

# *A Recurrent Neural Network for Musical Structure Processing and Expectation*

*Tim O'Brien e Iran Romá - Stanford University - 2016*

**Disciplina:** Aprendizagem profunda e geração de música

**Mestranda:** Lívia Gouvêa

# Introdução



Semelhante à **linguagem**, a **música** é um elemento universal da condição humana.



Devido à sua **natureza organizada**, todos os seres humanos expostos à música aprendem implicitamente as **dependências hierárquicas** que conectam os eventos na música.

Pesquisas em neurociência mostra que **todos os seres humanos** são capazes de **identificar** eventos que violem estruturas comuns na música.

O trabalho tem como objetivo **desenvolver um modelo que explique** como os **humanos** (redes neurais humanas) **reagem a eventos inesperados** em músicas.

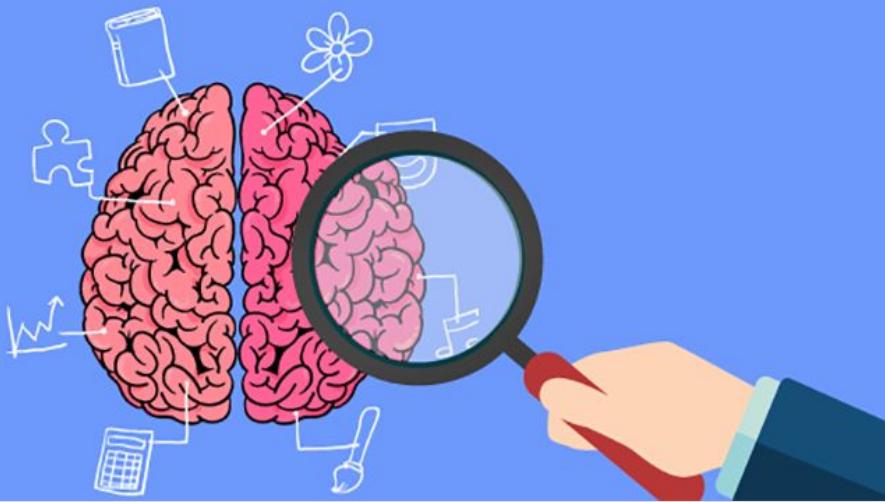
Uma rede neural recorrente (RNN) pode ser utilizada para realizar tal tarefa pois tem a capacidade de **desenvolver expectativas baseadas em eventos anteriores**.

Tal abordagem é semelhante aos **modelos de processamento de linguagem natural**.

Após o **treinamento**, tal modelo pode **reconhecer eventos inesperados** em uma sequência musical de maneira **semelhante aos humanos**.

Além disso, a rede pode **gerar seqüências musicais** após o treinamento, demonstrando as abstrações que aprendeu com os dados que foi treinada.

# Embasamento

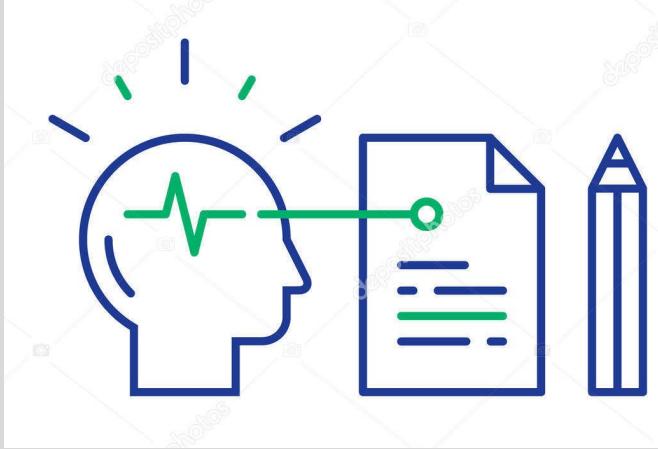


A neurociência cognitiva fornece evidências sobre como os seres humanos processam sequências de eventos ao longo do tempo

Esta ciência mostra também como **preparação sensorial e cognitiva** levam à construção de **representações mentais** de eventos musicais

E como o **histórico** de vida de um indivíduo **afeta sua expectativa** ao ouvir uma sequência de eventos musicais

Inspirados por essa evidência, **Berger e Gang\*** desenvolveram um **modelo** computacional composto por redes neurais recorrentes, capazes de **analisar sequências musicais e desenvolver expectativas** após o treinamento.



Considerando as semelhanças estruturais entre música e linguagem, o trabalho se baseou na **arquitetura do modelo** proposto por Berger e Gang para construir um **modelo de linguagem treinado em estruturas musicais** para mostrar **maior perplexidade** quando estimulado com **eventos inesperados** que violam as regras sintáticas do fluxo musical.

\* J. Berger and D. Gang, “A real time model of the formulation and realization of musical expectations.” 2010.

# Abordagem



O modelo foi treinado com um **corpus de seqüências de acordes**  
(Todos os corais de quatro vozes de Johann Sebastian Bach)

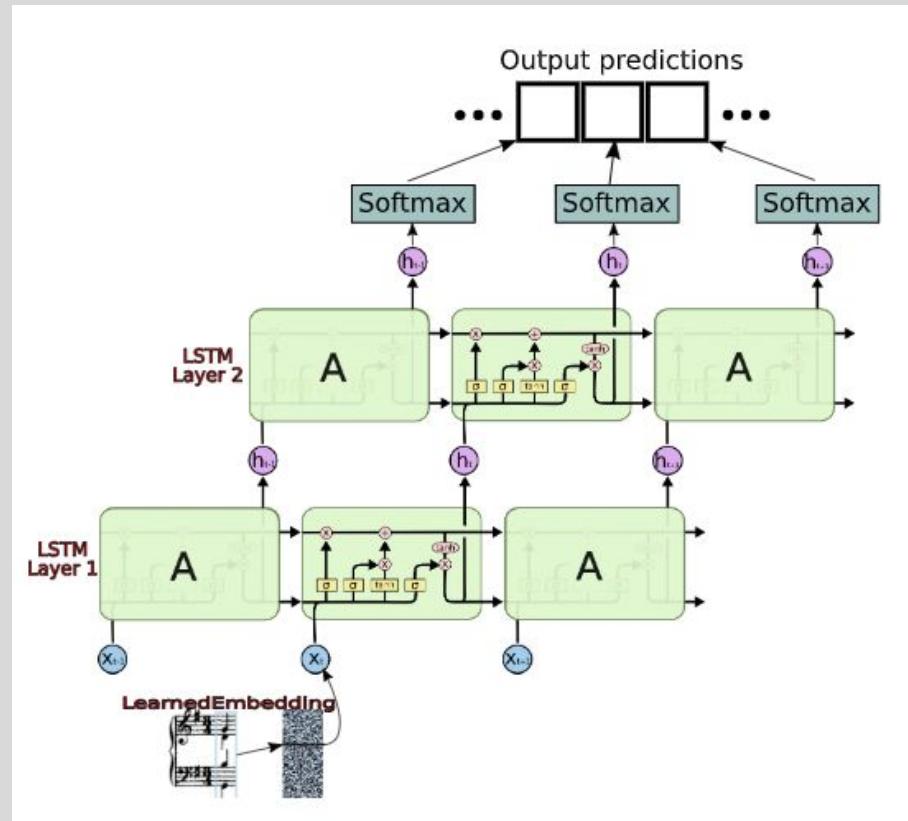
Os acorde são estruturalmente **análogas à linguagem**, já que os acordes são **eventos discretos**, semelhantes às palavras, e seqüências de acordes formam frases musicais, semelhantes às frases.

Assim, as **dependências entre acordes** simultâneos no conjunto de dados serão capturadas pelos **pesos nos módulos ocultos** da arquitetura de rede neural do modelo.

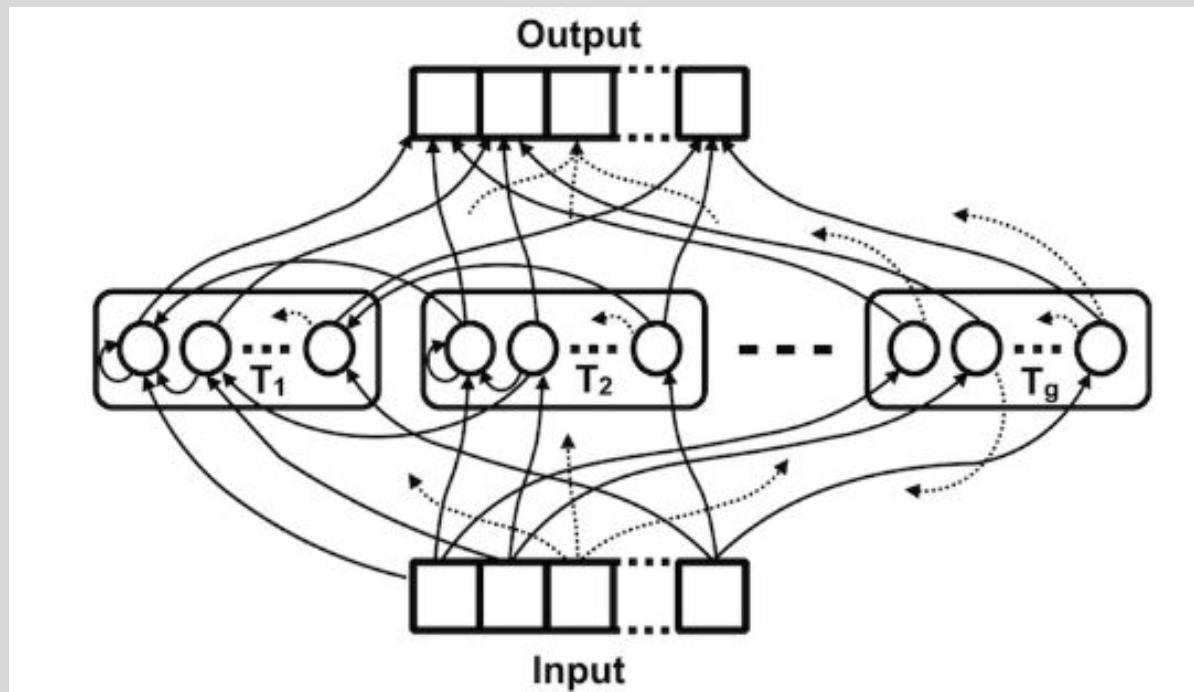
# Arquiteturas de Redes Neurais

Os experimentos se concentraram em dois tipos de redes neurais recorrentes:

**LSTM:** é um tipo de **rede neural recorrente** utilizada no campo da **aprendizagem profunda**. Ao contrário das redes neurais feedforward padrão, possui conexões de **realimentação**.



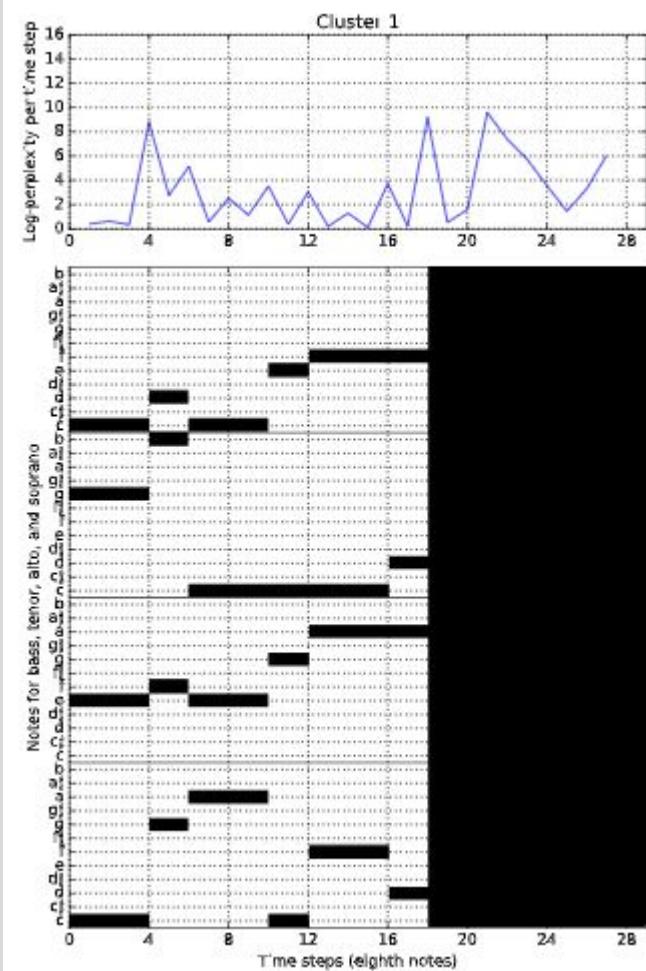
**Clockwork RNN:** É uma **RNN** padrão, mas possui **camadas ocultas** cujas matrizes de peso são atualizadas em diferentes etapas de tempo.



# Experimentos e Resultados

O primeiro teste mostra a **perplexidade** por **tempo** em uma progressão que se desenvolve normalmente até uma **cacofonia abrupta** (isso é como descansar o braço em todas as teclas de um piano) de tempos em tempos.

Como esperado, **observa-se maior perplexidade** nessa região da progressão. Curiosamente, a **perplexidade desaparece rapidamente, volta a subir e depois diminui mais lentamente**. No entanto, **esperava-se um aumento muito maior na perplexidade**, correspondendo à natureza surpreendentemente chocante do evento musical representado.

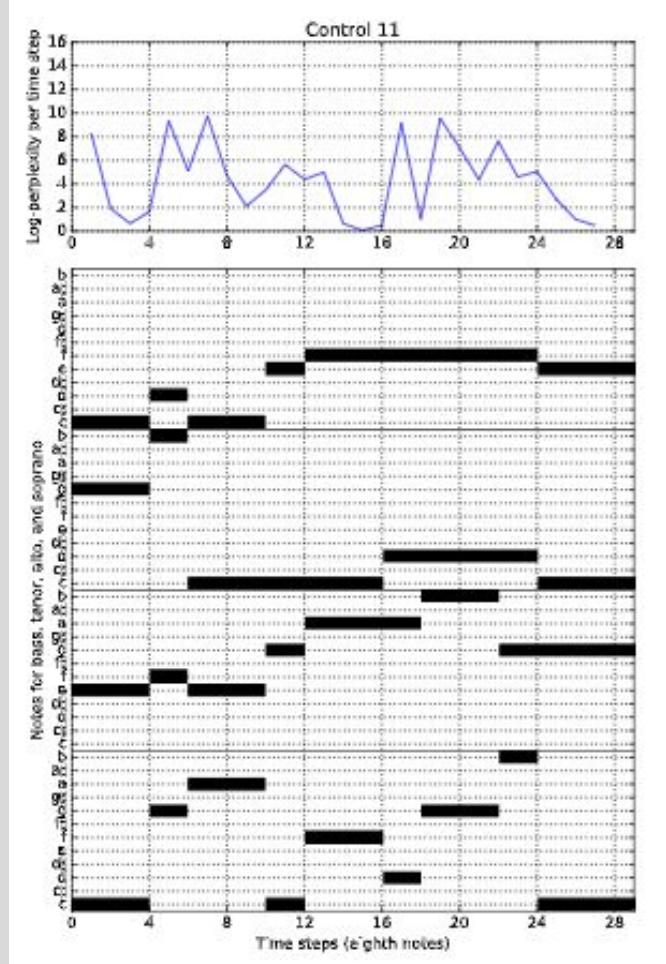


O segundo teste mostra uma progressão que segue todas as regras de harmonia tonal.

Esperava-se ver um mínimo de perplexidade. No entanto, observou-se alguns picos de perplexidade semelhantes ao exemplo anterior.

Porém, um nível mais alto de perplexidade no modelo é aceitável pois o terceiro acorde é um Lá menor, servindo como uma **cadência enganosa** (já que vem depois de Dó maior e depois do dominante G).

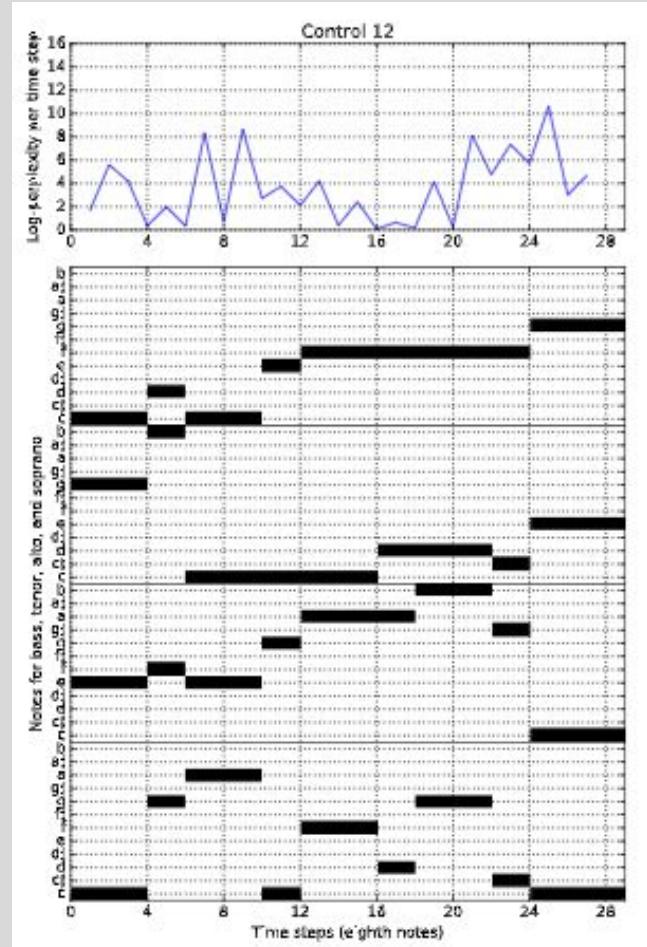
Apesar de ocorrer no conjunto de treinamento do modelo, a cadência enganosa continua a subverter a expectativa e surpreender.



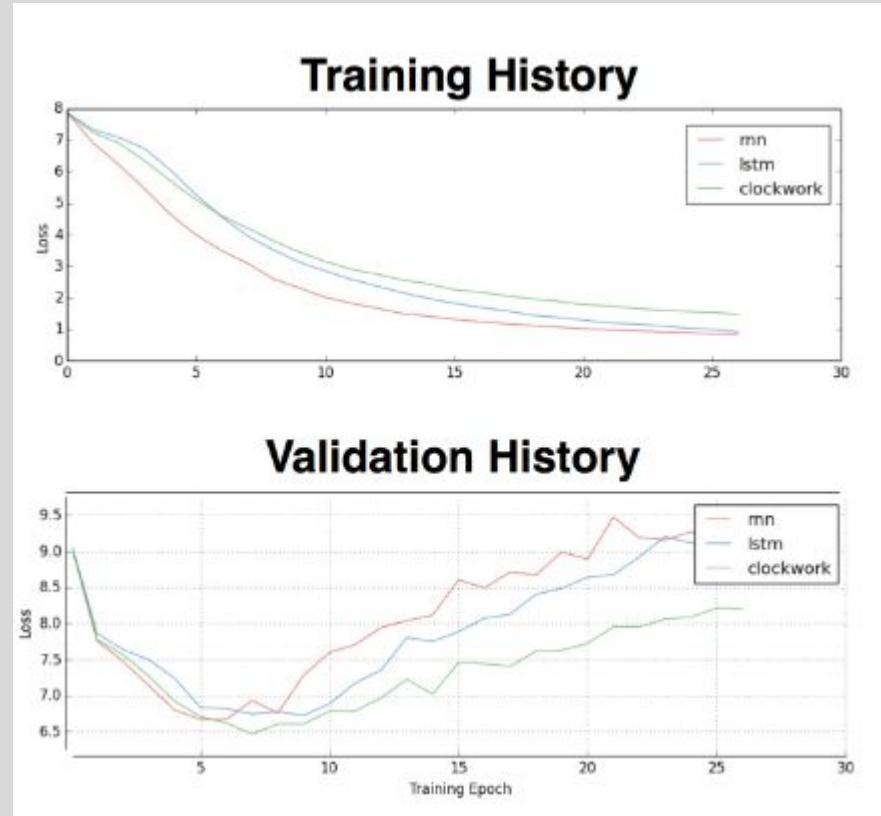
O terceiro teste mostra uma progressão harmônica perfeitamente fina, exceto para o penúltimo acorde.

Esse acorde, conhecido como napolitano, raramente era usado por Bach em seus corais.

Assim, um nível mais alto ou perplexidade em torno deste evento já era esperado.



**Três modelos** diferentes foram treinados: um **RNN**, um **LSTM** e um **clockwork RNN** para avaliar seus **desempenhos** no conjunto de dados. Após o treinamento, todos os modelos **foram capazes** de gerar progressões **de acordes**, que variaram em características estruturais.



Todos os modelos foram capazes de sobrepor os dados, mas **o clockwork RNN foi o que teve menor tendência a fazê-lo.**

Isso porque, arquitetura clockwork RNN é projetada para enxergar além das dependências de curto prazo, que geralmente são aquelas que um modelo RNN padrão captura.

No clockwork RNN, as **conexões de camadas ocultas** com períodos de tempo mais lentos para aqueles com períodos mais curtos, permitem **que ele capture relacionamentos em escalas de tempo maiores**

Finalmente, fize-se o cwRNN gerar música depois que foi treinado.

Observou-se que **modulou entre chaves diferentes e fez uso de harmonia sofisticada**.

Estas características não foram observadas quando um coral foi gerado pelo RNN ordinário treinado no mesmo conjunto de dados, como mostrado a seguir:



Na progressão de acordes gerada por um RNN após o treino observou-se que a **harmonia é incoerente** e parece mover-se pseudo-aleatoriamente.

Não há uma chave clara e o estilo e as regras da teoria musical não são seguidos.



# Conclusão

Redes recorrentes **LSTMs** e **Clockwork** são adequadas para modelar seqüências de tempo de dados musicais.

A abordagem do **modelo de linguagem** que foi utilizada para **previsão instantânea de erros** como uma forma de representar de surpresa de um ouvinte humano, foi **eficiente**.

O **clockwork RNN** parece ter um **desempenho melhor** em tarefas musicais, como a **geração de progressões harmônicas** musicalmente sofisticadas.