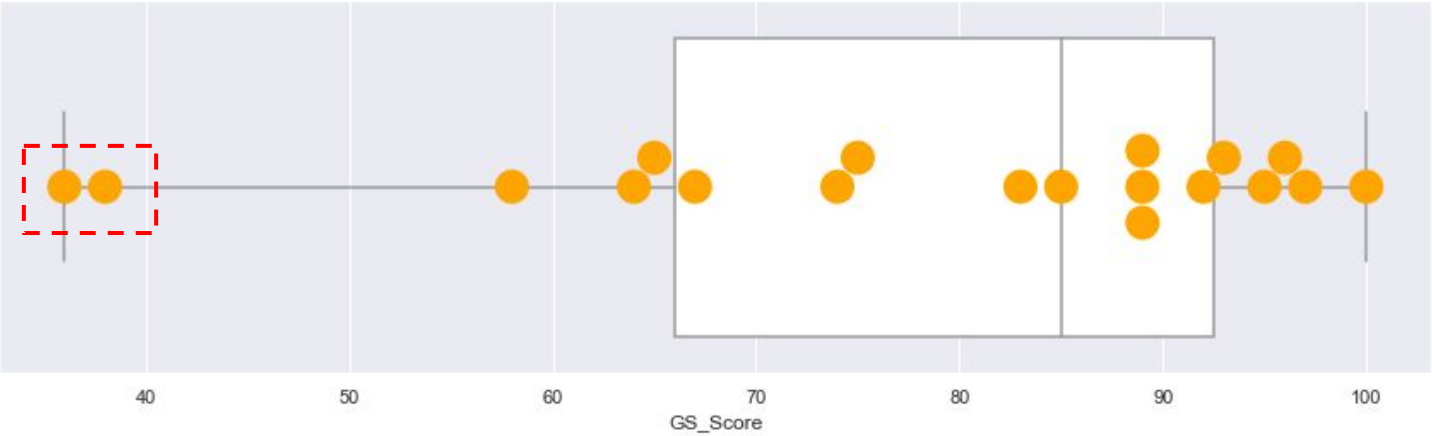
N=21 -> N=19

Groupings are similar to in the DeepBach paper.

Goldsmiths MSI Score between 36 and 100 *(M=78.16, SD=19.135)*.

Most participants scored > 60.

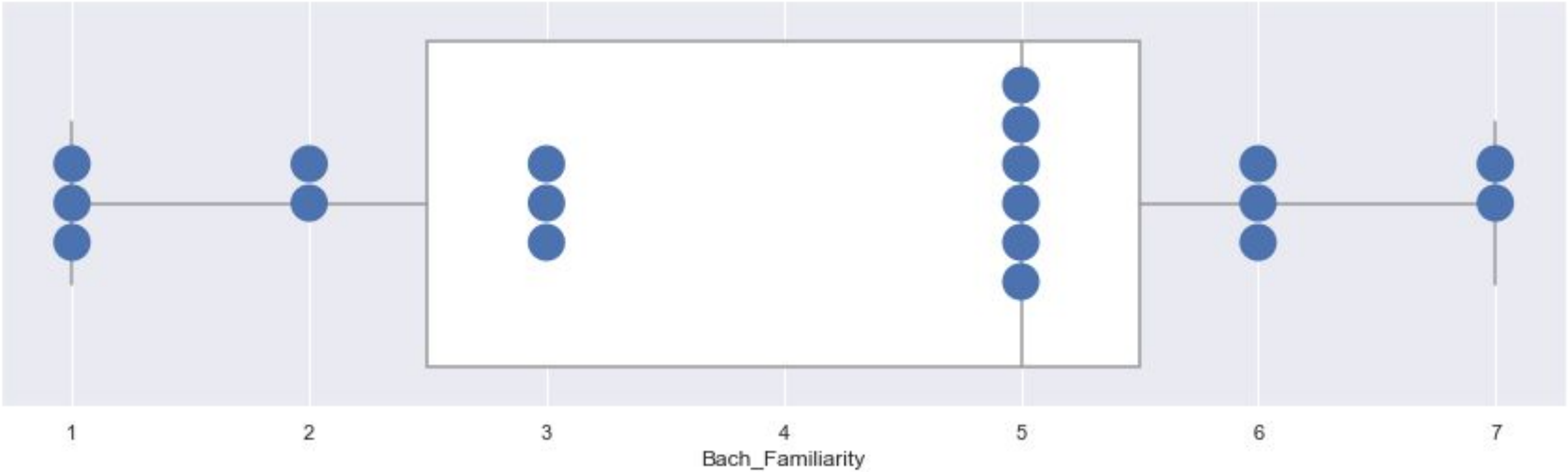




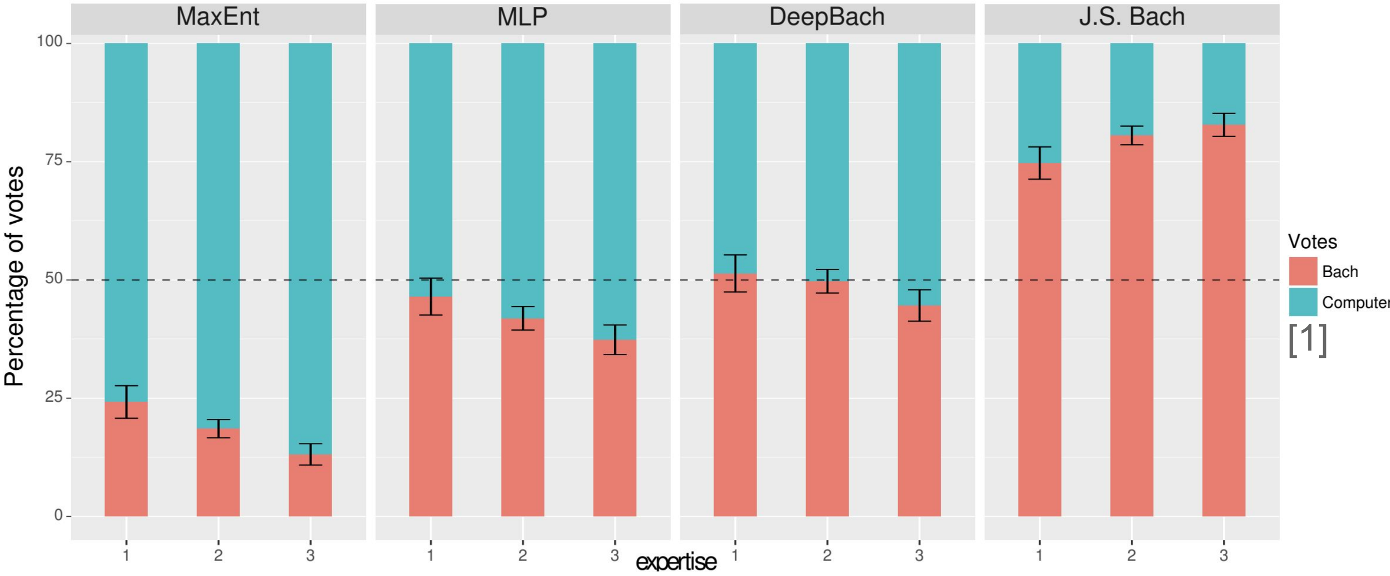
Results + Analysis

We also took a 7-point Likert-scale measure: “I am familiar with Bach Chorales”.

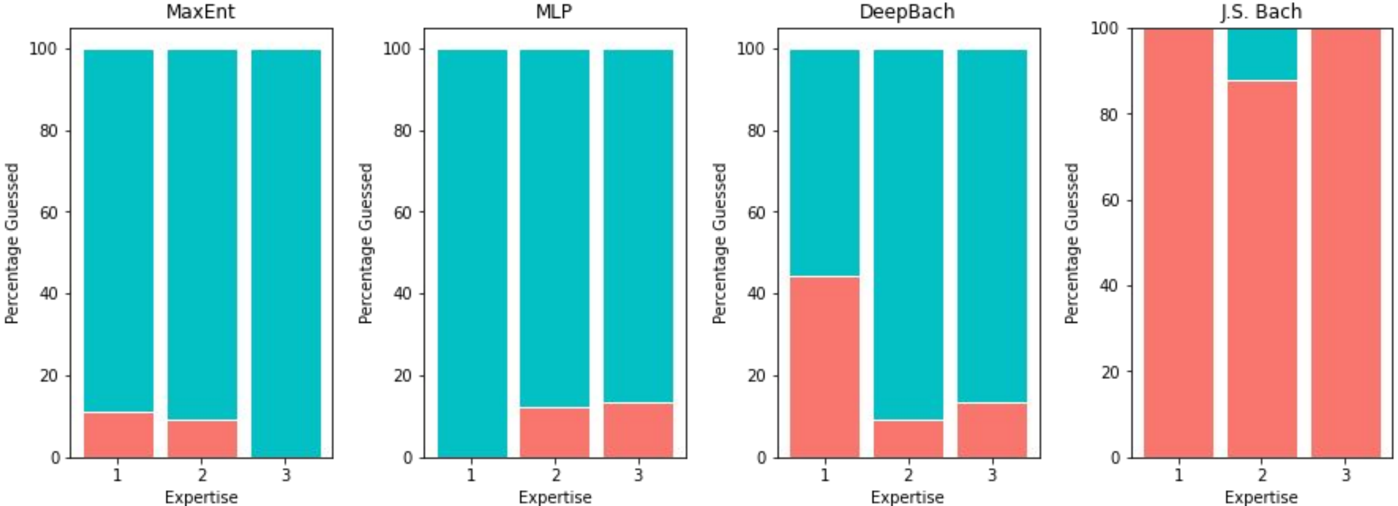
There were a range of responses *(M=4.11, SD=2.025)*. The data is split (those who agree and those who don’t agree). Data is not normal (top heavy, 5 mostly selected).



Original paper showed DeepBachs success and made links between expertise; we see this...

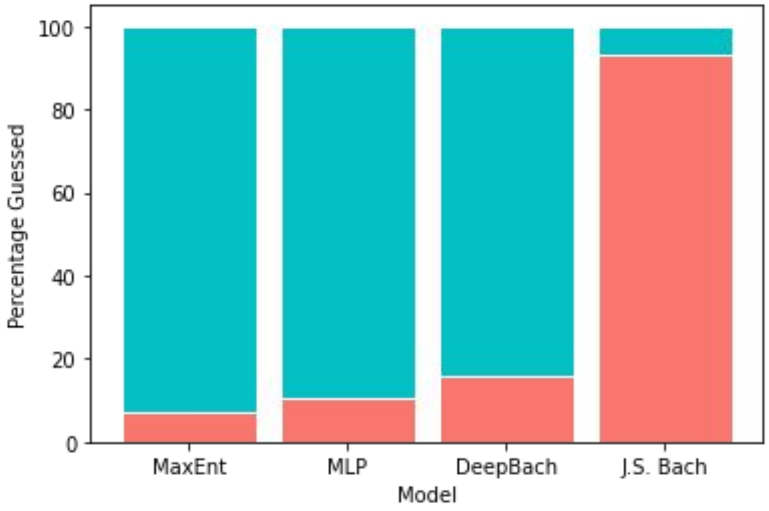


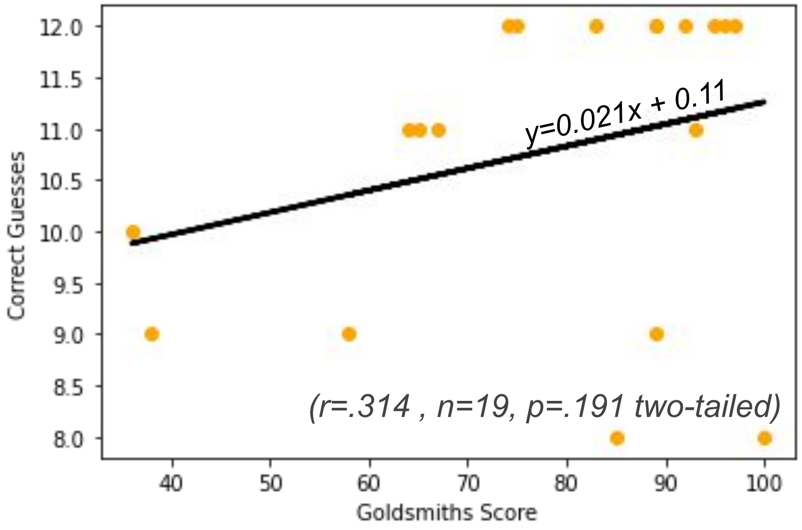
[1]

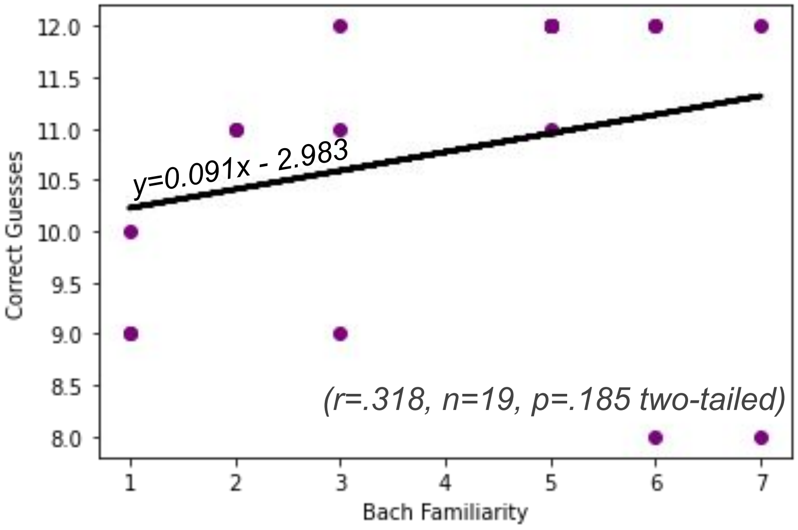


H_0: Distribution of correct guesses is the same across each model.

An Independent-Samples Kruskal-Wallis test **retains** the null hypothesis *(Kruskal-Wallis $H=.653$, $df = 3$, $p = .884$.);* do **not** reject H_0.





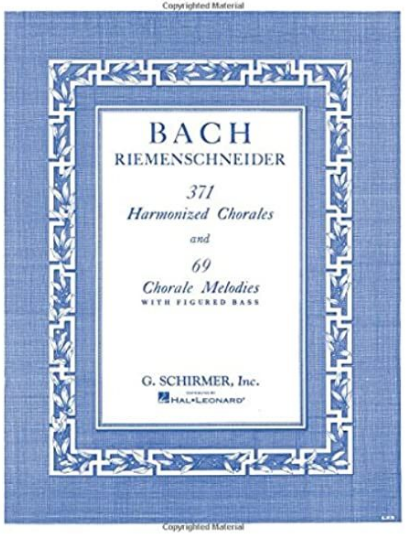


No significant correlation found between the total number of correct guesses and the total goldsmiths score/bach familiarity (calculated using Spearman's Rho).

However, within the Goldsmiths MSI:

- Significant (rank) correlation between **correct guesses** and **no. of years spent practicing an instrument daily** *(r=.617 , n=19 , p<0.005 two-tailed)*.
- Significant (rank) correlation between **correct guesses** and **people who consider themselves musicians** *(r=.502 , n=19 , p=0.029 two-tailed)*.

Basic Intro to Bach Chorales



Things to look out for in good harmonisations:

- Each part has strong **sense of line**
- **Contrary motion** between soprano and bass line
- **Passing notes** in ascending or descending step-wise motion
- **Suspensions**, which are resolved

Avoid:

- **Parallel 5ths or 8ves** (between every part!)
- Omitting **3rds**
- **Clashes** with dissonant note and resolved note

Keys and Cadences

- Starts off in a tonic key - modulates to related keys throughout
- Normally view cadences as three chord progression but for simplicity:

V - I **V^7 - I**	Perfect
I - V **IVb - V** **iib - V**	Imperfect

Interrupted cadences are also a thing!

[1]

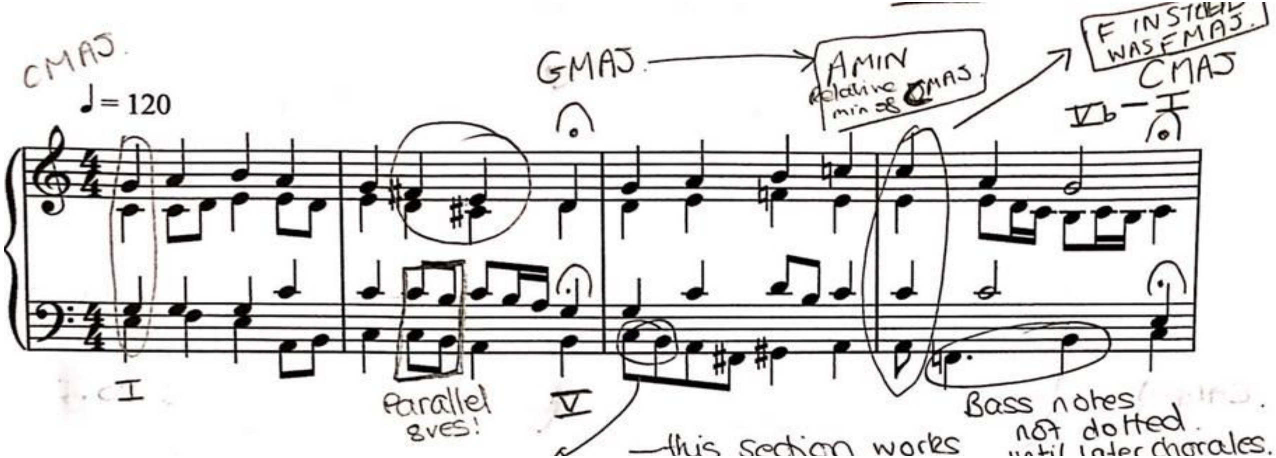
Chorale No.29 Soprano line:

Original:

- G Major

Generated output:

- Assumed C Major



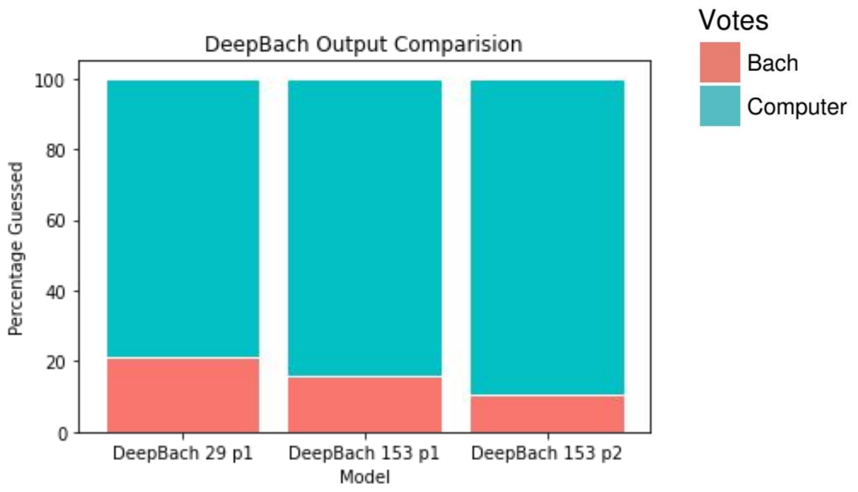
Analysis: What worked?

This generation demonstrated general sense of cadential harmony than others:

- Use of Perfect Cadence
- Use of relative keys to assumed tonic (including relative minor and dominant)

Better than other generations

Possibly why some thought it was Bach



Discussion: Still a long way to go

Generated outputs demonstrate no sense of musical line in each part

Lots of Parallel 5ths and 8ves

Soprano and Bass not complementary - little contrary motion if any

Model not able to properly distinguish appropriate tonic based on soprano input alone

Model not able to implement modulations within passages between each femata

Discussion/Conclusion

- The participants had reasonably high goldsmiths MSI scores. Perhaps, we would have found a bigger difference between models if we captured a more general set of people.
- Sample size is small also… perhaps Prolific or MTurk?
- Considering the correlations between music instrument skill and correct guesses. Do they have better listening skills? Do the recognise dissonance better?
  - “It was not as trivial as I thought to distinguish from both”
  - “For me, computer generated music sounded more harmonic.”
  - “I was expecting to hear more acoustic sounds when listening to Bach, organs, voices or something like that.”

Discussion/Conclusion

Harmony is really important! Could this be quantified?

- “the giveaway for some of the computer extracts was (I think) the use of non-chord tones”
- “...it might be better to compare with sections where Bach is modulating to a new key or similar.
- “a single note of deviation from this without harmonic justification would immediately alert me that a piece is not a Bach chorale”

Longer generations? Only model local interactions (one bar before and one after the note)

Thanks for listening!

References

[1] Alan Charlton. *Worksheet 3 - cadences - solutions.* 2015. isbn: 9780793525744. url: https://www.rhinegold.co.uk/wp-content/uploads/2015/10/MT10_scheme_KS5_Bach_chorale_harmonisation.pdf.

[2] Sam Gorree. *"DeepBach"*. Blog Post. 2017. url: https://samgoree.github.io/2017/01/29/DeepBach.html

[3] Gaëtan Hadjeres, François Pachet, and Frank Nielsen. "DeepBach: A steerable model for bach chorales generation". In: *34th International Conference on Machine Learning, ICML 2017*. Vol. 3. 2017, pp. 2187–2196. isbn: 9781510855144. arXiv: 1612.01010.

[4] Daniel Müllensiefen et al. "The musicality of non-musicians: An index for assessing musical sophistication in the general population". In: *PLoS ONE* 9.2 (Feb. 2014). Ed. by Joel Snyder, e89642. url: https://dx. plos.org/10.1371/journal.pone.0089642.

Find the dataset here: https://github.com/thecoreyford !

All researchers are part of the "UKRI Centre for Doctoral Training in Artificial Intelligence and Music". Corey Ford and Eleanor Row are funded by the UKRI and EPSRC, under grant EP/S022694/1. Jinging Tang is funded by the CSC (Chinese Scholarship Council).