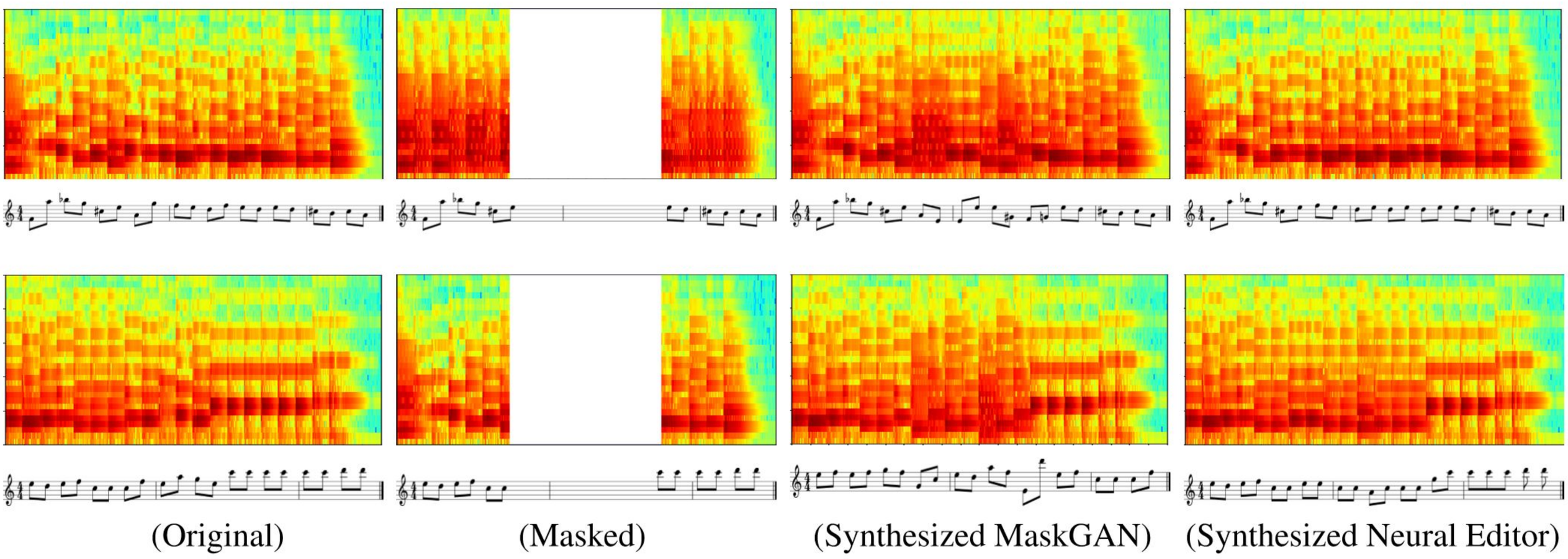
Explainable Musical Phrase Completion

Gregory W. Johnsen[1], Ling Lin[1], Lucia Yu[2], Andrew Dempsey[2], Vishwali Mhasawade[2], Daniel Jaroslawicz[1]
Iddo Drori[1,2]

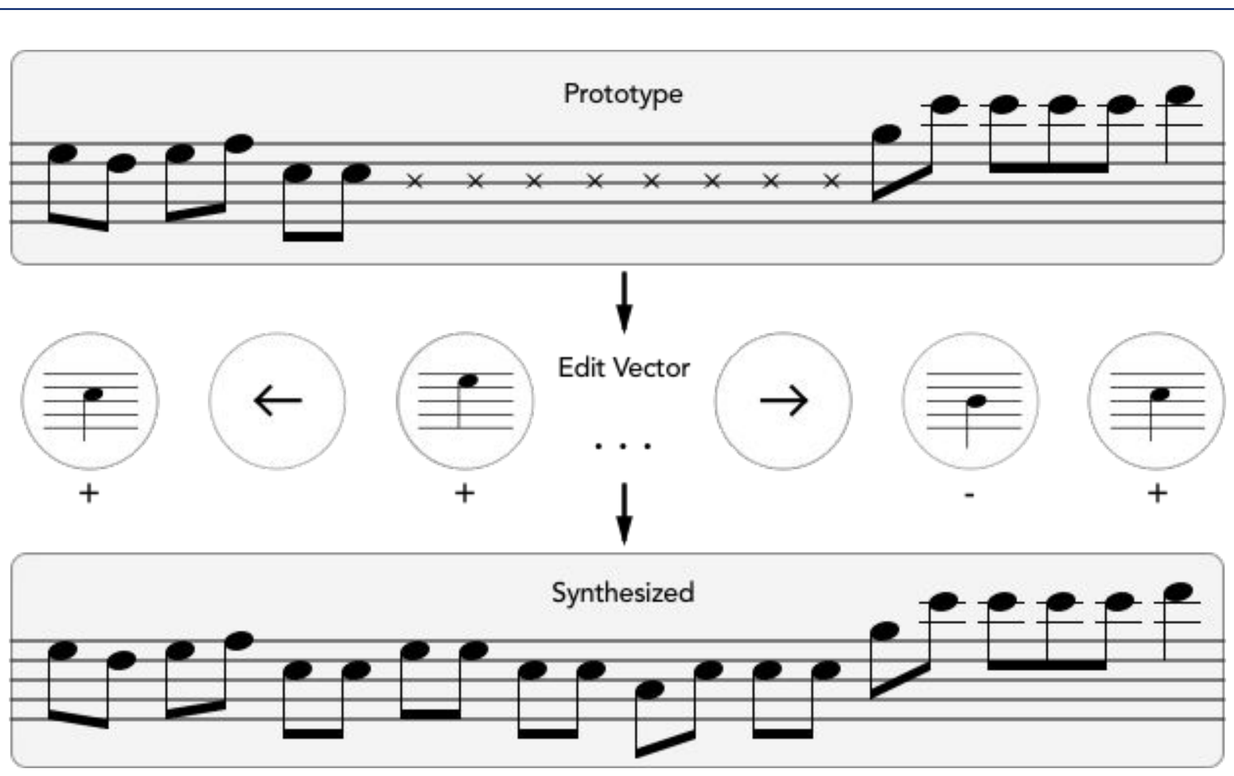
COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

[1]Columbia University, [2]New York University

NYU | TANDON SCHOOL OF ENGINEERING



Writers and composers do not create prose and music in a purely sequential manner. Composition is an iterative process that includes applying edit operations and shifting attention to and from different parts of the piece. We use a Neural Editor for completion and synthesis of musical phrases by editing prototypes. An advantage of this approach is that it is explainable by design: it maintains the provenance of all edits performed until reaching the synthesized output, which allows us to fully visualize the editing process.



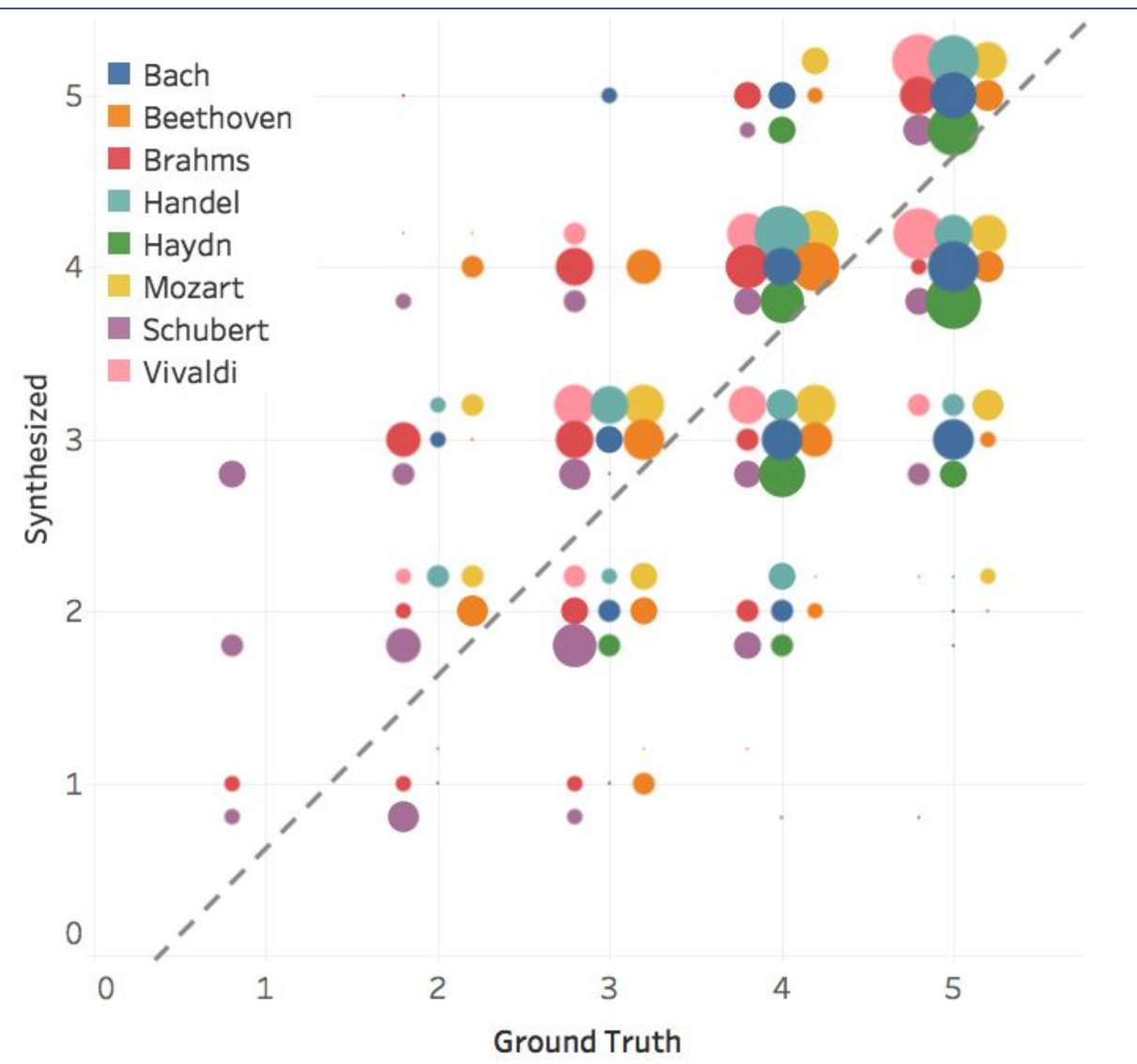
Prototype and edit approach to music completion.

Dataset:

- Collection of 3,428 classical music compositions, 8 composers.
- Tokenized main instrument, monophonic musical phrases.
- Vector representations for discrete musical note tokens.
- Split dataset into 95% training and 5% test sets.

Neural editor (Guu et al, TACL 2018).

- Bi-directional LSTM model
- Edit vectors apply various operations to musical notes:
  {insert, delete, replace, move left, move right}
- Model trained by maximizing the marginal likelihood.



Human evaluation of Neural Editor synthesized music

- Scatter plot of Likert scale of original music vs. the synthesized
- Each composer is colored and positioned in one of 9 positions.
- Circle size represents number of votes for each combination.
- Marks above diagonal: synthesized as good or better than original.

	Bach	Beethoven	Brahms	Handel	Haydn	Mozart	Schubert	Vivaldi
Original	4.3±0.9	3.5±1.0	3.4±1.0	4.0±0.9	4.4±0.7	3.9±1.0	3.2±1.3	4.2±0.9
Neural Editor	3.7±0.9	3.3±1.1	3.6±1.1	3.7±1.0	3.9±0.9	3.6±1.0	2.8±1.2	3.9±0.9
MaskGAN (ICLR 18)	3.4±1.1	3.0±1.1	3.1±1.1	3.0±0.9	2.9±1.2	3.1±1.0	2.8±1.2	3.6±0.9

Human evaluation of musical samples on a 1-5 Likert scale (mean ± standard deviation)

Source code: github.com/gjohnsen/musical-phrase-completion