Deep Neural Network for Real-Time EEG Decoding of Musical Rhythm Imagery

Irán Román

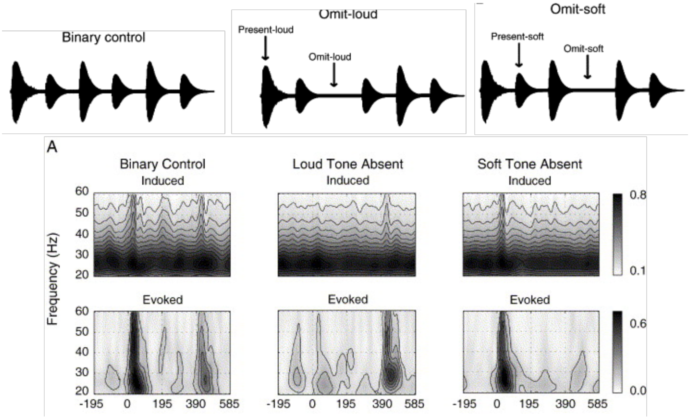
December 03, 2019

Pulse and Meter as Neural Resonance

- Pulse: (aka beat) the repeating, periodic *pulsation* that we *perceive* through time when we listen to music
  - Tempo: the pulse's frequency over time
- Meter: The patterns of accentuation between pulses (i.e. march or waltz)
- Neural Resonance: Music can trigger rhythmic bursts of high-frequency neural activity, which may enable communication between auditory and motor cortices (Large & Snyder 2009)

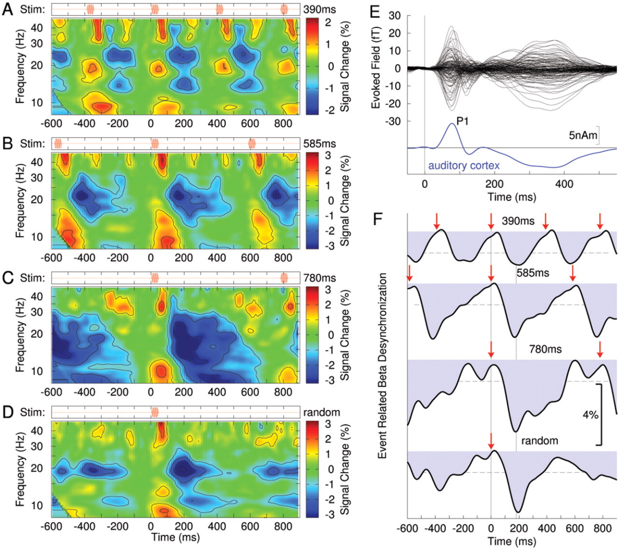
Early EEG Evidence for Neural Resonance

Induced and evoked oscilatory activity reflect the processing and expectation of periodic stimuli



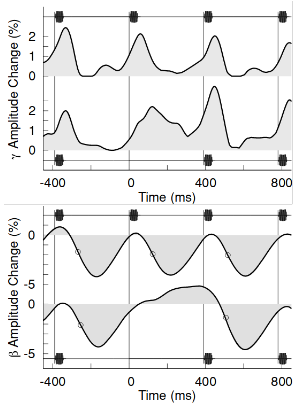
Snyder & Large 2005

Pulse timing is reflected in beta- and gamma-bands



Fujioka et al. 2009

Pulse timing is reflected in beta- and gamma-bands

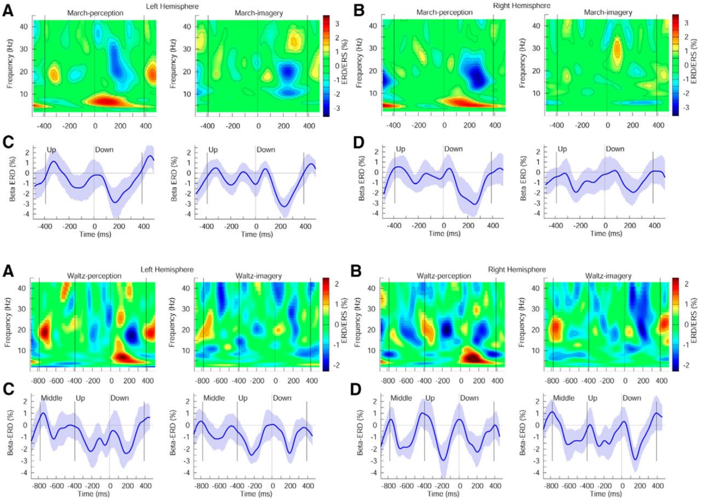


Fujioka et al. 2012

This kind of analysis involves the averaging of hundreds of trials

Imagined meters are also reflected in the beta-band

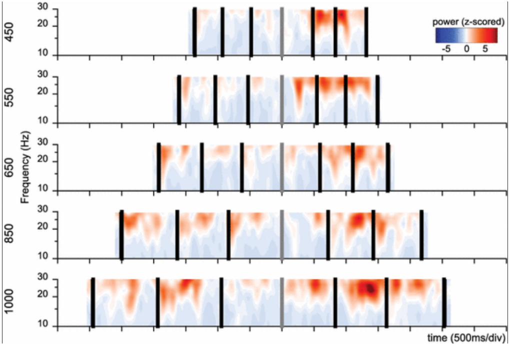
Imagination of different meters (i.e. binary march vs ternary waltz) results in different beta-band patterns



Fujioka et al. 2015

Primate oscillations reflect the metronome tempo

Dorsal putamen LFPs of macaques in a metronome tapping task



Merchant & Bartolo 2017

Bottomline: gamma reflects stimulus processing, while beta reflects the entrainment of large basal ganglia networks during internally driven pulse tapping

Research question and hypothesis

- Pulse, tempo and meter can be observed in the neural correlates measured with EEG and MEG.
- To maximize the SNR, these observations require the analysis of hundreds of trials.
- **Research Question:** could these features be identified in single trials? and if so, can we decode from human brain data, in real-time, perceived and imagined musical features like pulse, tempo and meter?
- **Hypothesis:** Deep neural networks, like CNNs, can learn to identify these musical features on a single-trial level.

Lit Review: What can CNNs learn from EEG data?

- Individual participant classification with resting-state data (Ma et al. 2018)
- Motor imagery classification: left vs rigt hand (Tang et al. 2017)
- Motor imagery with Riemannian Geometry Classifiers (Barachant et al. 2011)
- Music imagery information retrieval (Tan et al. 2018)
- The key problem: domain adaptation and transfer learning (Lotte et al. 2018)

Individual participant classification (Ma et al. 2018)

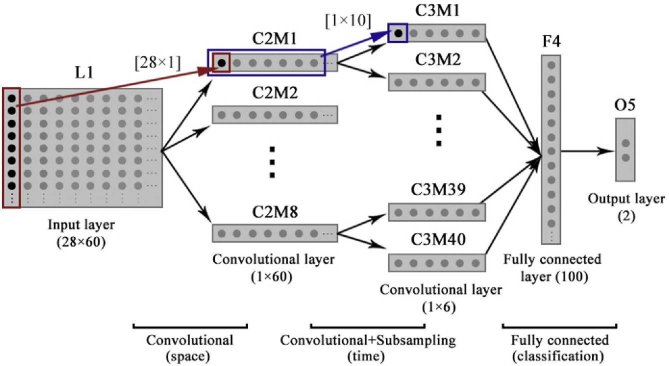
- Task: resting state with open and closed eyes
- CNN input: 64x160 EEG "images"
  - 64 EEG channels
  - 1 second-long epochs ($fs = 160$)
  - 50 (5) seconds of training (test) data per participant
  - 10 participants
- CNN architecture: LeNet (6 channels in each conv layer)
- Performance: **88% accuracy**
- Significance:
  - EEG for biometrics

Motor imagery classification: left vs right hand (Tang et al. 2018)

- Task: imagination of right or left hand movements during 5 second trials (460 total)
- CNN input: 28x60
  - 28 EEG channels
  - 3 seconds, 50ms fft frames (1 frequency band per participant)
  - 350 (100) trials of traning (test) data per participant
  - 2 participants

Motor imagery classification: left vs right hand (Tang et al. 2018)

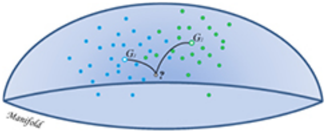
- CNN architecture



- Performance: **87% accuracy**
- Significance:
  - Using CNNs to classify motor imagery EEG

Motor imagery with Rimennian Geometry Classifiers (Barachant et al. 2011)

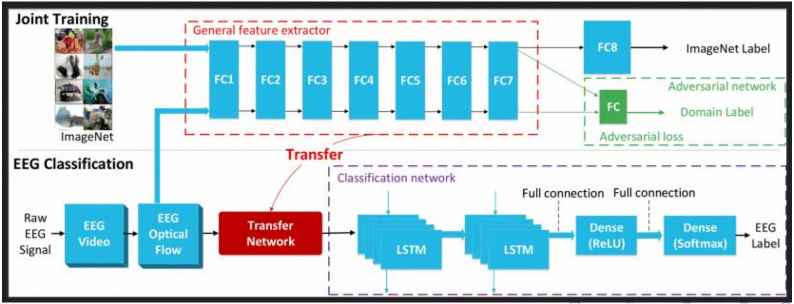
- Riemannian geometry studies smooth curved spaces that can be locally and linearly approximated with tangents
- Assumption: each EEG mental state has different power and spatial distribution
- The covariance matrix of the EEG data can code this information



- Performance: **70% accuracy** in a four-class, motor imagery task (four extremities)
- Significance:
  - Considered to be the *state-of-the-art* (Lotte et al. 2018)

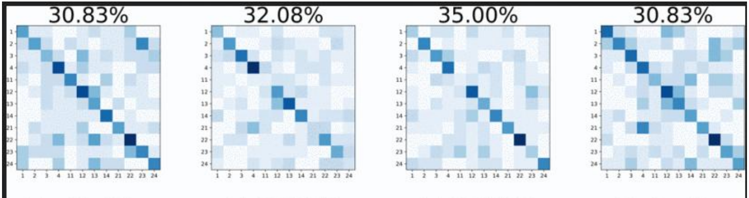
Music imagery information retrieval (Tan et al. 2018)

- OpenMIIR dataset: listening and imagination of 12 known musical stimuli (Stober et al. 2015)
- CNN input: raw EEG signal
  - 64 EEG channels
  - 240 trials per participant (12 stimuli x 4 conditions x 5 blocks). Avg trial length (1̃0 sec)
  - 10 participants
- CNN architecture



Music imagery information retrieval (Tan et al. 2018)

- Performance:



(a) AlexNet (b) VGG16 (c) VGG19 (d) ResNet

- Significance:
  - Deep transfer learning, exploiting EEG *multimodality*, and using joint training for knowledge transfer.

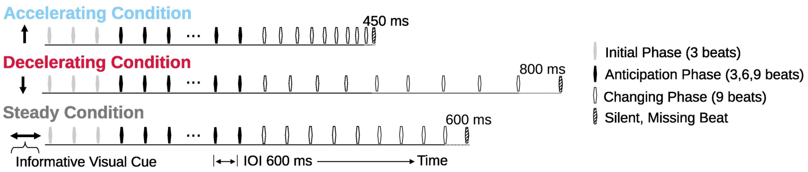
The key problem: domain adaptation and transfer learning (Lotte et al. 2018)

- In EEG, major changes in data distribution can occur between participants or across time.
- Solution: transfer learning (calibration) to optimize the algorithm across time and across participants
- The solution involves one or more of the following:
  - Learn the transformation that explains the change in the data distribution
  - Reweighting
  - Find the common features between the two distributions

Research question and hypothesis

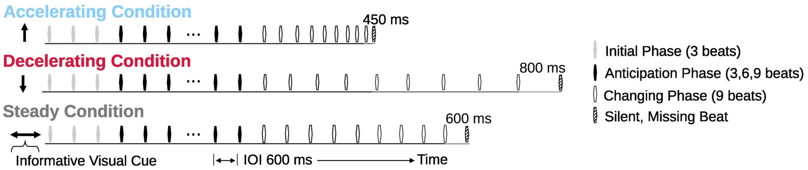
- Pulse, tempo and meter can be observed in the neural correlates measured with EEG and MEG.
- To maximize the SNR, these observations require the analysis of hundreds of trials.
- **Research Question:** could these features be identified in single trials? and if so, can we decode from human brain data, in real-time, perceived and imagined musical features like pulse, tempo and meter?
- **Hypothesis:** Deep neural networks, like CNNs, can learn to identify these musical features on a single-trial level.

Methods: EEG dataset task



- 21 Participants listened 450 trials, each consisting of $\sim$ 18 metronome clicks that either:
  - accelerate
  - decelerate
  - remain isochronous
- The metronome starts at 1.67Hz (600ms inter-stimulus-interval)
- The three types of trials are equally likely to appear and are presented randomly.

Methods: EEG dataset task



- Stimuli presentation was evenly divided across eight blocks
- At the beginning of each trial, participants see a visual cue that tells them what the metronome will do.
- No behavioral task related to metronome rate changes.

Methods: EEG data collection equipment

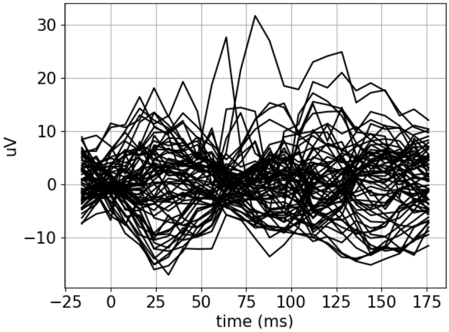
- 64 channel Neuroscan
- 500Hz sampling rate
- 2 electro-oculogram channels to record horizontal and vertical eye movements
- Stimuli delivered by insert earphones and a monitor

Methods: EEG data processing

- 10 noisy EEG channels were eliminated from all recordings.
- The noisy EEG channels were at the edge of the EEG cap.
- Eye movement artifacts were removed using the signal-space projection method.
- 200ms epochs were calculated starting 16ms before each click or visual cue.
- Epochs were baseline corrected with the mean of the eeg amplitude before each click (100ms).
- Epochs were downsampled from 500Hz to 125Hz sampling rate.
- Rejected epochs with EEG peak-to-peak difference greater than 100μV in any channel.

Methods: CNN input data

- 10% of all epochs were held-out as validation data
  - 133,813 training epochs (7,363 visual epochs)
  - 14,791 validation epochs (803 visual epochs)
- ∼6,300 epochs per participant (21 participants total)
- Epochs were multiplied by 10^6 and cast to be float32
- Data shape is: `[None, 54, 1, 25]`



Methods: LeNet architecture

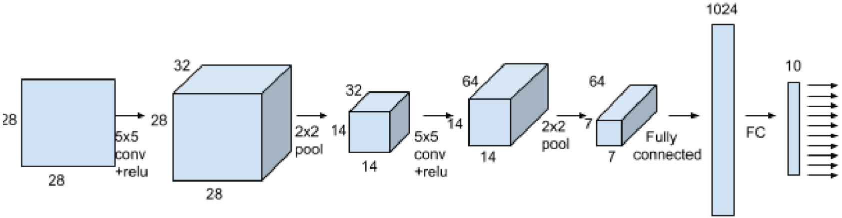


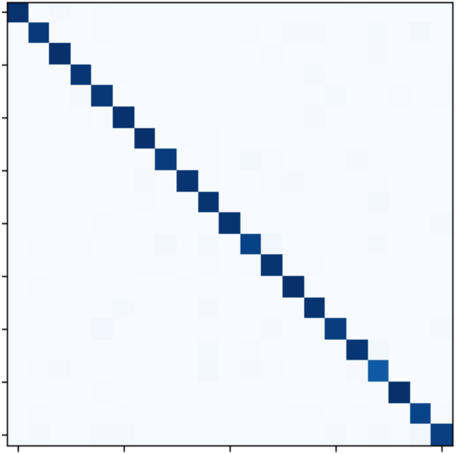
Figure D.2: Network architecture for MNIST classifier CNN

Methods: EEG classification experiments

- Subject classification
- Pulse vs visual stimulus classification
- Accelerating vs decelerating vs steady beat classification

Results: Subject classification

- **95% accuracy**
- Trained with pulse epochs (no visual stimuli epochs)
  - ~6,000 per participant
- Confusion matrix

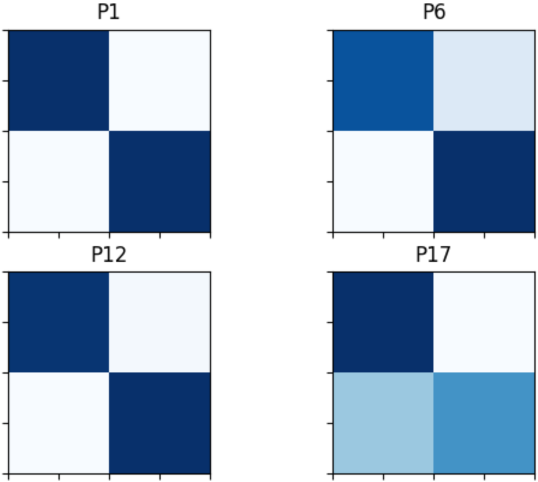


Results: Pulse vs visual stimulus classification

$\sim$350 visual epochs and $\sim$6,000 pulse epochs per participant

Accuracy per participant

P1	0.98
P2	0.82
P3	0.90
P4	0.85
P5	0.94
P6	0.78
P7	0.91
P8	0.94
P9	0.87
P10	0.86
P11	0.94
P12	0.99
P13	0.95
P14	0.88
P15	0.92
P16	0.93
P17	0.74
P18	0.86
P19	0.93
P20	0.83
P21	0.87

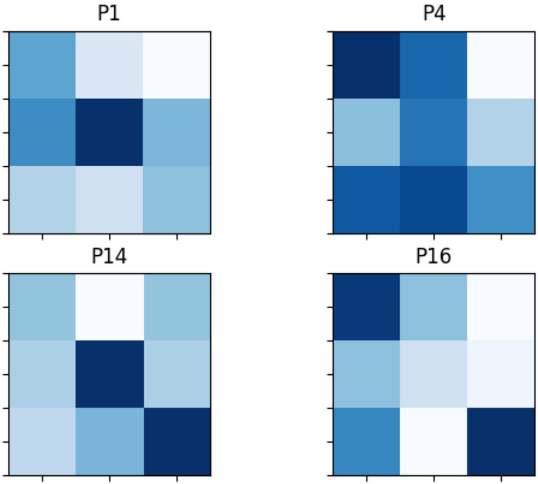


Results: Accelerating vs decelerating vs steady beat classification. "Tempo changing pulses"

∼2,800 "tempo changing" epochs per participant

Accuracy per participant

P1	0.43
P2	0.36
P3	0.38
P4	0.36
P5	0.37
P6	0.37
P7	0.39
P8	0.40
P9	0.37
P10	0.40
P11	0.41
P12	0.37
P13	0.40
P14	0.39
P15	0.40
P16	0.40
P17	0.36
P18	0.46
P19	0.42
P20	0.38
P21	0.37

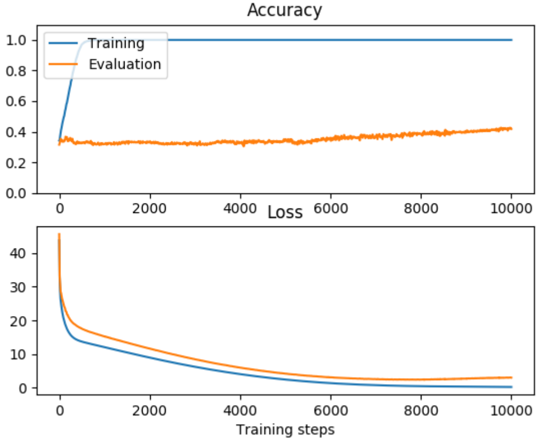


Results: Accelerating vs decelerating vs steady beat classification. "Tempo changing pulses"

∼2,800 "tempo changing" epochs per participant

Accuracy per participant

P1	0.43
P2	0.36
P3	0.38
P4	0.36
P5	0.37
P6	0.37
P7	0.39
P8	0.40
P9	0.37
P10	0.40
P11	0.41
P12	0.37
P13	0.40
P14	0.39
P15	0.40
P16	0.40
P17	0.36
P18	0.46
P19	0.42
P20	0.38
P21	0.37

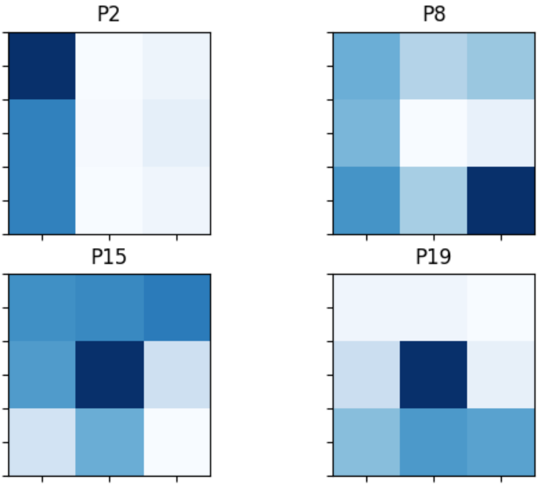


Results: Accelerating vs decelerating vs steady beat classification. Steady pulses

∼3,200 steady tempo epochs per participant

Accuracy per participant

P1	0.38
P2	0.41
P3	0.35
P4	0.37
P5	0.37
P6	0.39
P7	0.38
P8	0.35
P9	0.43
P10	0.42
P11	0.40
P12	0.36
P13	0.41
P14	0.38
P15	0.35
P16	0.41
P17	0.41
P18	0.35
P19	0.43
P20	0.41
P21	0.38



Ways to improve these results

- Use more available data
- Data augmentation
- Hyper-parameter search
- Use a different CNN architecture
- ...

Some conclusions (so far)

- EEG data allows for classification of individual participants on a single-trial basis, with high accuracy
  - Highlights the difference between individual participant's EEG data (Ma et al. 2018)
- On an individual participant level, accurate classification of visual vs auditory EEG data is also possible on a single-trial level.
- Classification of different "mental musical states" is a challenging problem and CNN performance needs to improve.

Next steps

- The elephant in the room: pulse identification (over 125,000 datapoints available)