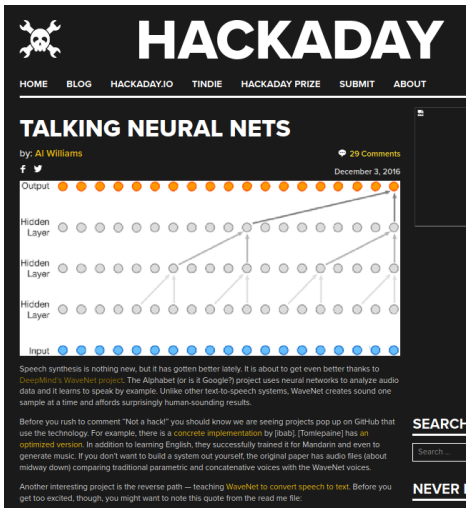
Deep learning for audio and music
Lecture 1

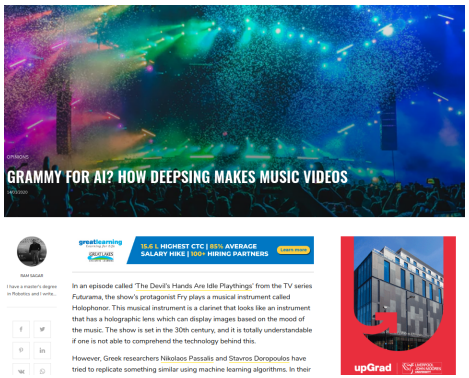
Callum Goddard

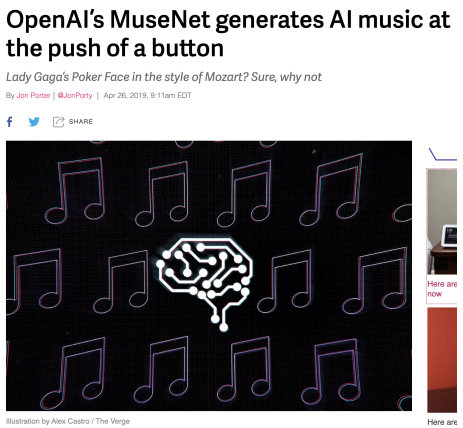
School of Elec Eng & Computer Science
Queen Mary University of London

(Thanks to Dan Stowell and Huy Phan)

Deep learning for audio and music







- https://deepmind.com/blog/article/wavenet-generative-model-raw-audio
- https://deepsing.com/
- https://openai.com/blog/musenet/
- https://magenta.tensorflow.org/

Introductions

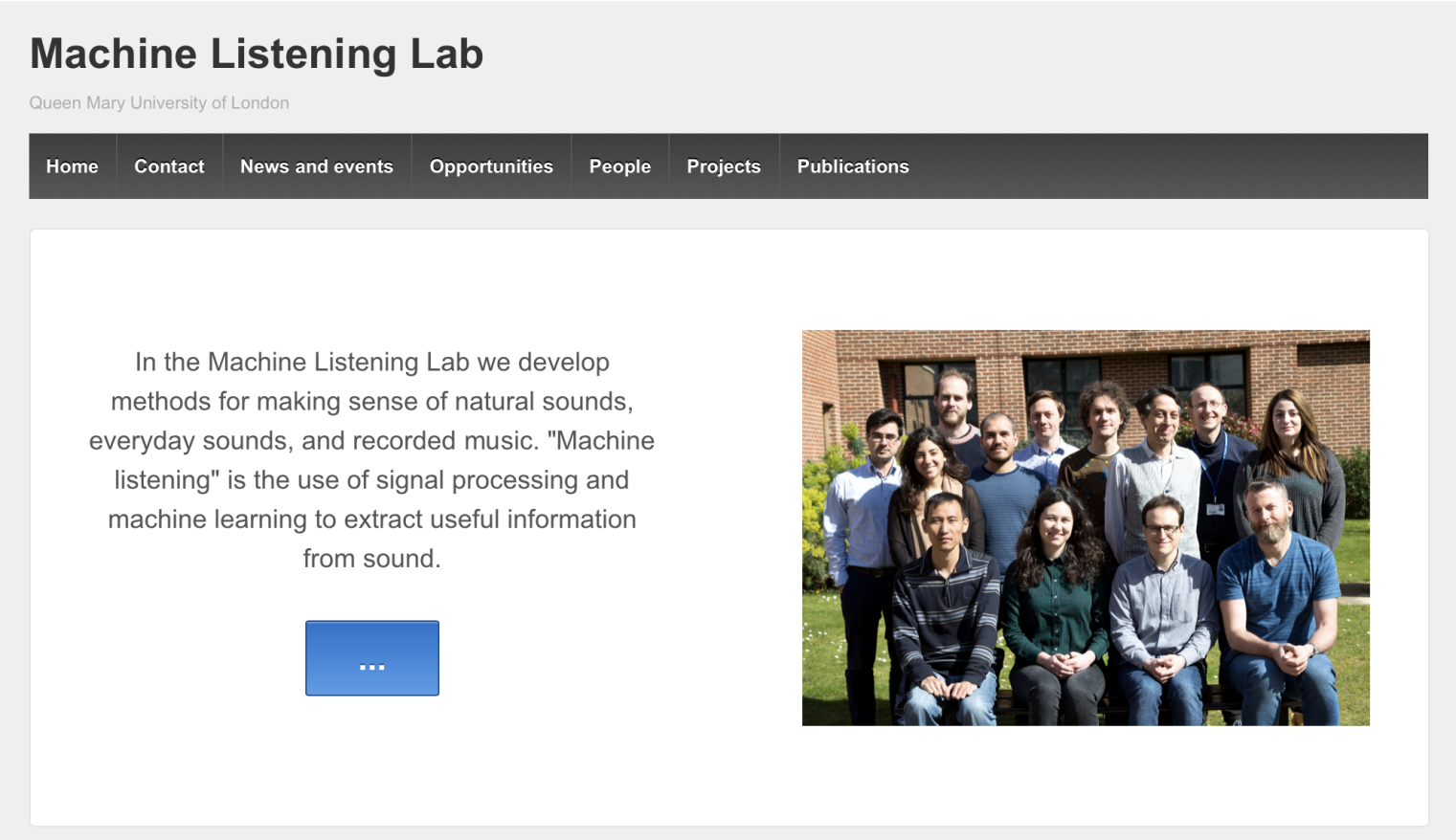
Done deep learning?

Introductions

What shall we do with deep learning?

Machine Listening Lab

Machine Listening Lab



- http://machine-listening.eecs.qmul.ac.uk

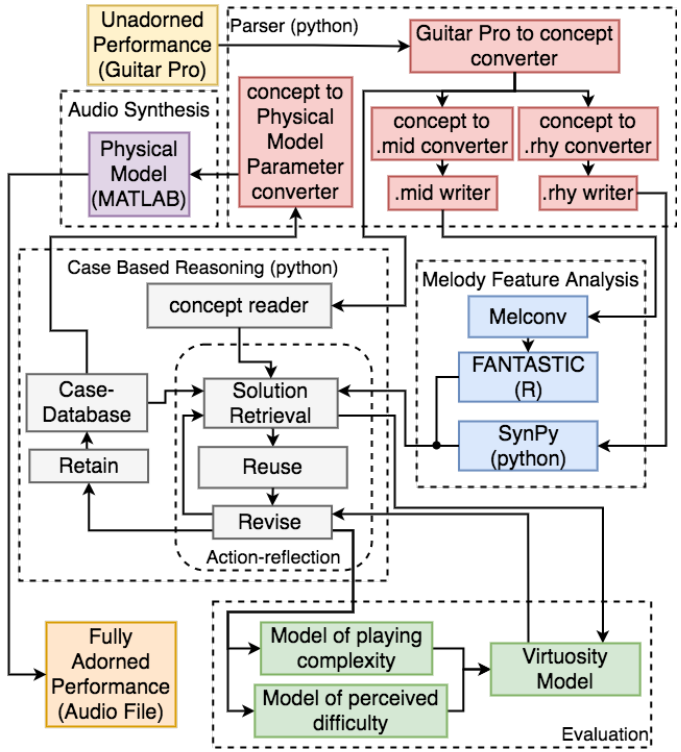
About me

About me

- Multi-interdisciplinary background in audio engineering, electronics, music performance and artificial intelligence
- MEng in electronic engineering and music technology systems
- Joined QMUL on the Media and Arts Technology PhD and was awarded my PhD in 2022

About me: PhD research

Computational Creativity, Music Performance Systems, Music Perception, Symbolic MIR



About me: recent work

Research Assistant working on music transcription tasks:

- Algorivm:
  A new startup developing automatic guitar performance assessment methods based on score informed transcription approaches (2021)
- Huawei: Multi-Instrument Music Transcription
  Tested and benchmarked state of the art multi-instrument transcription systems.
  Created new datasets (2022)

Teaching Fellow:

- Deep Learning for Audio and Music (2020)
- Procedural Programming (2019)

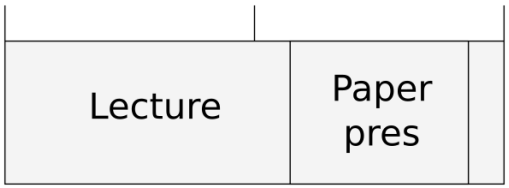
Schedule

- Lectures: Friday, 9am–11am, G.O.Jones UG1
- Labs: Friday, 4pm–6pm, ITL - Ground Floor

- Instructor: Callum Goddard, office hours: Fridays by Appointment/email
- Teaching Assistants: Shubhr Singh
  Yixiao Zhang

15 credits = 150 hours

Assessment



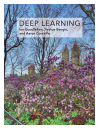
10%: Paper presentation in class (15 min)

10%: Mid-way revision test (45 min)

40%: Computer experiments with written report (1500 word)

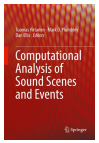
40%: Exam

Textbooks



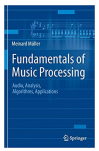
Goodfellow et al "Deep Learning"

deeplearningbook.org



"Computational Analysis of Sound Scenes and Events"

cassebook.github.io

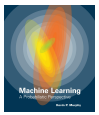


"Fundamentals of Music Processing"

https://link.springer.com/book/10.1007/978-3-319-21945-5

Zhang et al: "Dive into Deep Learning"

https://d2l.ai/



Murphy: "Machine Learning"

www.cs.ubc.ca/ murphyk/MLbook/

QM library: Hard copies AND ebooks

Textbooks

...plus blogs, tutorials, videos...

Online: QMplus, github, etc

QMplus:

https://qmplus.qmul.ac.uk/course/view.php?id=15474
(inc slides, Q&A, announcements, assessments)

Feel free to publish your code/slides online (e.g. blog, Github)

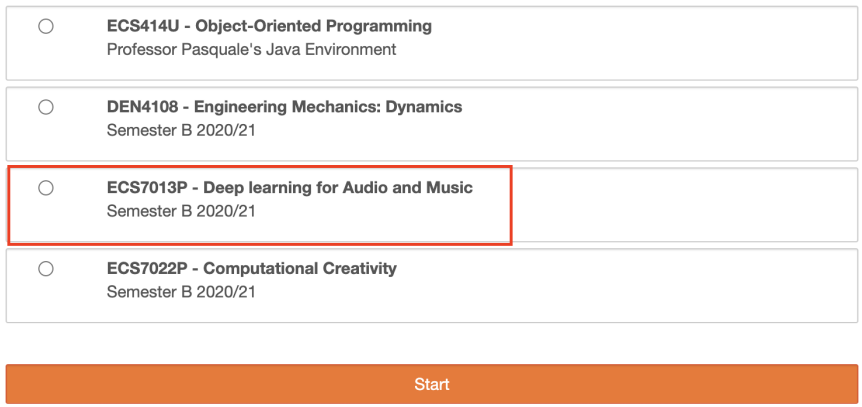
Student forum

Python, torch etc

Our programming framework:

- Python 3
- PyTorch
- Plus numpy, matplotlib, librosa

QMUL compute servers via `https://jhub.eecs.qmul.ac.uk/`
Login with your EECS credentials



Paper presentations

15 minute each, 10% of module mark

1. Read a research paper
2. Compose some simple slides on it
3. Present 10–12 minutes (plus questions)
4. (Also submit slides via QMplus **before the lecture time**)

Paper presentations

15 minute each, 10% of module mark

1. Read a research paper
2. Compose some simple slides on it
3. Present 10–12 minutes (plus questions)
4. (Also submit slides via QMplus **before the lecture time**)

Paper presentations

Suggested slide organisation:

1. Paper title, authors, and general **aim**
2. Method – what they did
3. Novelty – what was new/interesting
4. Limitations – in your opinion
5. Implications – what do we learn from this?

Approx 12 slides
You MAY reuse figures/tables etc from the paper

Paper presentations, "Season 1"

Fairbrass et al (2018)
: CityNet—Deep learning tools for urban ecoacoustic assessment

Davies and Boeck (2019)
: Temporal convolutional networks for musical audio beat tracking

Pandeya et al (2018)
: Domestic cat sound classification using learned features from deep n.n.s

Salamon et al (2017)
: Fusing shallow and deep learning for bioacoustic bird species classification

Sethi et al (2019)
: Combining machine learning and a universal acoustic feature-set yields efficient automated monitoring of ecosystems

Kim et al (2018)
: CREPE: A Convolutional Representation for Pitch Estimation

Cakir et al (2017)
: Convolutional recurrent n.n.s for polyphonic sound event detection

Dieleman et al (2014)
: End-to-end learning for music audio

Pons et al. (2019)
: Randomly weighted CNNs for (music) audio classification

Paper presentations

- You can choose:
  1. Present in class: upload your presentation to QM+ in advance
  2. Record video: (i) upload your presentation to the video QM+ in advance, (ii) upload the video (e.g. into SharePoint) in advance, (iii) share the link of the video via Student Forum on QM+, (iv) response to possible questions.

Sign ups by **17th February**!

Coursework

- Specification
  1. Apply **two** different deep learning methods to infer **two** different aspects of an audio/music task
  2. Explore empirically how to combine the two different inferences
  3. Pick an audio file that will be the case study to inspect the performance of the models in (1) and (2)
- **Coursework proposal** submission:
  1. Deadline: 12th March
  2. Proposals will not be accessed and your proposal can be adjusted
- **Coursework** submission:
  1. Deadline: 14th April
  2. What to submit: 1500-word report PDF.

Open science

We like:

Open source
— Reuse/study code of others; publish your own (e.g. Github)

Open data
— We'll use research datasets

Open access
— Research papers freely online

Be aware of **licences**: Creative Commons, GPL, MIT...

What is machine learning?

What is machine learning?

$$y = f_\theta(x)$$

$$\mathrm{x} \xrightarrow{f_\theta} \mathrm{y}$$

What is machine learning?

$$y = f_\theta(x)$$

$$\mathrm{x} \xrightarrow{f_\theta} \mathrm{y}$$

Parameters θ to be chosen with the aid of data

What is deep learning?

What is deep learning?

$$y = f''''_{\theta''''}(f'''_{\theta'''}(f''_{\theta''}(f'_{\theta'}(f_{\theta}(x)))))$$

$$\mathrm{x} \xrightarrow{f_{\theta}} \cdot \xrightarrow{f'_{\theta'}} \cdot \xrightarrow{f''_{\theta''}} \cdot \xrightarrow{f'''_{\theta'''}} \cdot \xrightarrow{f''''_{\theta''''}} \mathrm{y}$$

Industry examples — analyse

1. Shazam (1999)
   www.shazam.com
2. Google Voice Search (2012)
   en.wikipedia.org/wiki/Google_Voice_Search
3. Amazon Alexa (2014)
   en.wikipedia.org/wiki/Amazon_Alexa
4. Humtap (2014)
   www.humtap.com
5. Spotify 'Discover Weekly' (2015)
   bit.ly/36djQxM
6. FlowMachines 'Daddy's Car' (2016)
   bit.ly/2Rbg6bU
7. DeepMind 'WaveNet' (2016)
   deepmind.com/blog/article/wavenet-generative-model-raw-audio
8. Google 'Differentiable DSP' (2020)
   storage.googleapis.com/ddsp

Discuss: innovations, limitations

Industry examples — analyse

1. Shazam (1999)
   www.shazam.com
2. Google Voice Search (2012)
   en.wikipedia.org/wiki/Google_Voice_Search
3. Amazon Alexa (2014)
   en.wikipedia.org/wiki/Amazon_Alexa
4. Humtap (2014)
   www.humtap.com
5. Spotify 'Discover Weekly' (2015)
   bit.ly/36djQxM
6. FlowMachines 'Daddy's Car' (2016)
   bit.ly/2Rbg6bU
7. DeepMind 'WaveNet' (2016)
   deepmind.com/blog/article/wavenet-generative-model-raw-audio
8. Google 'Differentiable DSP' (2020)
   storage.googleapis.com/ddsp

Discuss: innovations, limitations

Next time...

CNNs (Convolutional Neural Networks)
and key concepts