LEARNING TO GENERATE MUSIC WITH BACHPROP

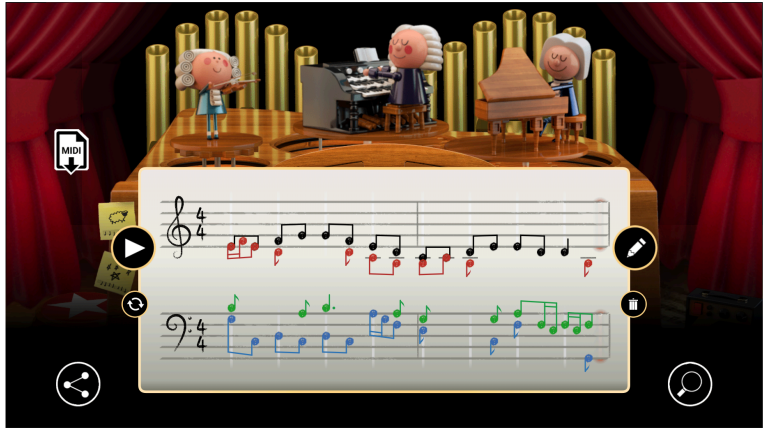
CS-456 Artificial Neural Networks | Florian Colombo | LCN

Deep Generative Models | Images

Karras et al. ICLR18

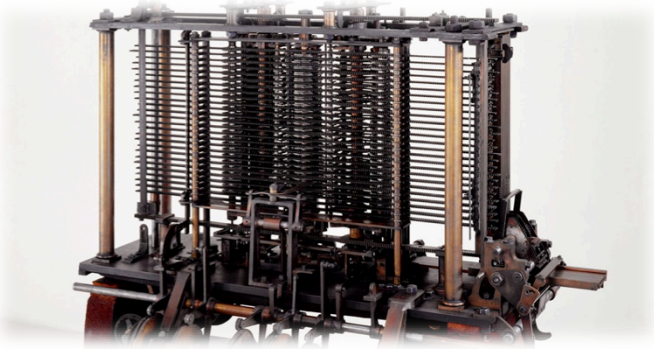
Deep Generative Models | Music

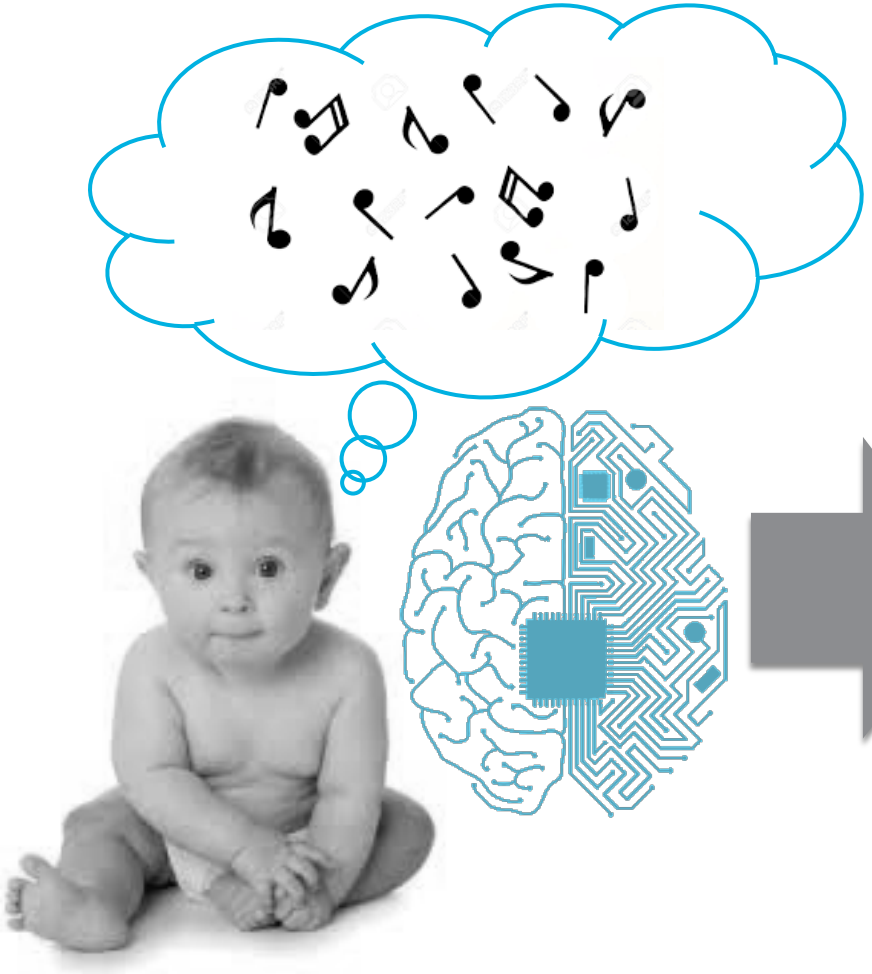
magenta



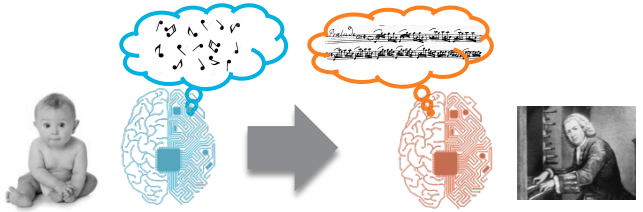
Ada Lovelace | First Programmer

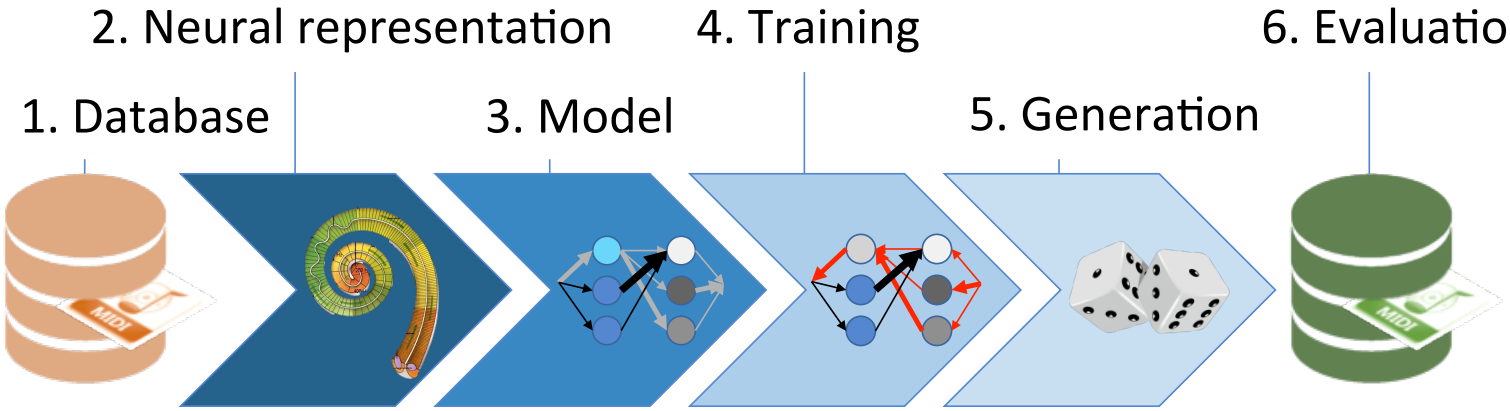
"Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might **compose elaborate and scientific pieces of music** of any degree of complexity or extent."




Prélude

BachProp | Algorithm

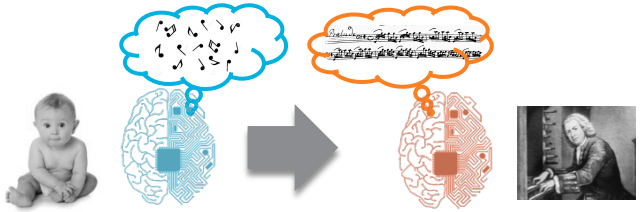


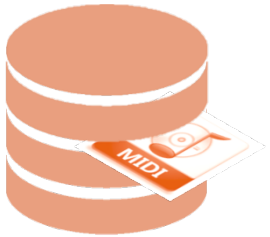


Original corpus

Generated corpus

Datasets

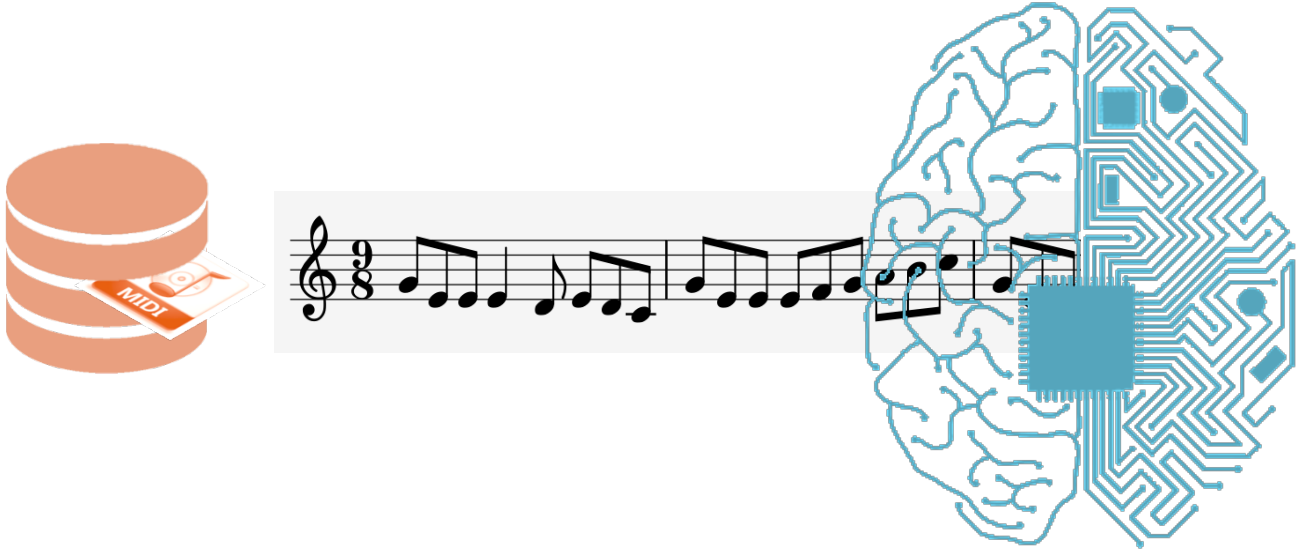




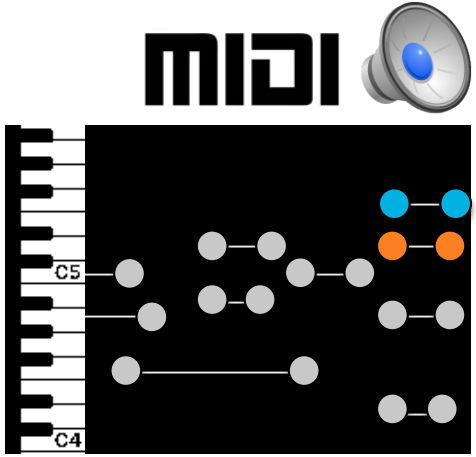
Original corpus

- **Chorales**
  - 381 4-voices chorales harmonized by J.S. Bach
- **John Sankey**
  - 135 keyboard pieces from J.S. Bach recorded live by John Sankey
- **Nottingham**
  - 1035 folk tunes with homogeneous structure
- **String quartets**
  - 215 Haydn and Mozart string quartets

Neural representation | Ear of an ANN



Preprocessing | MIDI 2 Score

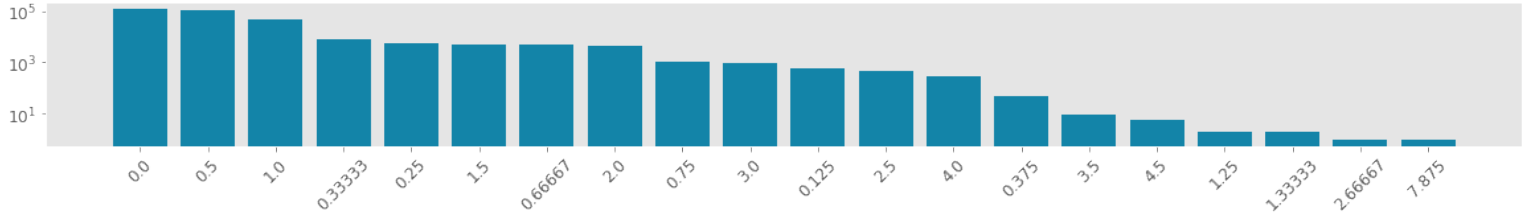


$$[dT_{\text{MIDI}} = 0, P = 74]_{\text{Note ON}}$$
$$[dT_{\text{MIDI}} = 2, P = 77]_{\text{Note ON}}$$
$$[dT_{\text{MIDI}} = 0, P = 74]_{\text{Note OFF}}$$
$$[dT_{\text{MIDI}} = 3, P = 77]_{\text{Note OFF}}$$

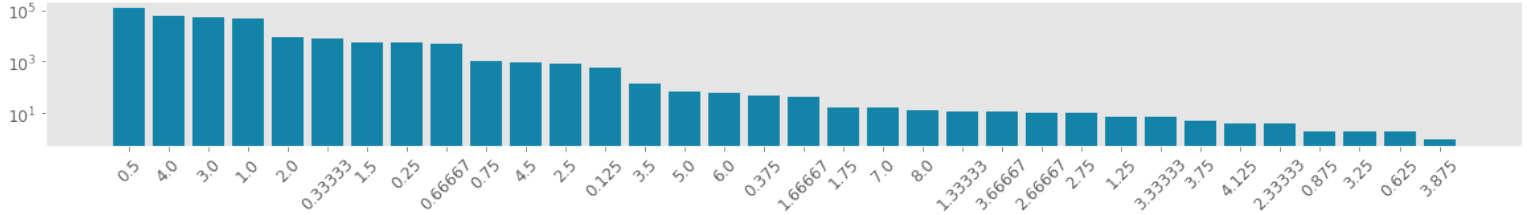
$$\begin{pmatrix} \mathbf{note}[1] \\ \vdots \\ \mathbf{note}[n] \\ \mathbf{note}[n+1] \\ \vdots \\ \mathbf{note}[N] \end{pmatrix} \rightarrow \begin{pmatrix} dT = 0 \\ T = \\ P = \text{D5} \end{pmatrix}, \begin{pmatrix} dT = 0 \\ T = \\ P = \text{F5} \end{pmatrix}$$

Feature dictionaries | Distributions

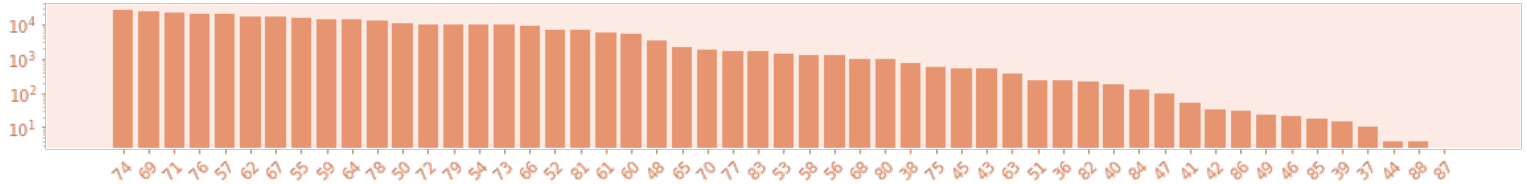
dT



T

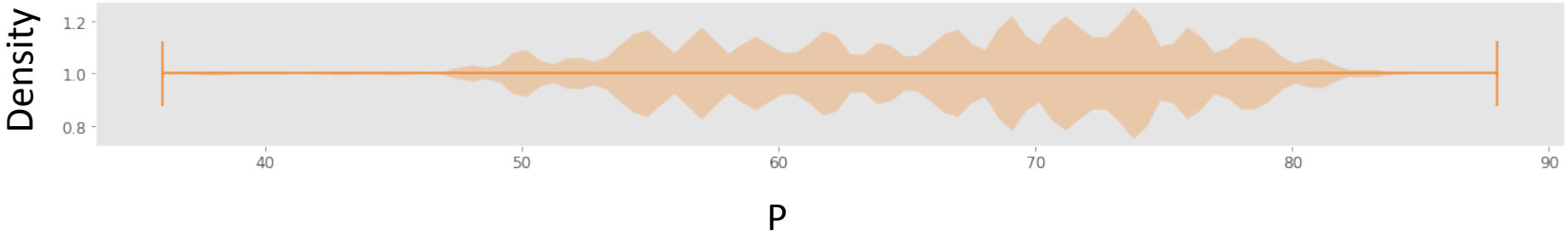


P



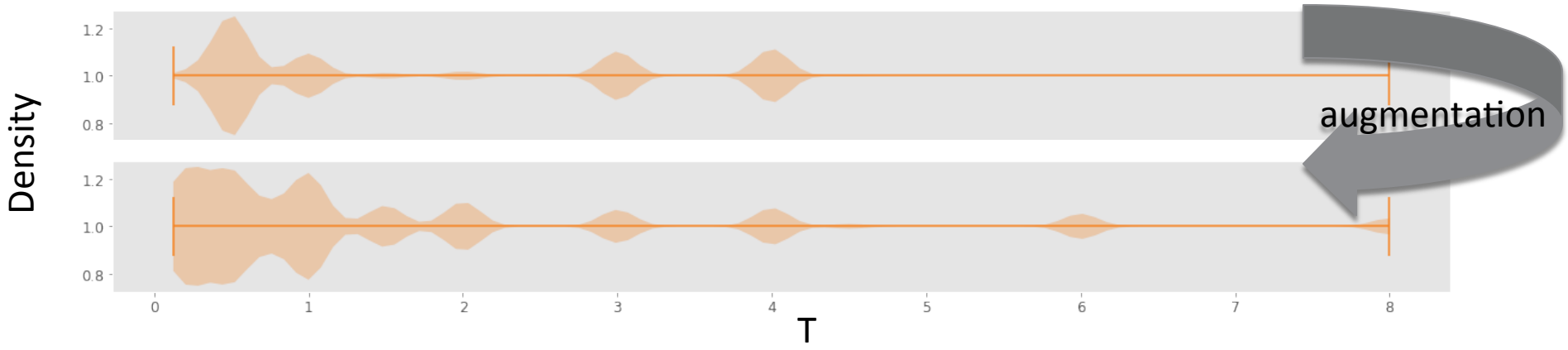
Data Augmentation | Pitch

- Music is transposition invariant
  - Transpose in every possible keys to increase the number of training samples

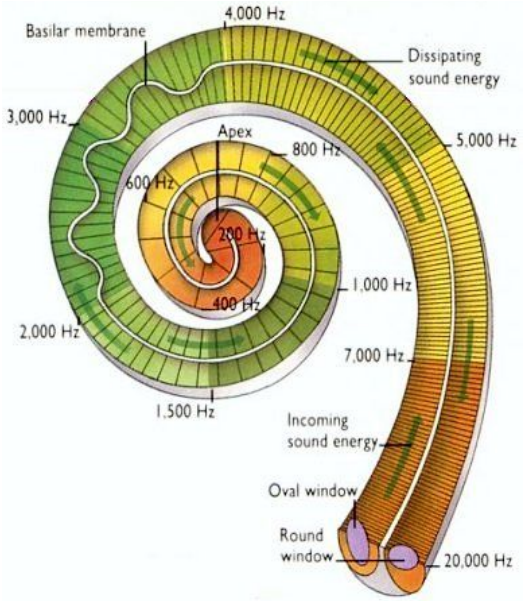


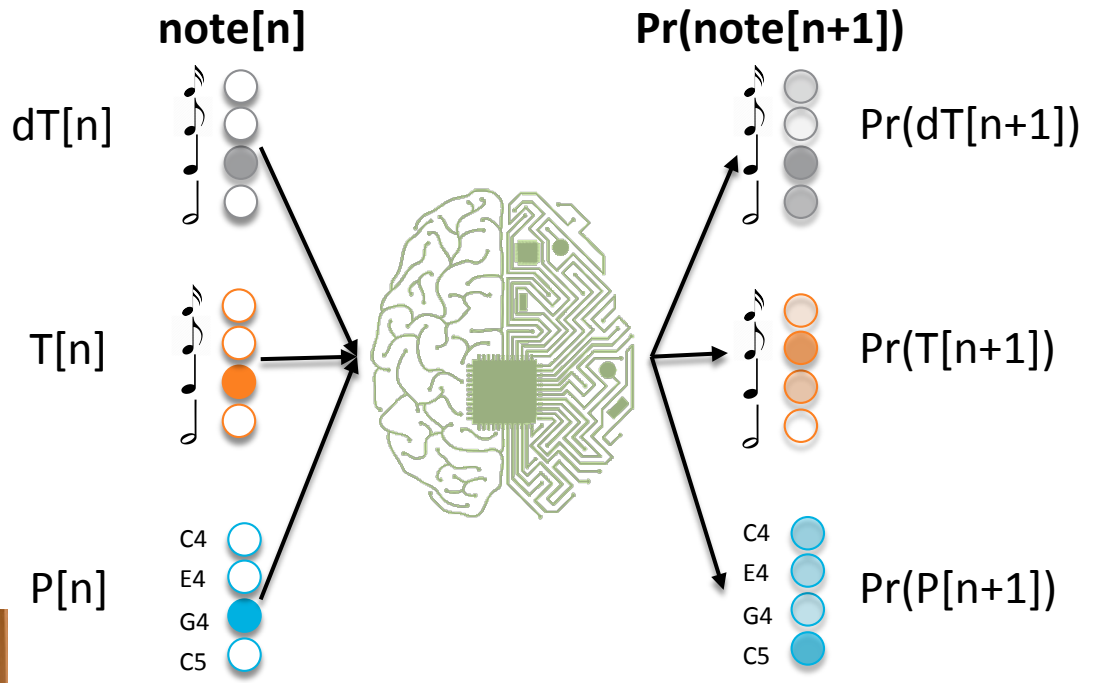
Data Augmentation | Rhythm

- Dilate and contract durations
  - Factors: [0.125, 0.25, 0.33, 0.5, 0.66, 1.0, 1.5, 2., 2.5, 3., 4.]
  - Constrain on the set of possible durations

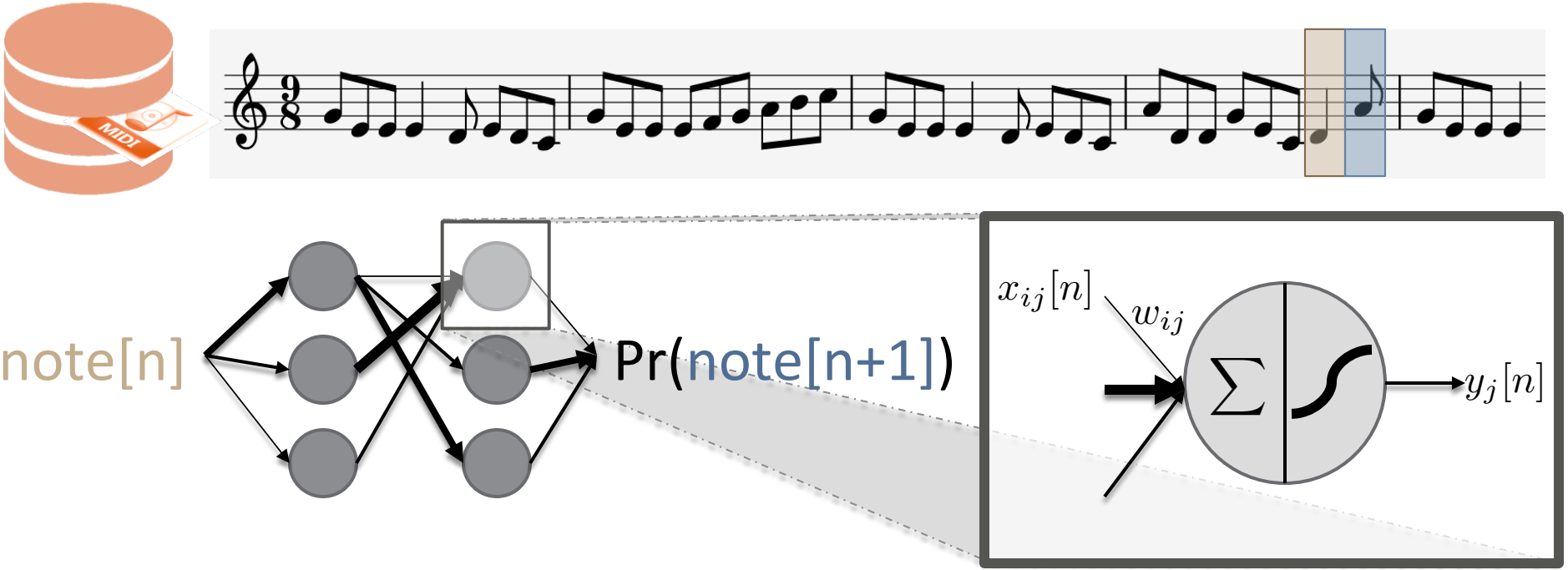


Data Representation | I/O

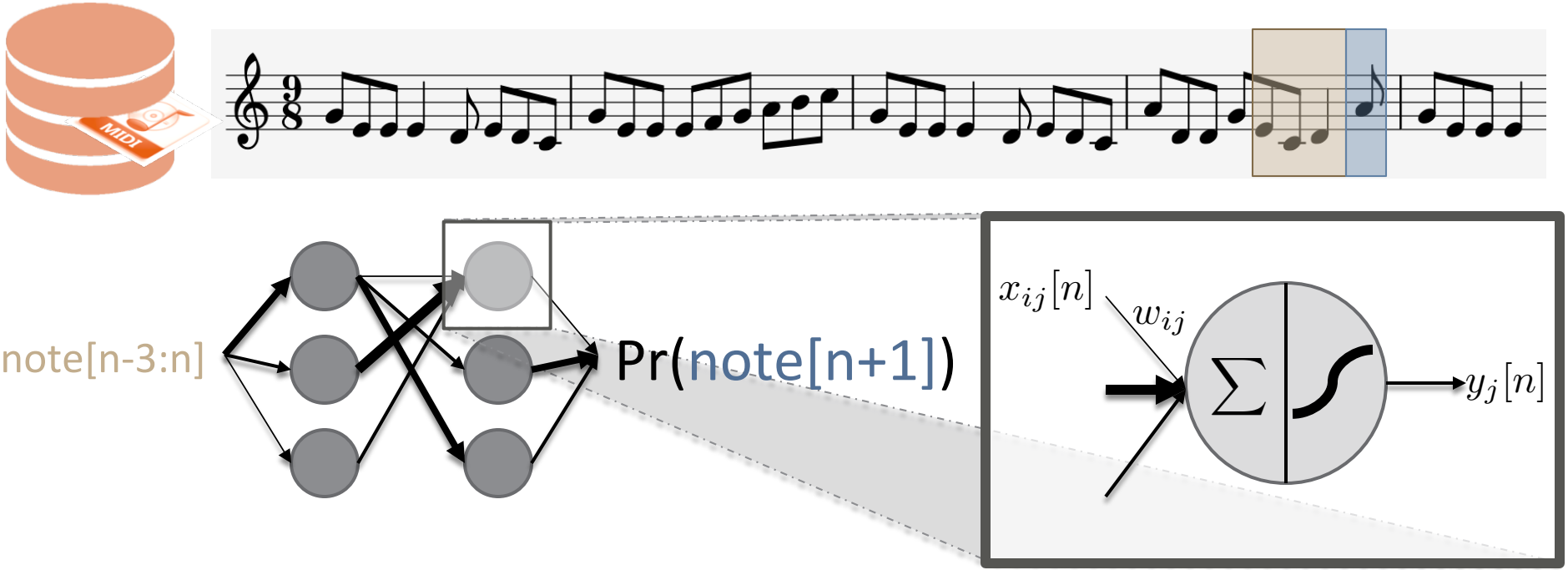




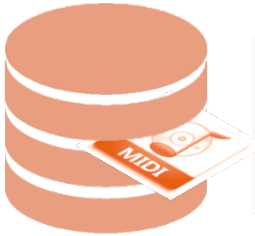
Model | Feedforward

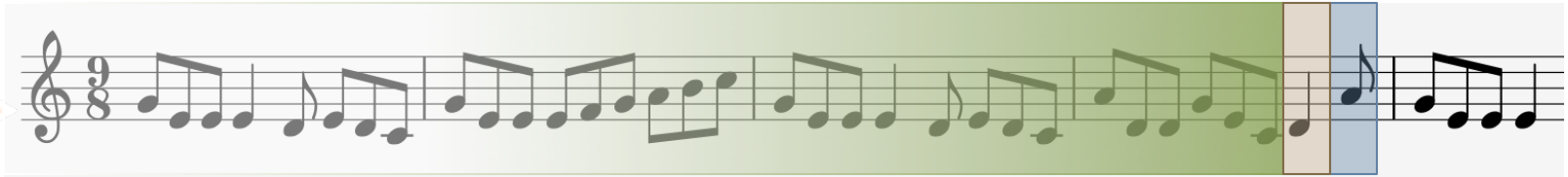


Model | Feedforward

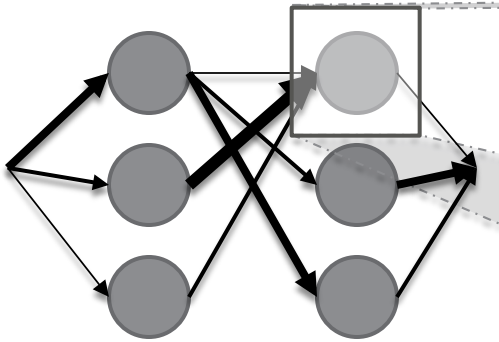


Model | Recurrent

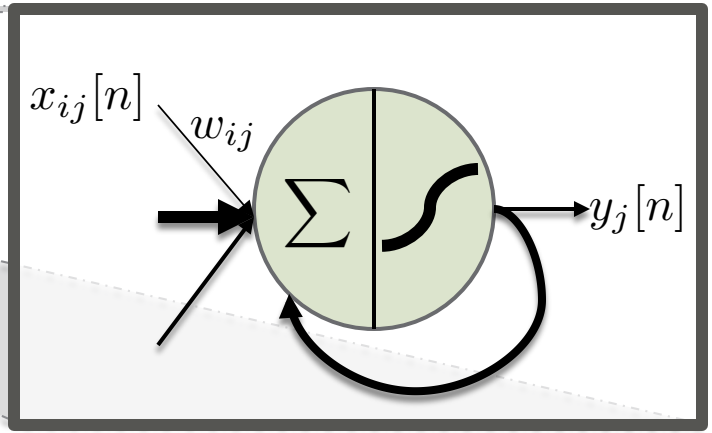




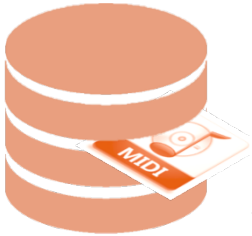
note[n]

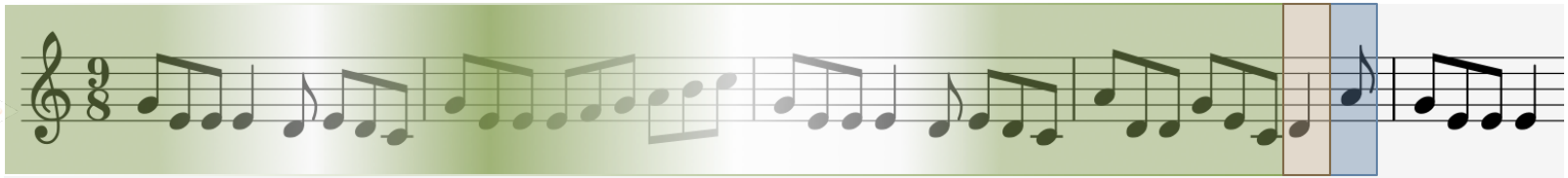


Pr(note[n+1])

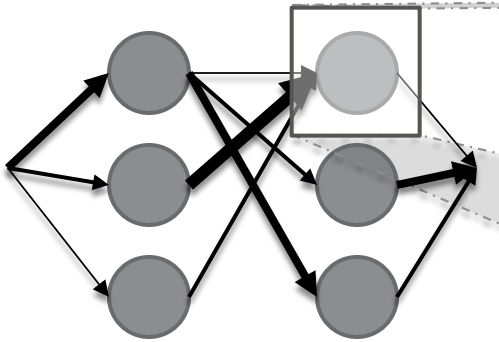


Model | LSTM (and GRU)

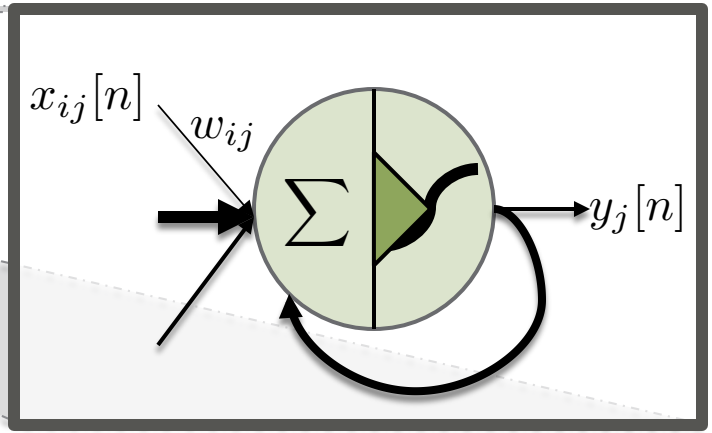




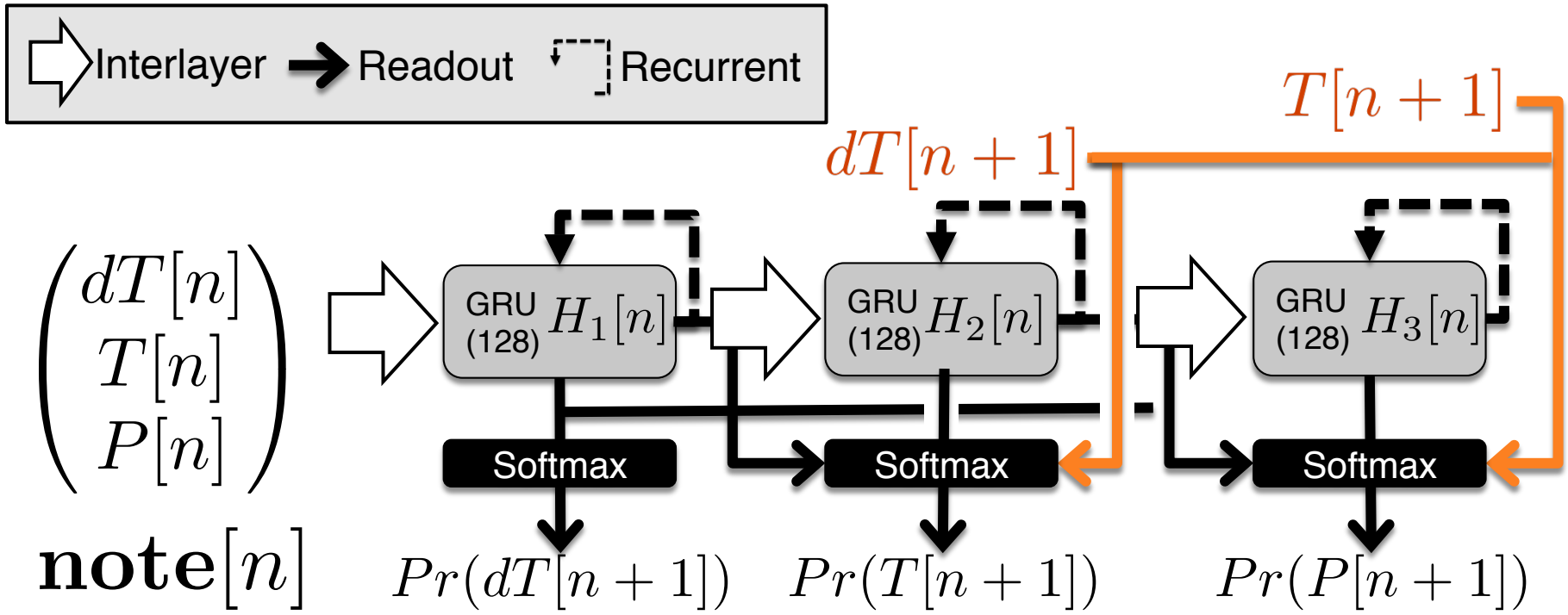
note[n]



Pr(note[n+1])

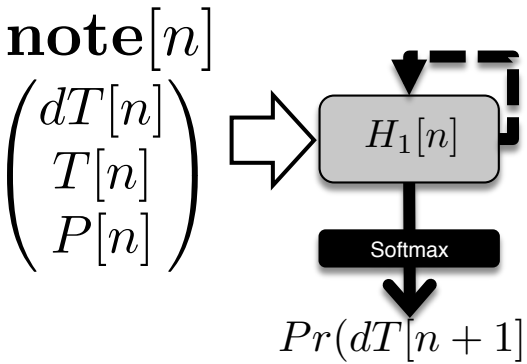


Neural Network Architecture



Probabilistic interpretation

$$Pr(\mathbf{note}[1:N]) = Pr(\mathbf{note}[1]) \prod_{n=1}^{N-1} Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n])$$



$$Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n]) =$$

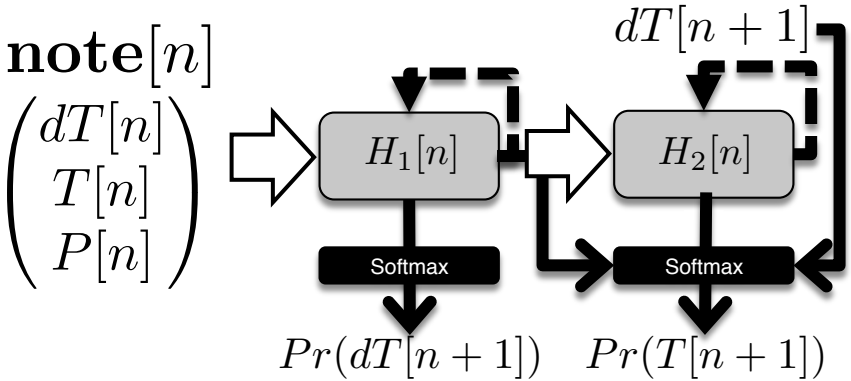
$$Pr(dT[n+1]|\mathbf{note}[1:n]) \times \approx Pr(dT[n+1]|H_{dT}[n])$$

$$Pr(T[n+1]|\mathbf{note}[1:n], dT[n+1]) \times \approx Pr(T[n+1]|H_T[n], dT[n+1])$$

$$Pr(P[n+1]|\mathbf{note}[1:n], dT[n+1], T[n+1]) \approx Pr(P[n+1]|H_P[n], dT[n+1], T[n+1])$$

Probabilistic interpretation

$$Pr(\mathbf{note}[1:N]) = Pr(\mathbf{note}[1]) \prod_{n=1}^{N-1} Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n])$$



$$Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n]) =$$

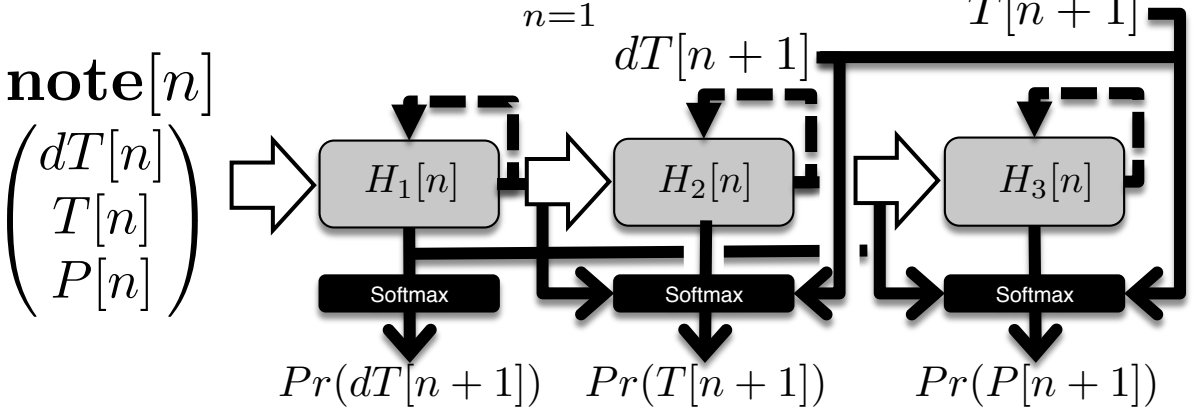
$$Pr(dT[n+1]|\mathbf{note}[1:n]) \times \approx Pr(dT[n+1]|H_{dT}[n])$$

$$Pr(T[n+1]|\mathbf{note}[1:n], dT[n+1]) \times \approx Pr(T[n+1]|H_T[n], dT[n+1])$$

$$Pr(P[n+1]|\mathbf{note}[1:n], dT[n+1], T[n+1]) \approx Pr(P[n+1]|H_P[n], dT[n+1], T[n+1])$$

Probabilistic interpretation

$$Pr(\mathbf{note}[1:N]) = Pr(\mathbf{note}[1]) \prod_{n=1}^{N-1} Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n])$$



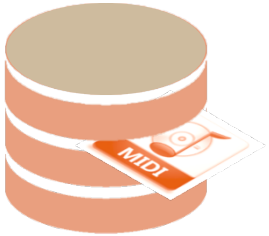
$$Pr(\mathbf{note}[n+1]|\mathbf{note}[1:n]) =$$

$$Pr(dT[n+1]|\mathbf{note}[1:n]) \times \approx Pr(dT[n+1]|H_{dT}[n])$$

$$Pr(T[n+1]|\mathbf{note}[1:n], dT[n+1]) \times \approx Pr(T[n+1]|H_T[n], dT[n+1])$$

$$Pr(P[n+1]|\mathbf{note}[1:n], dT[n+1], T[n+1]) \approx Pr(P[n+1]|H_P[n], dT[n+1], T[n+1])$$

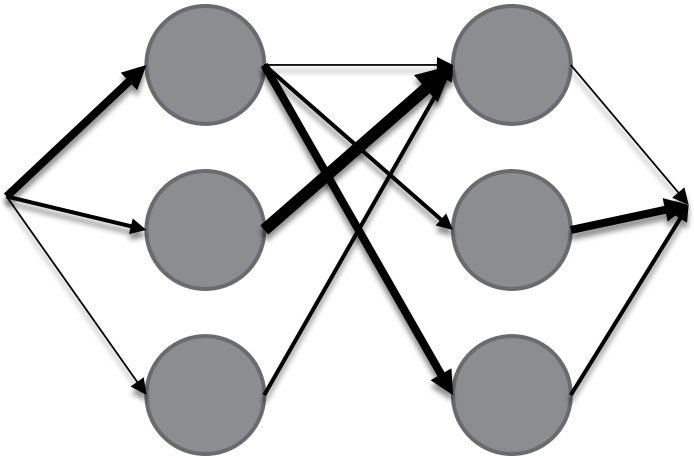
Training | Forward pass



10%

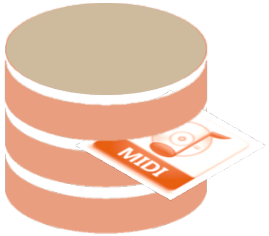
90%

note[n]



Pr(note[n+1])

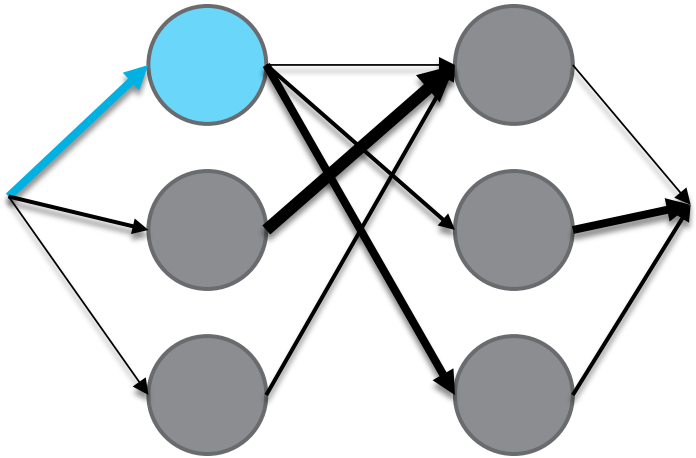
Training | Forward pass



10%

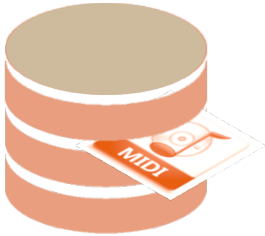
90%

note[n]



Pr(note[n+1])

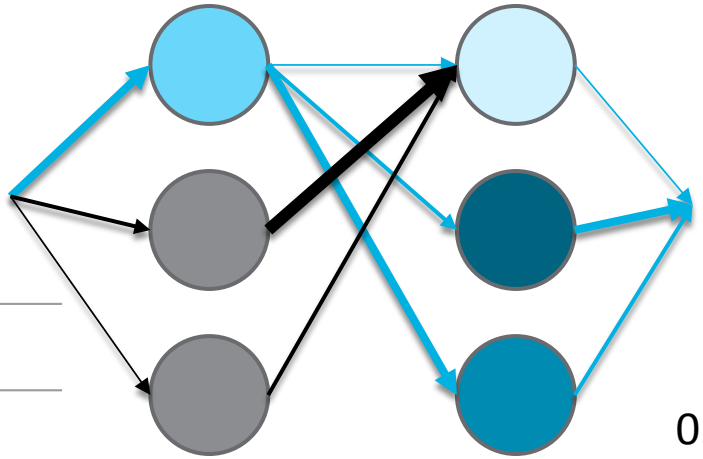
Training | Epoch performance



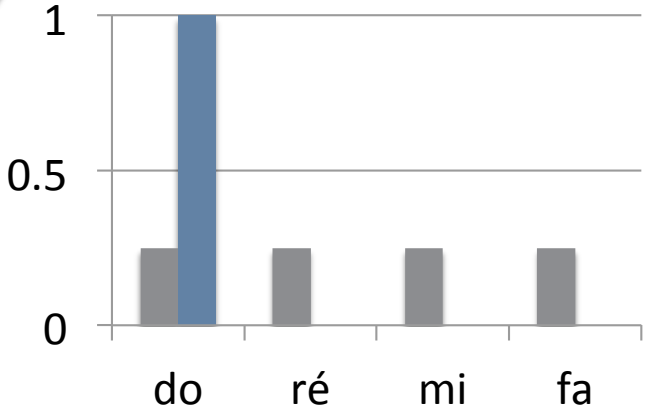
10%

90%

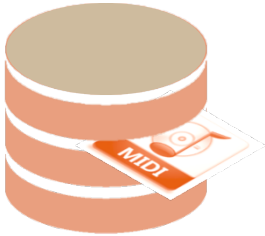
note[n]



Pr(note[n+1])



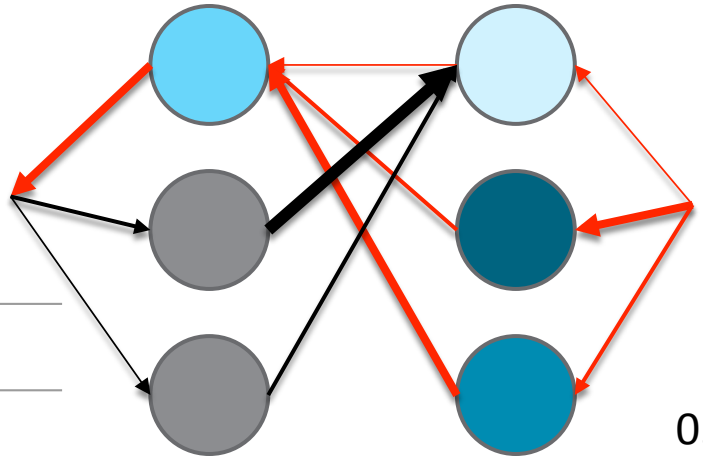
Training | BackProp through time



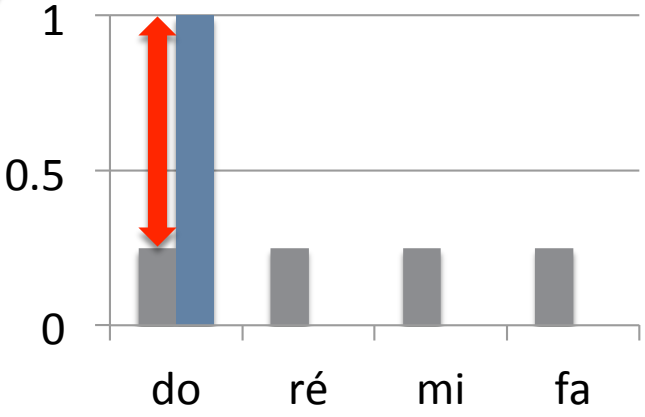
10%

90%

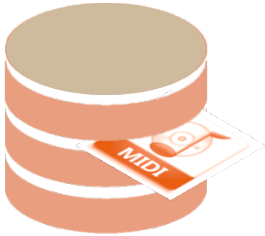
note[n]



Pr(note[n+1])



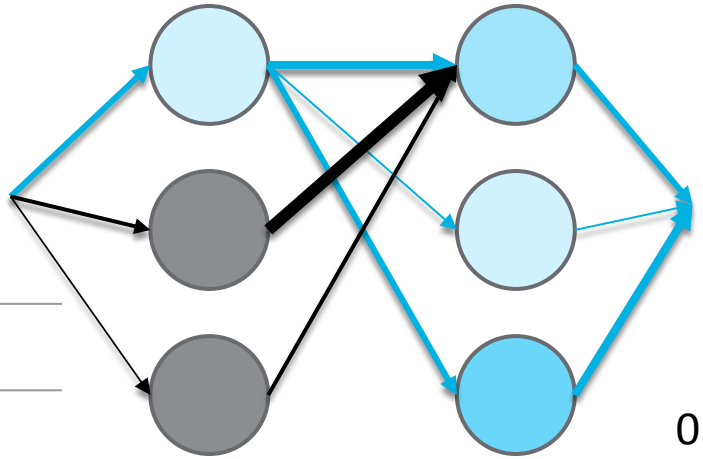
Training | Update parameters



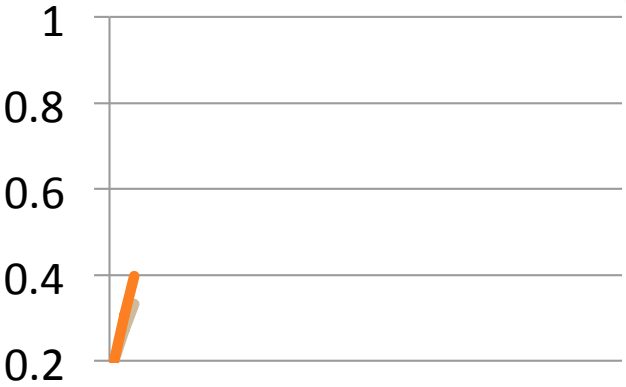
10%

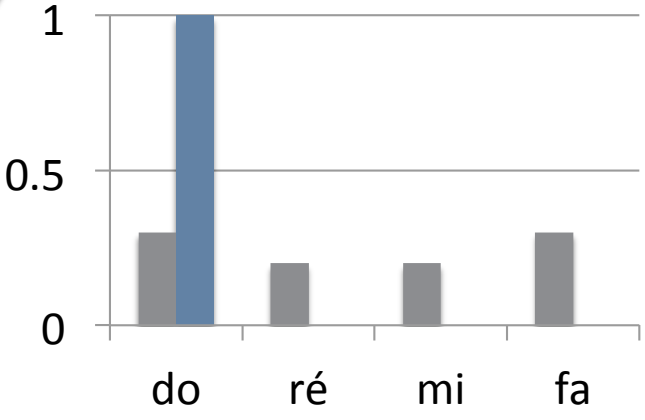
90%

note[n]

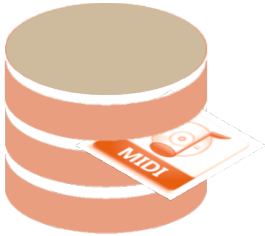


Pr(note[n+1])





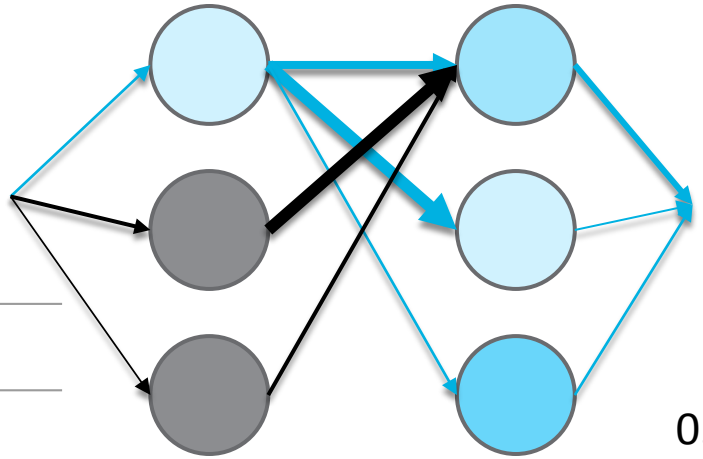
Training | Update parameters



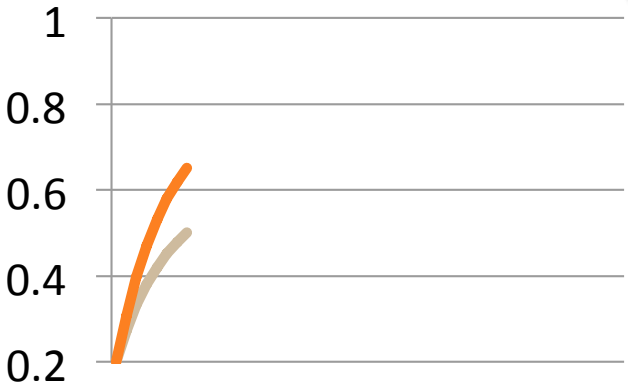
10%

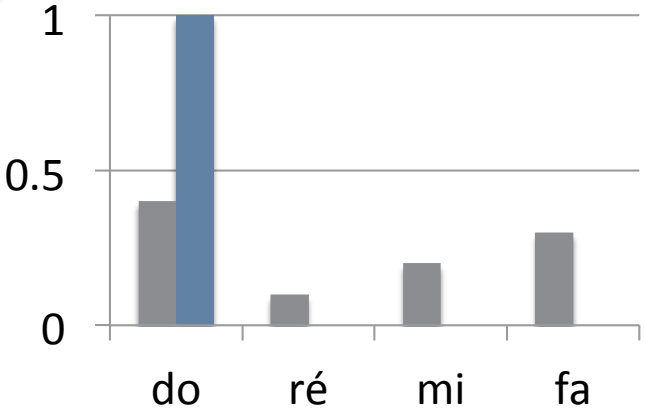
90%

note[n]

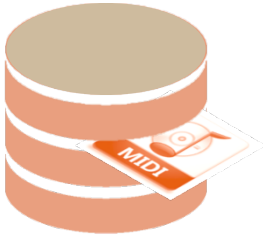


Pr(note[n+1])





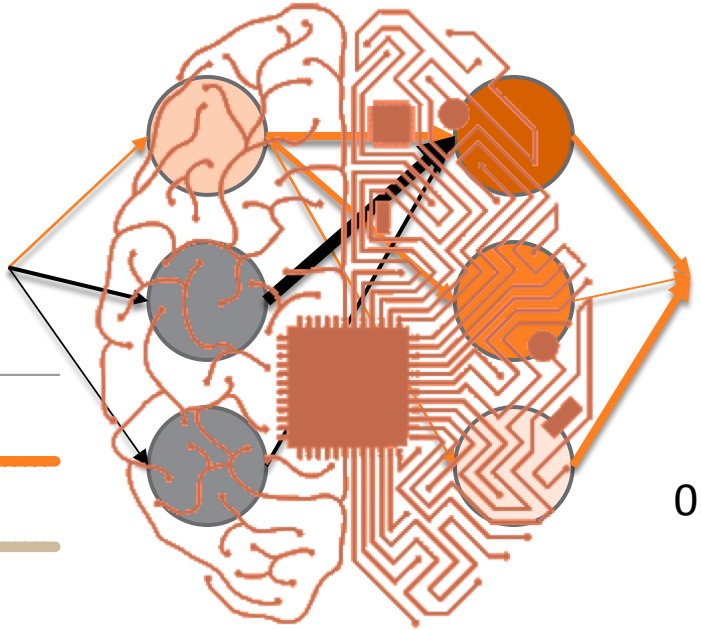

Training | End



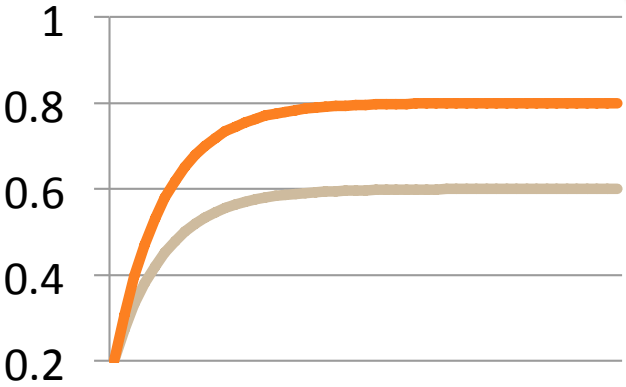
10%

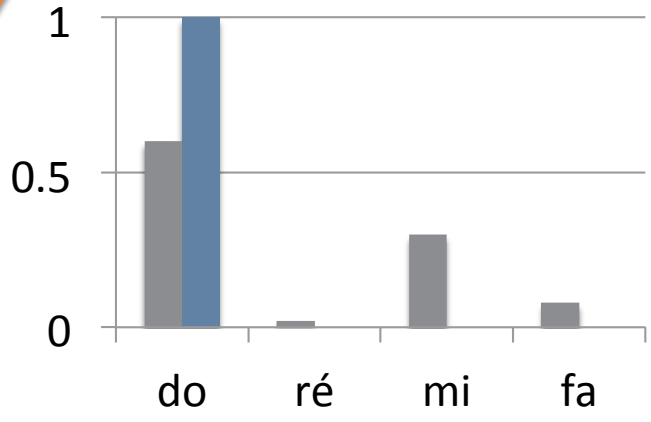
90%

note[n]

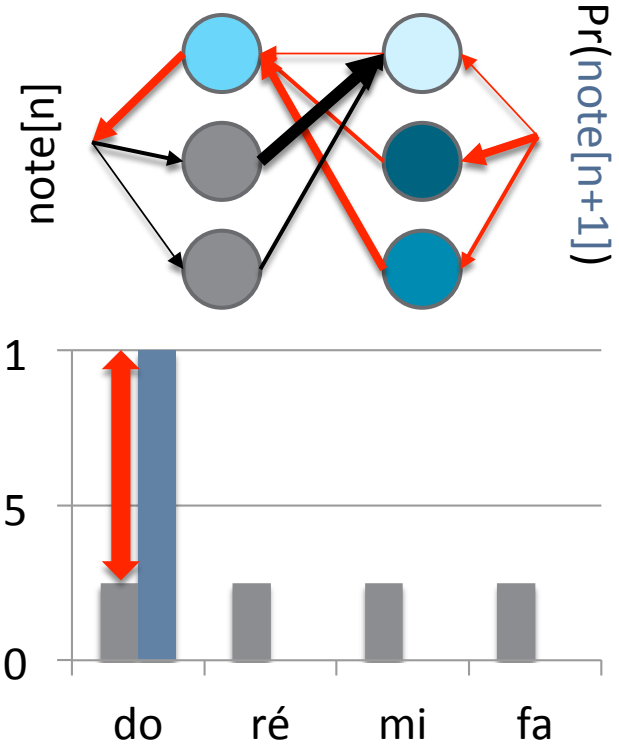


Pr(note[n+1])



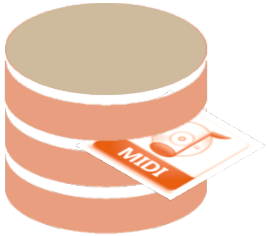


Training | Modalities



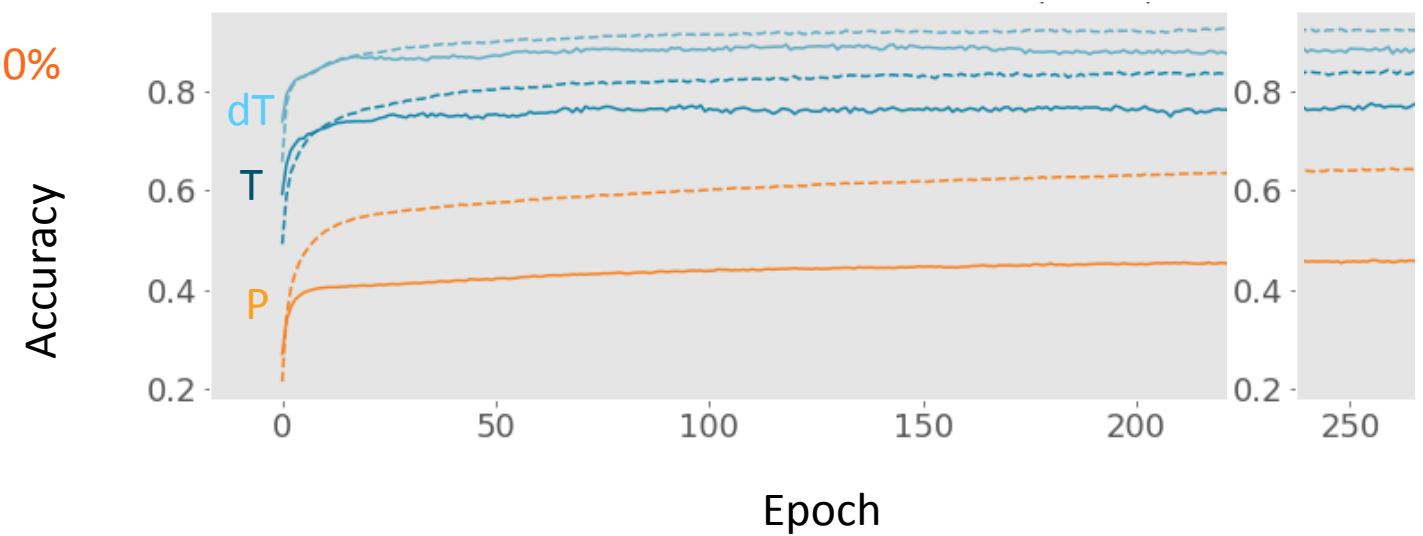
- TBPTT with 128 notes truncation
- Recurrent cells are stateful
  - Hidden state is maintained after truncation
- Adam optimizer
  - Weighted (dT: 0.1, T: 0.3, P: 0.6) cross-entropy loss functions

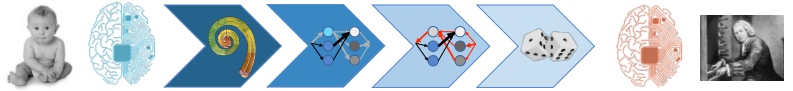
Training | Learning Curves



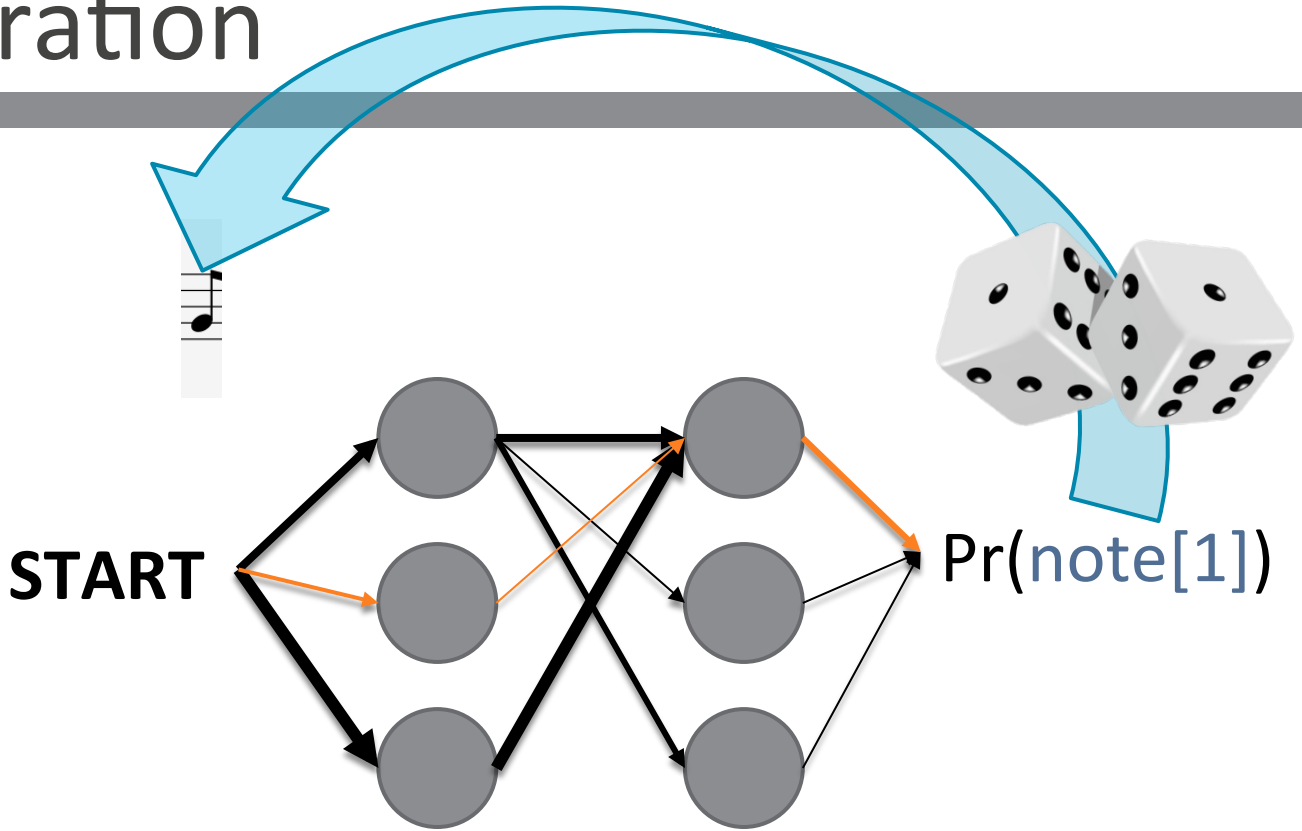
10%

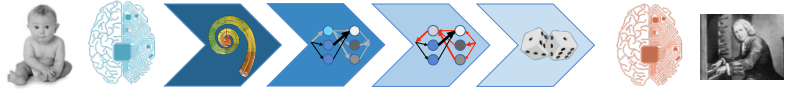
90%



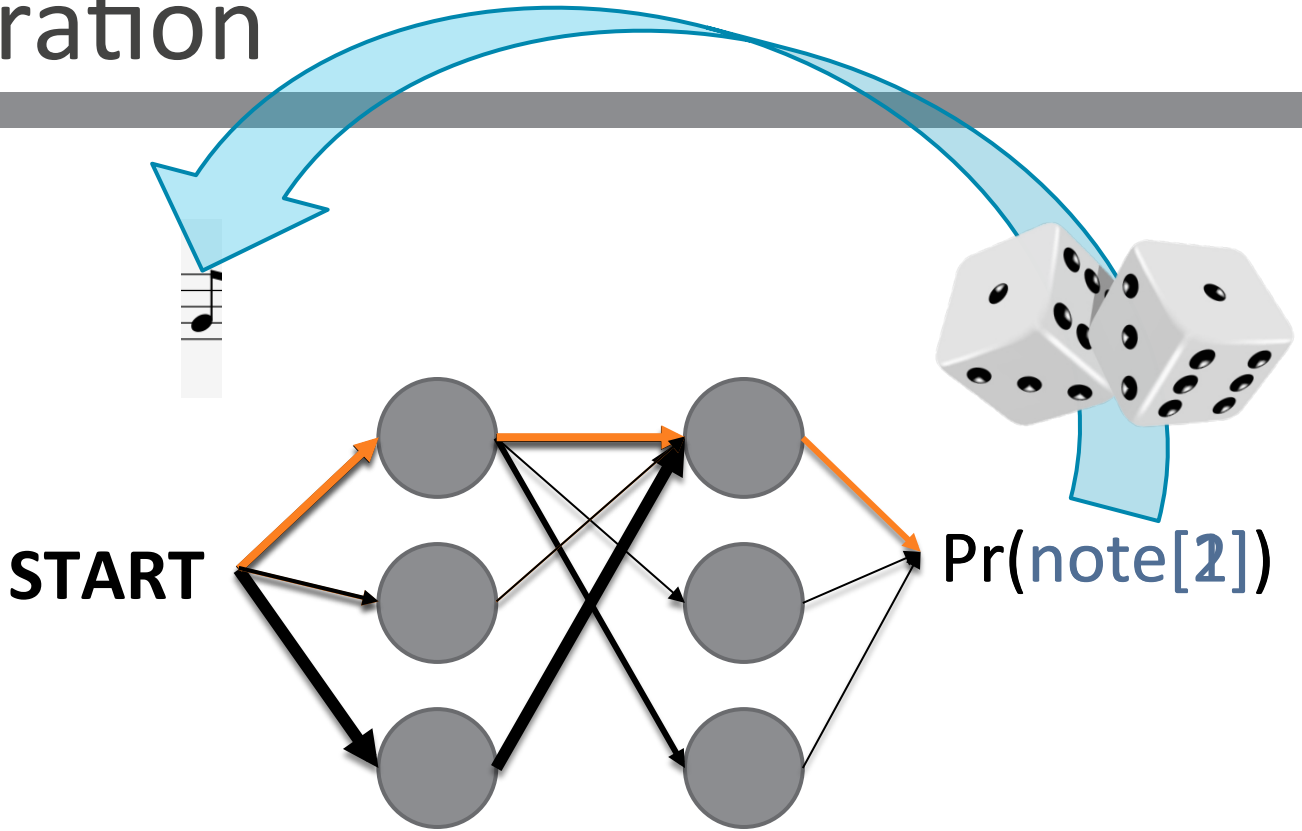


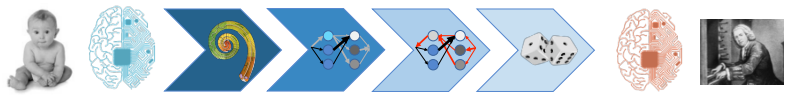
Generation



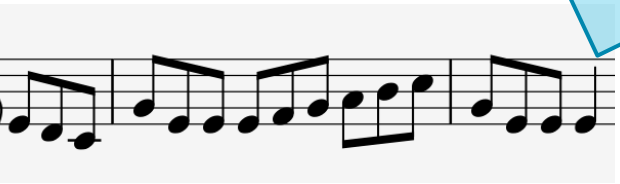


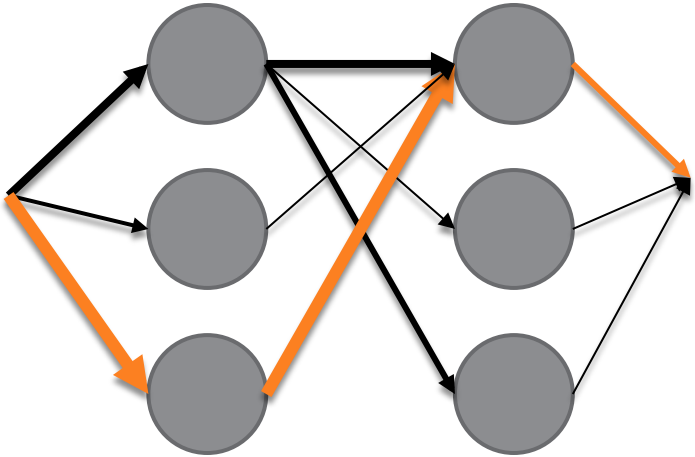
Generation

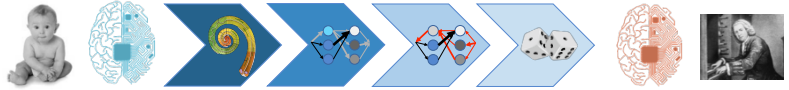




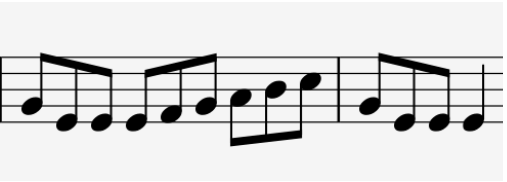
Generation



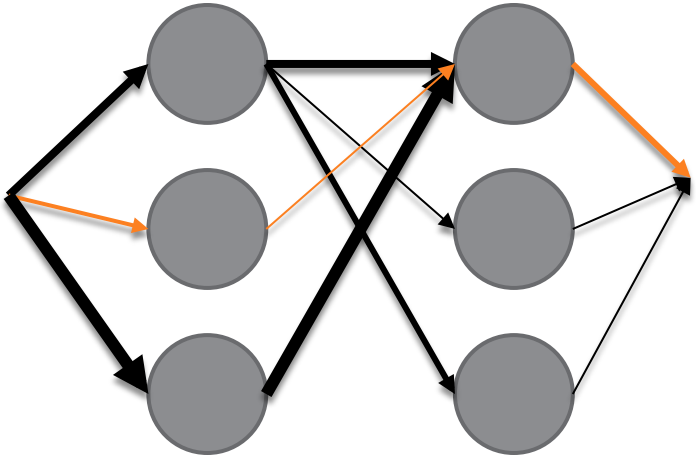




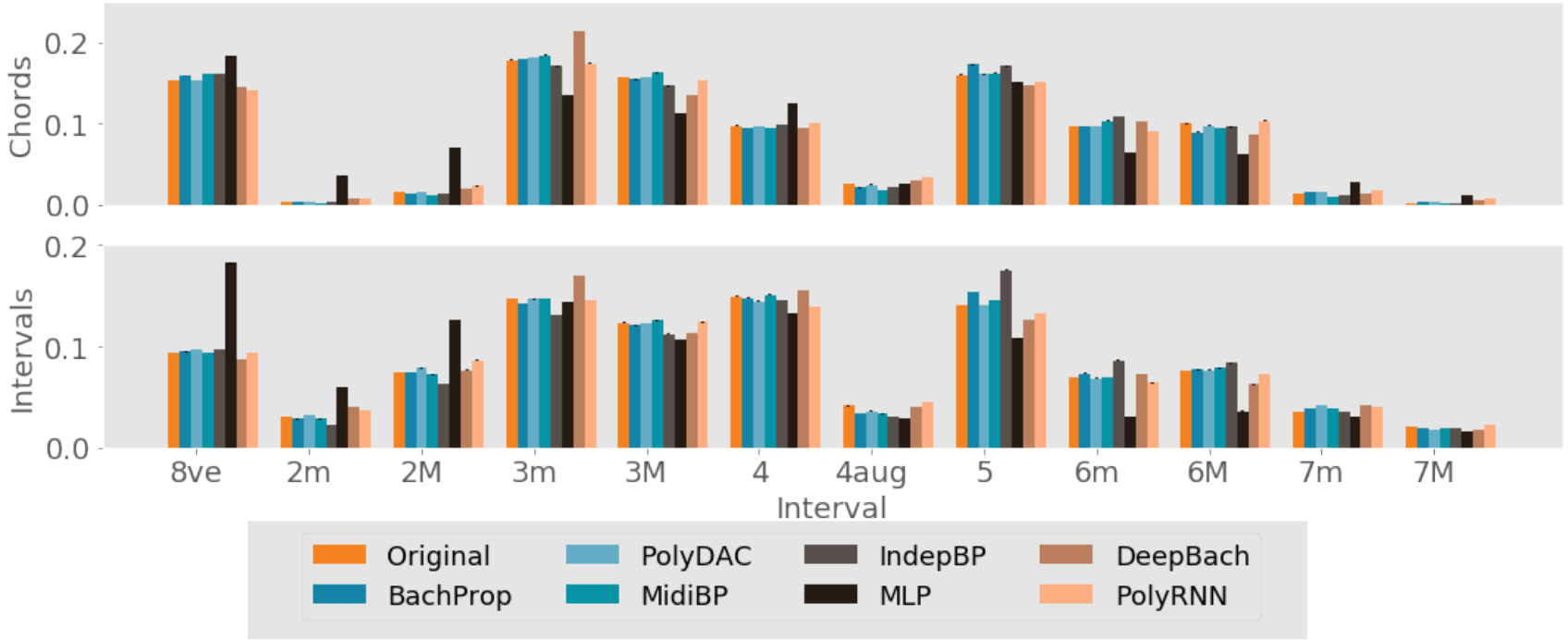
Generation



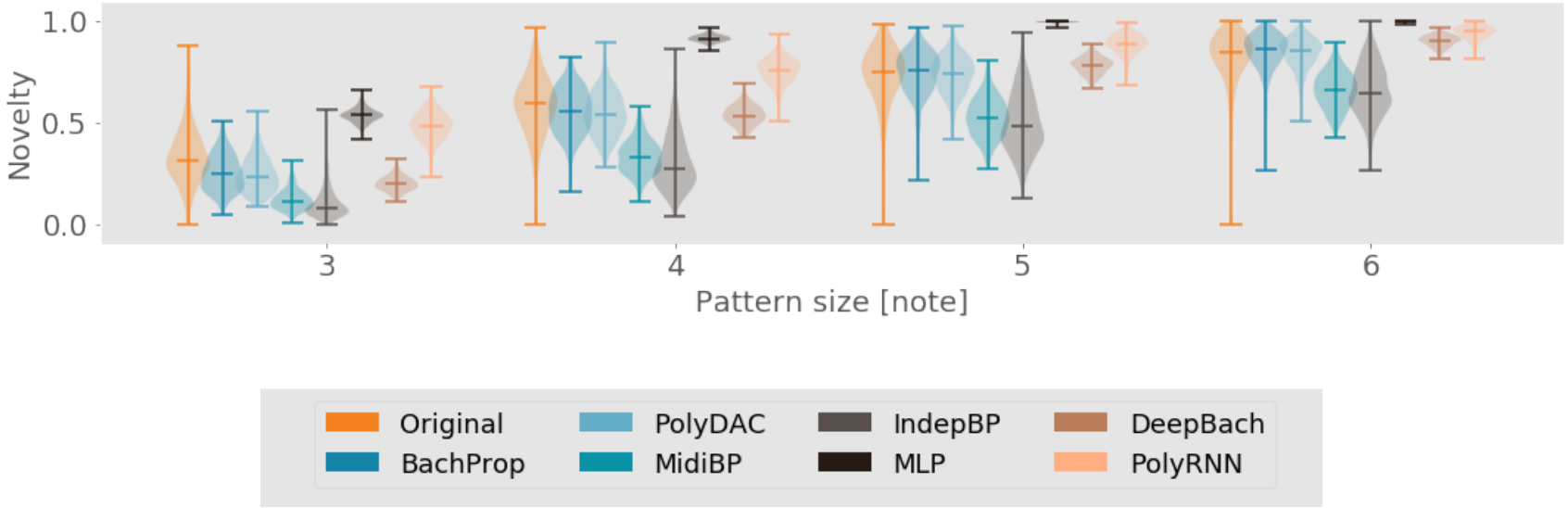
END



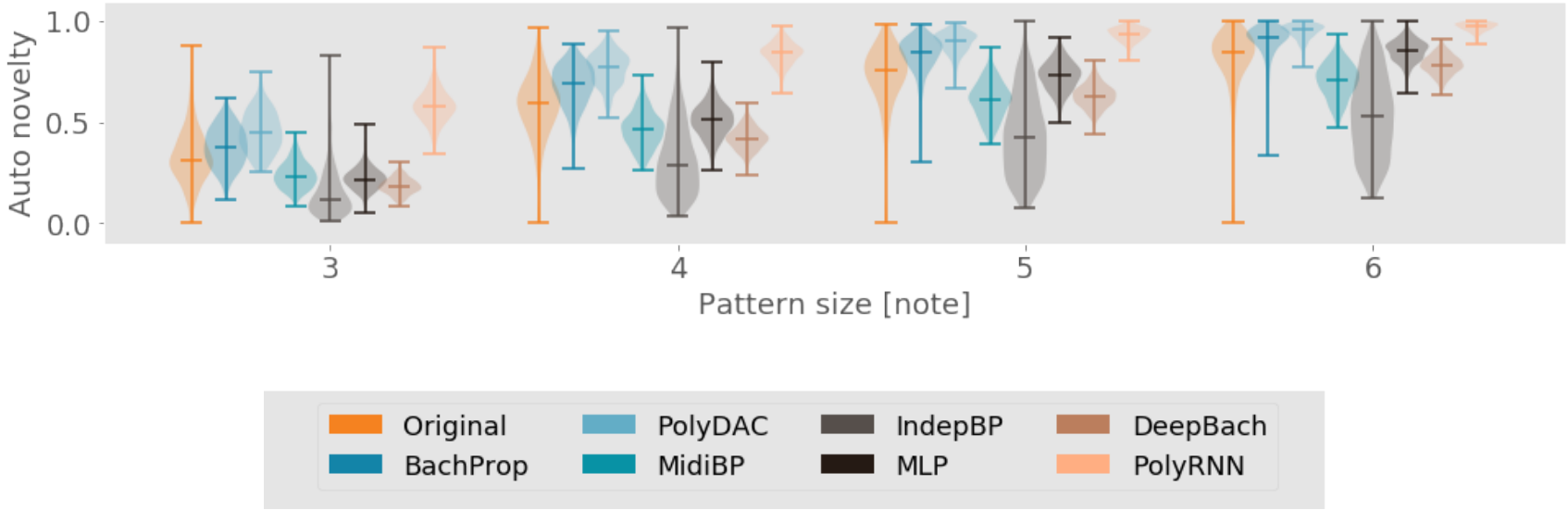
Evaluation | Intervals



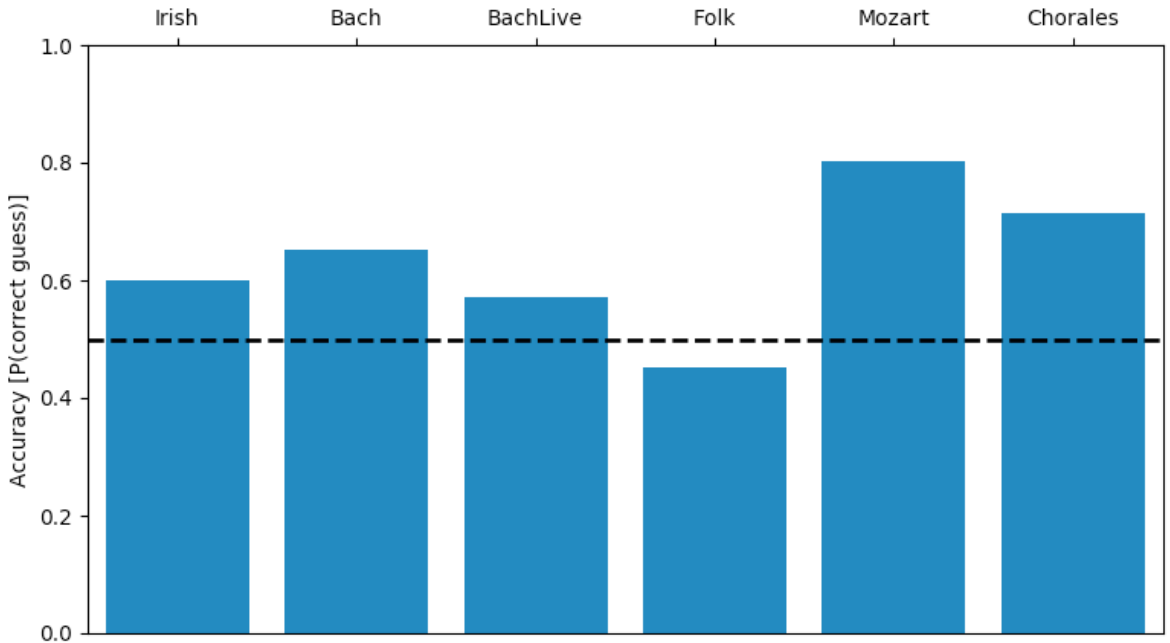
Evaluation | Originality



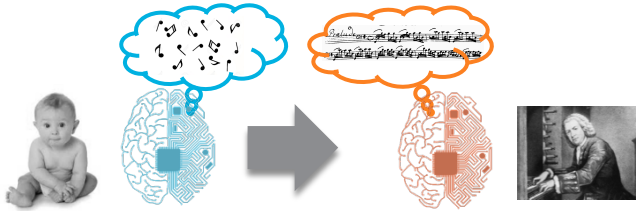
Evaluation | Diversity



Evaluation | Human VS Machine



Generated corpora

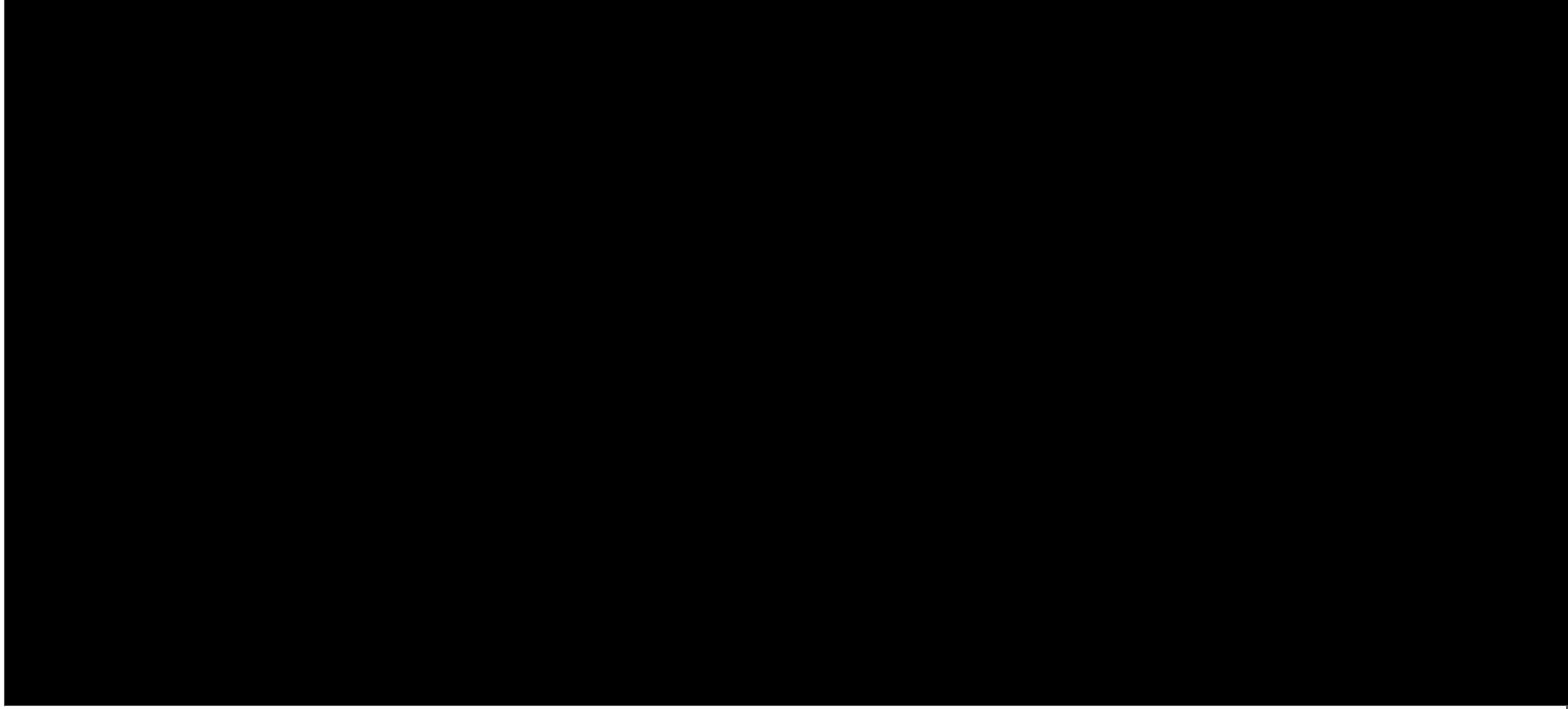


https://sites.google.com/view/bachprop/

https://www.youtube.com/watch?v=c_gj3VV0VbY (uneven numbers from BachProp)

Learning to Generate Music with BachProp

Florian Colombo, Johanni Brea, Wulfram Gerstner • Published in ArXiv 2018



Live interpretation of generated scores

AI Quartet [Ken Lila Ashanti, Xavier Mettan, Thomas Simonet, Florian Colombo]

Scores arranged for string quartet by Isabelle Sierro

Thanks for your attention!

Wulfram Gerstner

Johanni Brea
Samuel P. Muscinelli