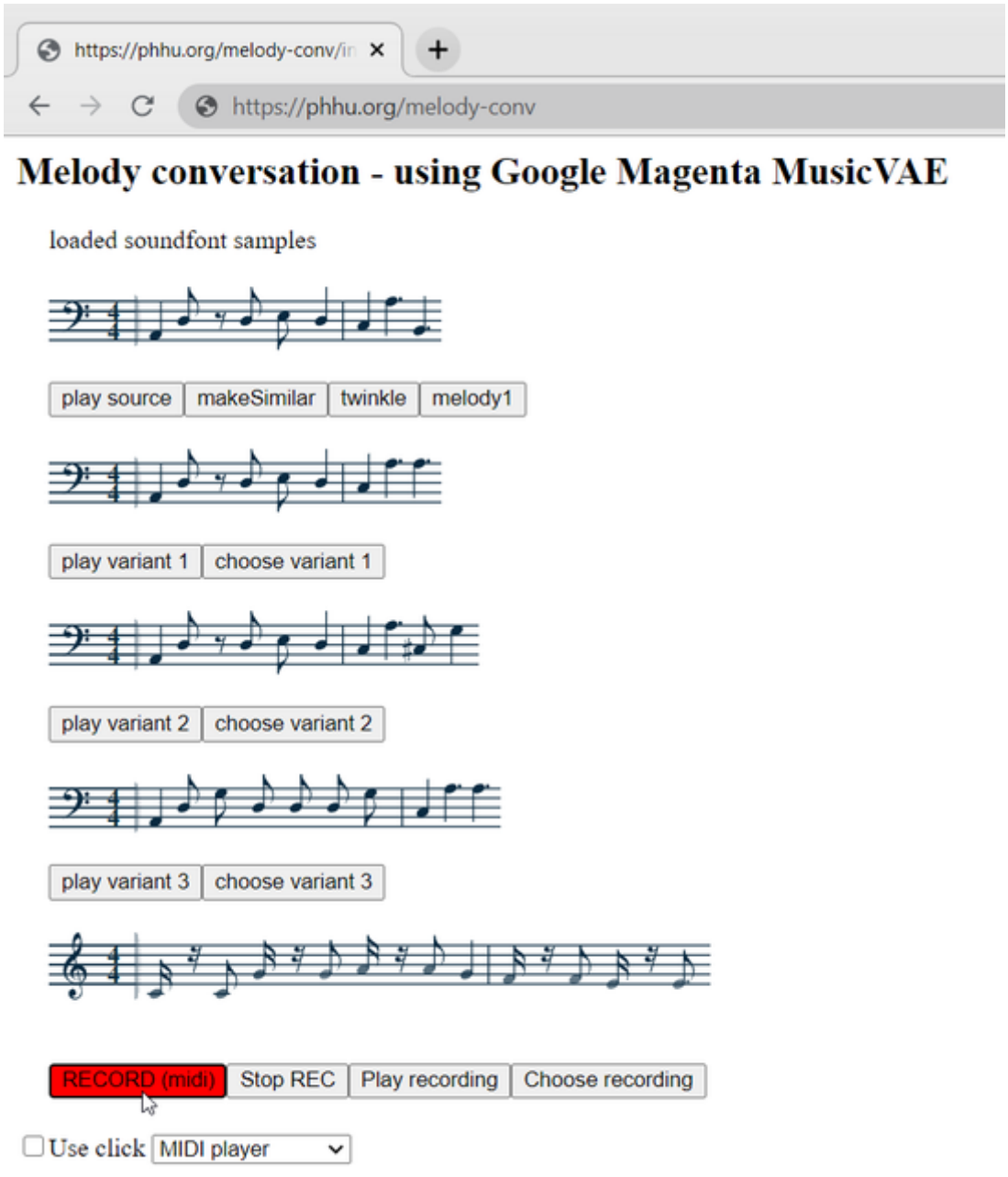
NYUx Theories of Media and Technology - Philip Hughson - philipfhughson@gmail.com - January 2021

Creation 3 - Conversational Cyborg Melody creation using Google Magenta MusicVAE

- My project is online at https://phhu.org/melody-conv.
- The code is available at https://github.com/phhu/melody-conv

Please note that this project is still a proof of concept! In particular the recording function will only work if you have a MIDI instrument connected to your computer: I did not have time to find and wire up a suitable QWERTY keyboard based MIDI controller. If you don't have one, please believe me that it works! You can still use the AI-powered melody variation functionality repeatedly to develop melodies this way.



Articulation

In this project overall, I have looked at the idea that our (creative) imaginations can considered as cyborgian processes just as more traditional human faculties such as strength, running speed, ability to sense and fly, etc. might be. I have been interested in the idea that by using technology, we can learn to see and express possibilities that might otherwise not occur to us. This idea contrasts with the idea that technology replaces our capacities, a common example of which might be our navigational capacities and instincts being replaced by satnavs (GPS / satellite navigation systems).

In my synthesis, I looked at two examples where I cyborgian imagination at work: Harry Yeff's beatbox music (for which see the references below); and the Google Magenta project (Google 2020), which is "an open source research project exploring the role of machine learning as a tool in the creative process.". I argued that both of these show an interesting interplay between the human and the artificial. Yeff's beatboxing is particularly interesting because it is a form of purely human vocal expression which doubtless would not have developed without the artifice of electronic music. Based on my reading of Paul Pangaro's exposition of Gordon Pask's Conversation Theory, I argued that conversation (metaphorically or not) is a very interesting idea for thinking about our relation with artificial intelligence: good conversation demands that both parties have some understanding of what the other person is talking about, and can understand their purposes to some extent. and that indeed such conversation is what is going on in some of Yeff's work, especially the "Second Self) performance (Yeff 2019). And furthermore such conversation is implied, if in a relatively undeveloped way, in the "AI Duet" project (see https://experiments.withgoogle.com/ai/ai-duet/view/). Both of these essentially involve musical conversation with a trained recursive neural network (RNN).

Based on this, I decided to try to use Google Magenta to create a project which would develop a musical idea - a melody - in a way that involves some sort of conversation between a human and an artificial neural network. I noticed that the Music-VAE project (see https://magenta.tensorflow.org/music-vae) might work for this. It uses "a hierarchical recurrent variational autoencoder for learning latent spaces for musical scores". The demos for the project mostly show interpolation between melodies and rhythms, and they often don't involve human interaction on the musical level.... but it is also possible with MusicVAE to ask the AI for melodies similar to a given melody, which is the central idea used here.

In my project, I have the AI offer three variations, all reasonably close to the original. The user can then play these melodies, and choose one of them, which is then set as the starting melody, creating an iterative cycle which allows for the development of new variations on a melody. (So far, this is somewhat similar to the predictive text poetry example I have looked at in my research).

Where is the conversation though? The user can also record themselves playing the melodies using a MIDI (musical instrument digital interface) instrument, such as the electric piano I have. If what you record matches one of the variations, it will be highlighted in green (based on a crude measure of similarity: the melody should have the same notes in the same order, irrespective of rhythm). Based on this, the user has to show the ability to express the same ideas that the AI comes up with, potentially before being allowed to move on to the next iteration (though this is

not enforced at present). The user can also override the initial melody with a recording, potentially introducing new variants.

This gives a situation where the eventual melodies produces are the product of the both the AI (in variations) and the user (in choice of variation, and also generation of new variations). Thus the melodies could reasonably be said to be cyborg melodies!

Furthermore, at possibly more interestingly, the idea that the user can learn to play - imitate - the melodies generated by the computer - gives an idea of AI being used to enhance human capacities, rather than remove them. The rate of variation can in principle be changed automatically, meaning that the AI could become a good teacher by pushing the user's level as they become more accurate. Perhaps if the interface were a little more automated, it would be getting closer to the ideas expressed by Paul Pangaro in his presentation "Less interference/More Dance" (2019).

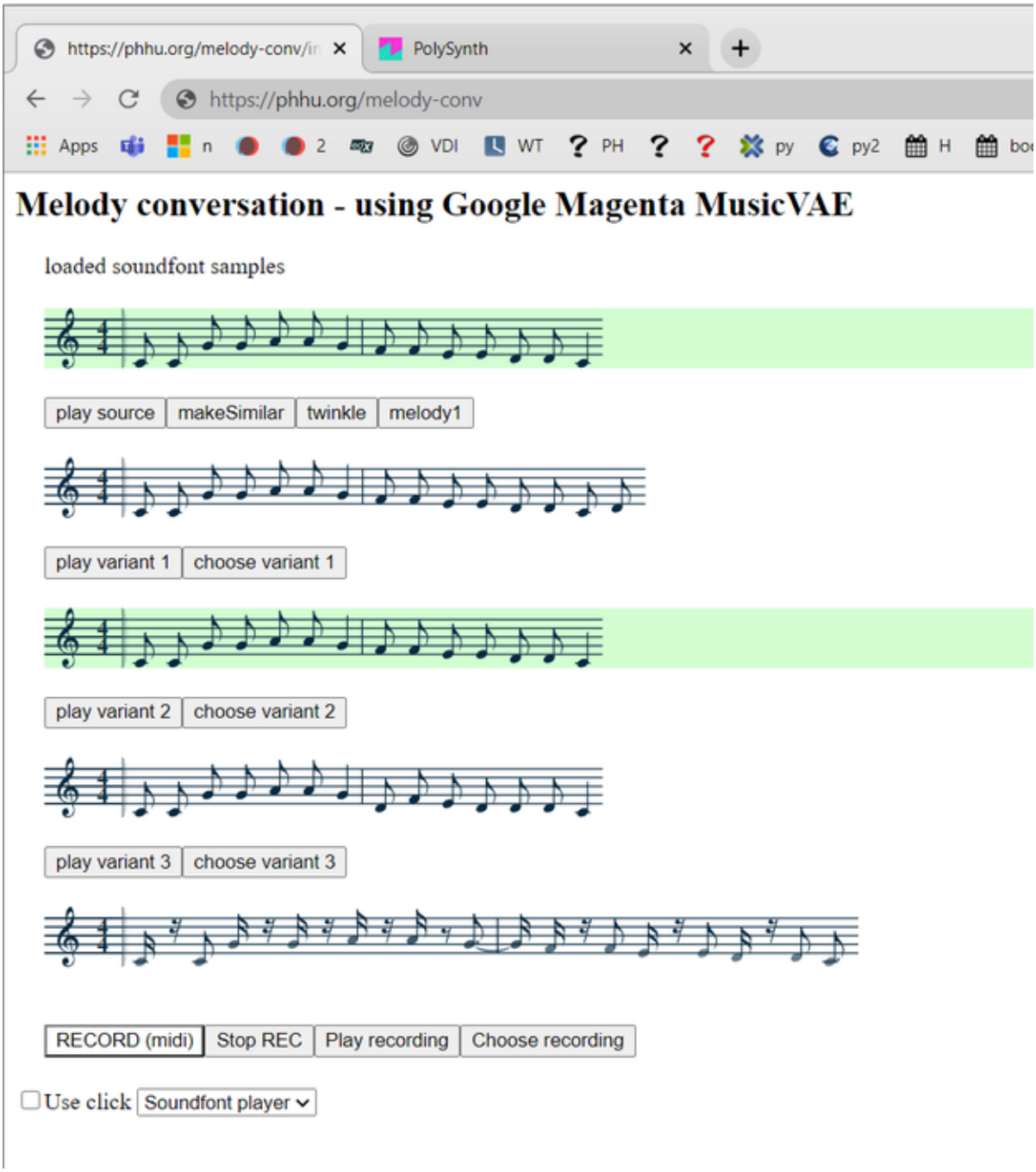
Another interesting idea here is the notion that AI can be used to generate meaningful (rather than arbitrary) variety. It is notable that the quality of the melodies produced by the AI, though sometimes erratic, is generally pretty good. If one of the negatives of mechanisation have been the standardisation of products (such as archtectural forms, or recording music), there is an idea here that a well trained AI could produce convincing variations that might make our post-industrialised world more varied place! (An interesting comparison is with the notion that 19th builders and architects made buildings with natural variations along common themes - a idea perhaps of conversation within a common language - which makes 19th century cities often pleasant to look at... time doesn't allow me to develop this idea further alas, but perhaps well conversed cyborg architects could do the same!

Usage notes

The main screen shows four or five musical staffs. The first represents a starting two bar melody, either "Twinkle Twinkle little Star" or "Melody 1". The next three represent variations on this initial melody. These variations are generated by the AI, which is pre-trained on common melodic patterns. If you click "choose variant x" for one of these, it will be selected as the starting melody, and new variants will be generated.

There are three sound players available: basic player, which is ugly; Soundfont player, which is a good piano sound; and MIDI player, which will use a connected MIDI instrument. The soundfont sounds take a while to load, so the basic player is used by default.

For recording, you need to connect a MIDI instrument. You can use https://tonejs.github.io/examples/polySynth to test these if you have one. Then you can record yourself, using an optional click to keep time. The recording will be automatically trimmed and quantized. If your recording matches one of the variations, that variation will show up in green, as shown on the screenprint below:



REFERENCES

Google (2020) Magenta (website) Available at https://magenta.tensorflow.org/.

Paul PANGARO (2019) "Less interference/More Dance". Available at https://pangaro.com/lasg2019/index.html and https://www.youtube.com/watch?v=GfP6CiAcDLs

Harry YEFF (2013) “Move” (beatbox performance) Available at https://www.youtube.com/watch?v=YH5ty3Kucz4

Harry YEFF (2019) “Reeps One ft. A.I. ‘Second Self’”. Available at https://www.youtube.com/watch?v=q981cTdL0_Y

Harry YEFF (2020) “How Battling A.I. Unlocked The Power of My Voice”. Available at https://www.youtube.com/watch?v=wTMMopLYJn4