MUSICA THEORICA 2022, V. 7.2
SCIENTIFIC ARTICLE
10.52930/mt.v7i2.232
Data do recebimento: 29/04/2022
Data da aprovação final: 03/08/2022

harmonIA: modelagem preditiva de sucessões cordais por aprendizagem computacional profunda

harmonIA: Predictive Modelling of Chord Sucessions using Computational Deep Learning

Fernando Rauber Gonçalves
Universidade Federal do Rio Grande do Sul

Resumo: Neste artigo, apresento um *software* prototípico para manipulação e visualização interativa de sucessões de acordes geradas por um modelo preditivo conexionista e discuto criticamente os resultados obtidos. Foram treinados modelos empíricos para tonalidades, compositores e estilos a partir de um *corpus* de 52.203 canções populares cifradas, coletado por *scraping* (extração automatizada) de um site colaborativo de cifras de acordes. Os dados brutos passaram por filtragem e tratamento para obtenção de melhor consistência e eficiência na modelagem. A aprendizagem computacional empregada utilizou Redes Neurais Recorrentes com células LSTM (*Long Short-Term Memory*), arquitetura escolhida pela sua capacidade em modelar dependências e relacionamentos em dados sequenciais. Em busca de indícios do sucesso da modelagem, foram avaliadas a capacidade de generalização dos modelos de tonalidades aprendidos em sucessões harmônicas com implicações funcionais claras – extraídas da música popular brasileira – bem como a capacidade de contextualização a partir dos elementos prévios.

Palavras-chave: geração musical. Modelagem preditiva. Aprendizagem profunda em música. Harmonia tonal. Musicologia empírica.

Abstract: In this article, I present a prototype software using connectionist predictive models for interactive music generation of chord sequences and critically discuss the results. Models for specific keys, composers, and styles were trained on a corpus of 52,203 popular songs scraped from a collaborative chord sheets site. Raw data underwent filtering and treatment to obtain better consistency and efficiency in the modeling. The computational learning employed used Recurrent Neural Networks with Long Short-Term Memory (LSTM) cells, an architecture chosen for its ability to model dependencies and relationships in temporal series (sequential data). To evaluate the predictive modelling, we tested the generalization capacity of the resulting tonality models in harmonic sequences– selected from Brazilian

popular music – with clear functional implications as well as the model's ability to contextualize output based on previous elements.

Keywords: Music generation. Predictive modelling. Deep learning for music. Tonal harmony. Empiric musicology.

* * *

1. Introdução

Neste artigo, apresento um *software* prototípico para modelagem preditiva de sucessões de acordes por aprendizagem computacional (*machine learning*) – um subcampo da inteligência artificial – e avalio criticamente os resultados obtidos. O ímpeto inicial deste projeto de aprendizado[1] foi a curiosidade em *experimentar* e *avaliar* o alcance da aprendizagem profunda (*deep learning*) com redes neurais artificiais na construção de modelos empíricos baseados em dados musicais. Poderiam estes modelos ajudar de alguma forma na compreensão de repertórios diversos, inclusive aqueles menos sistematizados pela literatura teórico-analítica? A modelagem preditiva pode ter relevância para a criação musical assistida ou para pedagogia musical? Que padrões e regularidades são de fato aprendidas pela máquina? Como avaliar os indícios de sucesso ou fracasso nesta modelagem? Qual o papel do teorista-analista nessa avaliação?

Para investigar algumas destas questões, foram coletadas sucessões de acordes musicais em canções populares para modelagem preditiva com redes neurais. Esta modelagem resultou no software *harmonIA*, um ambiente de geração musical assistida que possibilita ao usuário[2] percorrer – interativa e musicalmente[3] – as continuações sugeridas para qualquer sequência de acordes em diferentes contextos preditivos. A modelagem em *harmonIA* tenta lidar com

---

[1] Designo este projeto como de aprendizado por se tratar de um primeiro *experimento* do pesquisador com a aprendizagem computacional. Escrito sob a perspectiva empática de um aprendiz, este artigo compartilha os resultados atingidos e introduz conceitos desse campo em nível introdutório, mirando a promoção da literacia digital como componente fundamental da prática acadêmica em música.

[2] Como usuário final, teve-se em vista o músico ou pesquisador em música sem conhecimento técnico aprofundado nas Tecnologias da Informação e Comunicação (TICs).

[3] Há uma geração sonora rudimentar que fornece uma contextualização musical mínima do processo.

o seguinte problema: dada uma sequência qualquer de acordes como entrada, quais as continuações mais prováveis em diferentes contextos musicais? Estes contextos musicais modelados foram, especificamente, diferentes tonalidades, a obra de determinados artistas/compositores e certos estilos musicais[4] na música popular.

Mais do que obter uma previsão *correta*, o objetivo desta modelagem é fornecer um resultado ao menos contextualmente *plausível* para a geração musical assistida e, potencialmente, instigar novas perspectivas sobre os processos harmônicos em música[5]. Programar estes modelos preditivos por um sistema fechado de regras (*rule-based*) seria uma tarefa demasiado complexa ou mesmo impossível, pois cada contexto musical – como estilos específicos – demandaria regras distintas. Neste caso, a aprendizagem computacional[6] (*machine learning*) é uma ferramenta especialmente adequada: recorremos a esta abordagem para inferir – a partir de certos conjuntos de dados – regras e comportamentos que não conhecemos plenamente[7] ou complexos demais para serem formalizados.

Enquanto que, em um paradigma computacional imperativo, heurísticas – como regras sobre sucessões harmônicas – são programadas proceduralmente, de forma explícita e sequencial, na aprendizagem computacional por paradigmas conexionistas[8] os comportamentos são modelados a partir de um processo de aprendizado por (auto)*treinamento*, uma aprendizagem computacional que gera

---

[4] Estas categorias contextuais aproveitaram metadados coletados de um site colaborativo de "cifras" musicais.

[5] Este propósito ainda se encontra distante: a complexidade das representações aprendidas por redes neurais resulta em uma baixa *explicabilidade* de seus comportamentos.

[6] Um programa de computador para aprendizagem computacional "aprende pela experiência E, com respeito a algum tipo de tarefa T e medida de performance P, se sua performance P nas tarefas em T, medida por P, melhoram com a experiência E" (Mitchell, 1997 *apud* Goodfelow; Bengio; Courville 2016, p. 99).

[7] "*In most cases, we do not know a priori what the intended behavior of the algorithm is. In fact, the entire point of using machine learning is that it will discover useful behavior that we were not able to specify ourselves*" (Goodfellow; Bengio; Courville, *op. cit*., p. 431).

[8] O termo *conexionismo* designa as abordagens com redes neurais. Mais recentemente, o termo *Deep Learning* (aprendizagem profunda) tem sido utilizado para designar este tipo de abordagem, sinalizando o emprego de redes em múltiplas camadas (Goodfellow; Bengio; Courville, *op. cit*., p. 13–18).

modelos probabilísticos[9]. Treinar um modelo designa, nesta abordagem conexionista, o processo de (auto)calibragem dos parâmetros[10] de uma rede neural a partir de um conjunto de dados (*dataset* ou *corpus*) de treinamento, tendo como objetivo diminuir o *erro*[11] do modelo em produzir os resultados desejados. Dependendo da arquitetura de aprendizado, da representação[12] e do alcance dos dados, diversos comportamentos podem ser modelados[13]. Em *harmonIA*, o objetivo do aprendizado foi criar modelos de geração musical capazes de estimar probabilisticamente a continuação – o próximo elemento simbólico – de uma sequência[14] de acordes. Para tanto, foi empregada uma arquitetura de aprendizado computacional com Redes Neurais Recorrentes com mecanismo LSTM (*Long Short-Term Memory*), solução consolidada para modelagem de elementos sequenciais com dependências de longo prazo[15] (Goodfellow; Bengio; Courville 2016, p. 411).

Uma característica de grande impacto fundamental do aprendizado computacional reside no reaproveitamento da mesma arquitetura de

---

[9] Para uma comparação sobre a geração musical por algoritmos baseada em regas (*rule-based*) e modelos conexionistas com redes neurais, veja Miranda (2002, p. 99–118). Um panorama geral sobre modelos da geração musical pode ser encontrado também em Carvonali e Rodà (2020).

[10] Efetivamente, os pesos sinápticos (*weights*) e vieses (*bias)* de cada célula da rede neural. Uma impactante representação gráfica desde processo de treinamento foi produzida pelo matemático Grant Anderson (disponível em <https://youtu.be/aircAruvnKk>).

[11] O erro do modelo é uma penalidade atribuída ao desvio do resultado pretendido. No processo de treinamento, esta medida indica a necessidade de correção na calibragem do modelo para que melhores previsões sejam realizadas, tendo o conjunto de dados de treinamento como modelo de conduta. Quando um limiar de erro aceitável é atingido, o treinamento costuma ser interrompido.

[12] A representação consiste na natureza e formato dos dados utilizados no treinamento e geração musical (Briot; Hadjeres; Pachet 2020, p. 12).

[13] Veja Briot, Hadjeres e Pachet (2020), Siphocly, El-Horatby e Salem (2021) e Bayle (2019) e para panoramas abrangentes sobre diferentes abordagens para a geração musical com abordagens conexionistas e aprendizado computacional.

[14] O termo "sequência de acordes" empregado neste artigo se refere a um conjunto ordenado de acordes consecutivos, sem qualquer relação com o conceito de sequência harmônica. O termo "sucessão de acordes" é também utilizado com sentido idêntico, enquanto que "progressão harmônica" é um termo evitado pela possível implicação de funcionalidade e direcionalidade tonal.

[15] Ou seja, cujos elementos prévios, mesmo os mais distantes, ainda exerçam influência na probabilidade do elemento subsequente, um fator fundamental na modelagem da linguagem e, analogamente, em eventos musicais sequenciais.

aprendizado para o treinamento de diferentes sistemas de geração musical, modelando estilos musicais diversos (Briot; Hadjeres; Pachet 2020, p. 5). Em *harmonIA*, recortes diversos de um conjunto massivo de dados coletados por *scraping* (extração automatizada) possibilitaram a modelagem de vários contextos musicais. Estes dados brutos consistiram em cerca de 52.203 sequências de acordes (após filtragem preliminar), cada uma representando uma canção popular completa. Destes dados, foram inferidos modelos empíricos para tonalidades, artistas e estilos, bem como variantes destes modelos para fins de aprendizado, explorando técnicas de expansão e normalização dos dados, variações nos hiperparâmetros[16] de treinamento e recortes de comportamentos especializados – por exemplo, sequências compostas por acordes com extensões.

Ainda que o potencial impacto do aprendizado computacional em diversos campos de conhecimento seja amplamente difundido, menos conhecido é o valor da *expertise* sobre determinado domínio de conhecimento na construção e avaliação de modelos preditivos. Para que a máquina "aprenda", é necessário um recorte de dados consistentes e bem-definidos – um *dataset* ou *corpus* – a partir dos quais se possa modelar algum comportamento. Cabe – na melhor das hipóteses – ao *expert* de um campo específico de conhecimento supervisionar a seleção e recorte destes dados, bem como avaliar os resultados obtidos[17] nesta modelagem. Essa avaliação é especialmente desafiadora, tendo em vista que modelos preditivos com redes neurais são frequentemente descritos como sistemas *black box* (Briot; Pacheres; Pachet *op. cit*., p. 218), cujo funcionamento interno e representações aprendidas não podem ser facilmente convertida em regras e condições explícitas por conta do número elevado de parâmetros envolvidos. Por conta dessa dificuldade, Goodfellow, Bengio e Courville (*op. cit*., p. 436–440) enfatizam a importância da observação direta dos resultados obtidos para avaliar se medidas quantitativas como o *erro* preditivo obtido ao final do treinamento de fato refletem o sucesso do modelo obtido. Esta necessidade de

[16] Hiperparâmetros definem a estrutura da rede (como a quantidade camadas e de células) ou controlam aspectos do treinamento (como a taxa de aprendizado) (Goodfellow; Bengio; Courville, *op. cit*., p. 120). Diferente de tarefas convencionais em classificação e previsão, que possuem métricas objetivas de avaliação, o ajuste dos hiperparâmetros para sistemas de geração musical é de natureza *qualitativa*, a partir da avaliação subjetiva dos resultados (*ibid*., p. 82)

[17] Por exemplo, em *BachBot* (Liang 2016), a parceria do desenvolvedor (com conhecimentos musicais bastante limitados) com um musicólogo resultou em uma avaliação preliminar do possível significado de certos padrões de ativação da rede neural.

observação direta foi um dos aspectos que motivaram a criação de uma interface de geração musical assistida em *harmonIA*, propiciando um mecanismo de controle e avaliação através da visualização e escuta da geração musical resultante em tempo real.

O *software* vem sendo desenvolvido desde janeiro de 2018, com uma primeira versão pública lançada em novembro de 2021, em sessão de comunicação do *IV Congresso da Associação Brasileira de Teoria e Análise Musical* (TeMA). Desde o princípio, o esforço de desenvolvimento caminhou em duas direções paralelas: 1) coleta de dados, organização de *datasets* e treinamento de modelos preditivos e 2) construção da interface ao usuário com ferramentas de visualização e manipulação. Estas duas direções se retroalimentaram ao longo do desenvolvimento: o avanço das ferramentas de visualização e da interface de criação musical assistida motivou novas organizações e experimentos com os dados coletados, em sucessivos ciclos iterativos.

A seguir, detalho alguns aspectos deste projeto experimental: 1) a escolha da arquitetura de aprendizagem computacional; 2) a construção da interface ao usuário; 3) a coleta, limpeza e tratamento dos dados; 4) a avaliação dos modelos preditivos. Como conclusão, compartilho considerações gerais e delineio algumas propostas futuras.

2. A arquitetura de aprendizagem computacional

Um aspecto crítico para geração musical por modelos preditivos reside na escolha de uma arquitetura adequada para a tarefa proposta. Pela sua natureza *sequencial*, eventos musicais são especialmente propícios para modelagem enquanto probabilidades condicionais (Choi; Fazekas; Sandler 2016). Por conta desta propriedade, uma abordagem empregada com bastante sucesso na geração musical por modelos probabilísticos[18] emprega processos baseados em *Cadeias de Markov*, que modelam matrizes das probabilidades de transição[19] com base nos eventos passados. Por mais que obtenham resultados satisfatórios em certos

[18] Veja o capítulo "*Probabilities, grammars and automata*" de Miranda (*op. cit., p.* 61–82) para uma introdução sobre a geração musical por métodos probabilísticos, incluindo processos *markovianos*.

[19] Para um exemplo recente de modelagem teórico-analítica com Cadeias de Markov e uma explicação sobre este processo, veja Miccolis *et al*. (2022).

contextos, cadeias markovianas apresentam potencial limitado de generalização[20] na geração musical e, na previsão de elementos baseados em séries longas, tendem ao *plagiarismo* (Briot; Hadjeres; Pachet, *op. cit.*, p. 8), reproduzindo os dados originais.

Por conta destas limitações, a geração musical por abordagens probabilísticas tem centralizado seus esforços – a partir da década de 2010 (Ji; Luo; Yang 2020, p. 6) – na aprendizagem profunda (*deep learning*) com redes neurais artificiais[21]. O aprendizado dito *profundo* se refere às redes com estruturas em múltiplas camadas[22], capazes de processar múltiplos níveis hierárquicos de abstrações, inferidas dos dados de treinamento. Briot, Hadjeres e Pachet (*op.* cit., p. 8) elencam algumas vantagens da aprendizagem profunda sobre a geração com Cadeias de Markov: a) possibilidade de captura de tipos variados de relações, contextos e regularidades; b) aprendizado de dependências de longo prazo e de nível mais elevado; c) maior poder de generalização através do uso de representação distribuída[23]. Há, no entanto, alguns inconvenientes para este tipo de abordagem, como a maior complexidade de implementação e o elevado número de exemplos de treinamento necessários[24] para que uma boa capacidade de generalização seja alcançada.

Em *harmonIA*, optou-se por empregar uma arquitetura de aprendizagem com Rede Neural Recorrente e unidades LSTM (*Long Short-Term Memory*), também designada mais concisamente como "LSTM *Networks*". A arquitetura

---

[20] A generalização designa a capacidade de funcionamento do modelo em dados desconhecidos, que não fizeram parte do processo de treinamento (Briot; Hadjeres; Pachet, *op. cit.*, p. 79).

[21] Veja Goodfellow, Bengio e Courville (2016) para uma explicação aprofundada, Miranda (*op. cit.*) para uma introdução enfocando a geração musical e Briot, Hadjeres e Pachet (*op. cit.*) para uma referência mais aprofundada de métodos e procedimentos específicos à geração musical.

[22] Composta tradicionalmente por camadas de entrada (*input layer*), saída (*output layer*) e um número variável de camadas intermediárias ocultas (*hidden layers*). Estas camadas compõe uma rede de funções aninhadas (*nested functions*) que buscam aproximar um determinado comportamento não-linear.

[23] A representação distribuída é uma propriedade fundamental das arquiteturas de aprendizado profundo em múltiplas camadas. Nestas estruturas, as camadas intermediarias aprendem representações compartilhadas por diversos elementos, de tal forma que uma entrada no sistema seja representada por muitas características, e cada característica esteja envolvida na representação de diversas entradas possíveis (Goodfellow; Bengio; Courville, *op. cit.*, p. 17).

[24] Esta propriedade será discutida na avaliação dos modelos preditivos treinados, na qual ficou evidente a superioridade dos modelos treinados com *datasets* mais abrangentes.

LSTM se consolidou como solução para a modelagem de informações sequenciais[25] pela sua capacidade de aprender dependências de longo prazo em séries temporais. O termo dependências no contexto da modelagem de informações sequencias designa os relacionamentos entre os diferentes elementos ao longo do tempo. Por exemplo, a previsão de continuações probabilisticamente plausíveis para a sequência textual "Os legumes frescos foram comprados por Tia Amélia na" envolve a modelagem de um contexto complexo, no qual as informações mais relevantes – com resultados implicativos menos entrópicos – estão dispersas ao longo do tempo. Na proposição original da arquitetura de aprendizado LSTM por Hochreiter e Schmidhuber (1997), esta capacidade está bem representada nos Experimento 6a/6b, que demonstram a identificação de regularidades simbólicas de natureza sequencial – mesmo quando separadas por dezenas de elementos. Ressaltamos novamente que, para o aprendizado destas implicações, ou *regularidades*, é necessário um *dataset* de treinamento suficientemente robusto, contendo uma quantidade massiva de exemplos para subsidiar a construção de representações adequadas durante o processo de aprendizado computacional.

A arquitetura LSTM tem sido utilizada para geração de melodias (Sturm; Santos; Korshunova 2015), sucessões de acordes e estruturas rítmicas (Choi; Fazekas; Sandler, *op. cit.*), modelagem polifônica (Liang 2016), bem como em diversos outros protótipos (veja a compilação de Briot; Hadjeres; Pachet, *op. cit.* e Bayle 2019) que demonstraram a viabilidade dessa abordagem no aprendizado de diferentes representações sequenciais de dados musicais.

Escolhida a arquitetura de aprendizado propícia, uma busca na plataforma *GitHub* chegou ao código-fonte de uma solução que demonstra a geração textual – com tokenização[26] por palavras – utilizando Redes Neurais Recorrentes com mecanismo LSTM (Kim 2017). Este projeto, escrito com

---

[25] No momento inicial da escrita do software, em 2018, esta era a solução mais consolidada para esta tarefa: *"As of this writing, the most effective sequence models used in practical applications are the called gated RNNs. These include the long short-term memory and networks based on the gated recorrente unit"* (Goodfellow; Bengio; Courville, *op. cit.*, p. 411). Mais recentemente, arquiteturas com mecanismos de *atenção* – veja o artigo "*Atenttion is All You Need"* (Vaswani *et al.* 2017) – apresentam capacidade significativamente mais elevada de generalização em séries temporais mais longas.

[26] A tokenização, neste contexto, se refere à unidade mínima de representação simbólica dos elementos no processo de treinamento e geração.

propósitos didáticos pelo professor de ciências da computação Sung Kim, foi utilizado como ponto de partida – por bifurcação (*fork*) – para o desenvolvimento de *harmonIA*. Evitar uma implementação do zero, partindo de uma base de código pré-existente com licença permissiva, possibilitou que o pesquisador se concentrasse, desde o início, na coleta e organização de *datasets* para treinamento e, posteriormente, na interface de criação musical assistida para avaliação dos resultados, tarefas de maior interesse para a especialidade do autor enquanto *usuário final* e *treinador*[27] de um modelo de aprendizado computacional como ferramenta em seu domínio específico de conhecimento – a teoria e análise musical. Diversos critérios subsidiaram a escolha dessa solução em específico como ponto de partida: *a)* licença permissiva MIT, permitindo a modificação e redistribuição livre; *b)* uso da biblioteca *TensorFlow*, que não só propicia ajustes na arquitetura a partir de níveis de abstração mais elevados como também dispõe de documentação extensiva e comunidade ativa de desenvolvedores; *c)* uso da linguagem *Python*, com a qual o autor possuía familiaridade prévia; *d)* tokenização por palavras, de fácil adaptação para a natureza da representação simbólica pretendida, conforme discussão na seção 3 deste artigo e na nota de rodapé n° 37; e) código-fonte conciso e bem documentado.

3. *Datasets*: coleta, representação e tratamento dos dados

Para o treinamento de uma rede neural, é necessário organizar um conjunto (massivo) de exemplos representativos – um *dataset* ou *corpus* de treinamento. É a partir destes exemplos que o algoritmo de treinamento calibrará os pesos e vieses da rede neural, até chegar a uma configuração da rede que apresente um *erro* preditivo aceitável. Estes exemplos não só devem ser numerosos o suficiente como devem representar um comportamento

[27] Strobelt *et al*. (2018) apontam três categorias de interessados na visualização de redes neurais profundas: *arquitetos*, *treinadores* e *usuários finais*, cada qual enfocando dimensões e níveis de abstrações distintas. Ao *usuário final* pouco importam os detalhes do funcionamento interno – a *caixa-preta* – mas tão somente os resultados obtidos. *Treinadores* constroem modelos e refinam os hiperparâmetros, sem um conhecimento aprofundado de cada componente, enquanto que *arquitetos* dominam o funcionamento interno de cada unidade da rede e projetam suas interações em uma determinada arquitetura de aprendizado.

razoavelmente consistente[28] para que a modelagem possa inferir regularidades significativas presentes neste conjunto de dados. Portanto, a obtenção de dados em quantidade satisfatória para o aprendizado com redes neurais é um dos desafios envolvidos neste tipo de modelagem.

Em *harmonIA*, optou-se pela coleta de dados a partir de uma fonte inédita na literatura[29] – um site colaborativo de cifras brasileiro[30] –, tarefa para a qual foram construídas ferramentas específicas[31] de extração automatizada (*web scraping*). O material bruto obtido consistiu em sequências de acordes para 82.600 músicas. Foram coletados também metadados disponíveis para cada música, tais como classificações por artista, estilo musical e tonalidade[32]. A partir destes dados e metadados, foram treinadas três categorias de modelos preditivos:

[28]Ainda assim, uma quantidade considerável de ruído e heterogeneidade em datasets massivos é tolerável (ou até mesmo desejável) em virtude da modelagem de dependências prévias, aspecto que será abordado na avaliação dos modelos preditivos.

[29] Briot, Hadjeres e Pachet (*op. cit*., p. 48–49) ressaltam o número limitado de *datasets* de referência para a geração musical por *Deep Learning* e apresentam uma compilação das principais fontes de dados utilizadas em projetos neste campo. Ji, Luo e Yang (2020) apresentam uma compilação com propósito similar. Em *harmonIA*, o uso de uma fonte inédita também satisfez o propósito de aprendizado do projeto, instigando o autor a se aprofundar nos procedimentos de coleta automatizada e tratamento de dados brutos.

[30] <https://www.cifraclub.com.br>

[31] O código-fonte destas ferramentas bem como uma descrição sucinta da função de cada *script* pode ser encontrado em: <https://github.com/frauber84/harmonIA/tree/main/utils>

[32]A classificação das tonalidades nos metadados coletados apresentou elevado índice de erros. Alguns mecanismos simples de filtragem foram desenvolvidos, descartando sequências com alta probabilidade de erro na classificação da tonalidade. Ainda assim, persistiram nos *datasets* construídos por tonalidade um índice significativo de erro, cujo impacto aparentemente foi minimizado pela contextualização das dependências prévias na modelagem.

a) tonalidades maiores e menores; b) sucessões de acordes nas músicas de artistas específicos[33] e c) estilos musicais específicos[34].

Antes do treinamento, os dados passaram por filtragem preliminar – descartando exemplos fora de critérios mínimos[35] – e processamento para uniformização das cifras. Esta uniformização consistiu no agrupamento de diversas variantes de cifragem em representações simbólicas unificadas. Como resultado, o vocabulário global (conjunto das cifras distintas) foi reduzido pela metade, otimizando a modelagem ao evitar a dispersão de símbolos de significado idêntico em representações distintas. A Fig. 1 apresenta um excerto[36] deste tratamento para uniformização, demonstrando seis (de um total de cerca de 460) chaves de substituição textual para propósitos de unificação da cifragem. Após esses tratamentos e filtragens, chegou-se a um *dataset* global de 52.303 músicas distintas, com 4.500.000 elementos individuas (acordes) e um vocabulário de 6.500 acordes distintos.

A Fig. 2 mostra dois exemplos de sequências coletadas, representando a sucessão de acordes em duas canções populares completas. Na construção do *dataset* de treinamento, estas sequências foram tokenizadas (separadas) no nível do acorde[37], e cada acorde foi indexado a partir do vocabulário global, composto

---

[33] Esta modelagem foi problemática pelo baixo número de exemplos disponíveis para cada artista e pela inconsistência dos dados, demasiado heterogêneos pela natureza colaborativa do site. Os dados coletados a partir das obras de determinados artistas não conseguiram produzir generalizações satisfatórias com a abordagem de aprendizado utilizada, mesmo após o emprego de técnicas de ampliação de dados (transposição para os 12 tons e normalização para tonalidade única). Neste caso, métodos como as Cadeias de Markov provavelmente seriam mais efetivos para orientar a modelagem.

[34] Também aqui os resultados foram geralmente inconsistentes. Uma das exceções foi o modelo do estilo *samba*, cujo comportamento consistentemente funcional da harmonia resultou em uma modelagem mais satisfatória.

[35] Alguns critérios de exclusão: quantidade insuficiente de acordes, quantidade insuficiente de acordes distintos, músicas repetidas e classificações de tonalidades com alta probabilidade de erro.

[36] O *script* completo pode ser acessado em <https://github.com/frauber84/harmonIA/blob/main/utils/cifras_helper_batch.py>.

[37] Seria possível separar (tokenizar) a sequência textual no nível do caractere, como nos exemplos de Karpathy (2015). Choi, Fazekas e Sandler (2016) compararam a geração de sucessões harmônicas com as duas abordagens em redes LSTM, indicando preferência pela tokenização por acordes. Em *harmonIA*, a tokenização por acorde foi escolhida por representar a mínima unidade de sentido musical pretendida: ainda que a aprendizagem computacional conseguisse elaborar

por todos os acordes distintos de cada modelo. A Fig. 3 demonstra uma representação numérica hipotética resultante da tokenização por acordes como unidade simbólica mínima.

```
def UniformizaCifra(x):
    chaves = { 'sus4/7 ' : '7sus4 ' ,
               '7/4 '    : '7sus4 ' ,
               '7(4) '   : '7sus4 ' ,
               '7/sus4 ' : '7sus4 ' ,
               '7SUS4 '  : '7sus4 ' ,
               # etc ...totalizando cerca de 460 elementos
             }

    # substitui a partir da lista de chaves
    for key in sorted(chaves, key=len, reverse=True):
        x = x.replace(key, chaves[key])
    return x
```

Figura 1: Excerto do *script* para uniformização de cifras

tom=C C Dm7 G7 C6 Am7 Dm7 G7 C6 Em7(b5) A7 Dm7 Dm7(b5) C7M Am7 Dm7 G7 C6 C7M A7 Dm7 G7 Dm7 G7 C6 C7M C7 F7M F#o7 C7M Am7 Dm7 G7 C6 Dm7 G7 C6 Am7 Dm7 G7(9) C6 Em7(b5) A7 Dm7 Fm6 C7M Am7 Dm7 G7 C7M Dm7 G7 C6 Am7 Dm7 G7(9) C6 Em7(b5) A7 Dm7 Fm6 C7M Am7 Dm7 G7 C7M

tom=C C7M Am7 Dm7 G7/E G Em Am7 F G Gm7 C7 F7M G Em Am D7 G7 C7M Am7 Dm7 G Em Am7 F G Gm7 C7 F7M G Em Am D7 G F G C Am7 Dm G Gm7 C7 F G C Am7 Dm7 G7 C

Figura 2: Exemplo de sequências de acordes[38]

uma cadeia complexa de condicionantes para utilização do caractere como unidade morfológica, a representação aprendida não seria benéfica no propósito da geração musical, sendo inclusive potencialmente prejudicial à modelagem das dependências temporais ao exigir a produção de sequências muito mais longas. A tokenização por acordes também garante que todos elementos da camada de saída sejam acordes válidos (presentes no conjunto de treinamento), evitando *garbage output* – neste caso, a concatenação de caracteres que não constituam um símbolo cordal válido.

[38] A classificação "tom=C" no início é um metadado, ignorado durante o treinamento.

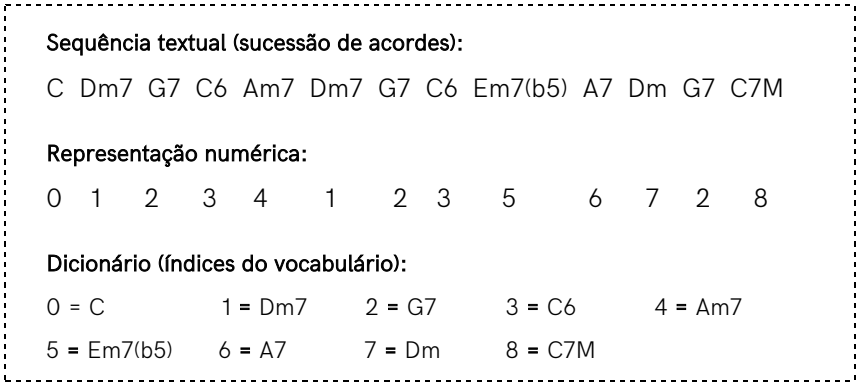

Figura 3: Representação numérica usando o acorde como unidade simbólica

É importante refletir sobre as implicações da representação dos dados para a modelagem preditiva. Algoritmos de aprendizagem computacional são agnósticos em relação aos dados: tudo que puder ser representado como uma variável quantitativa discreta pode ser potencialmente modelado no aprendizado da máquina, através da descoberta de diferentes regularidades e padrões associativos inerentes a esses dados. Desta forma, a escolha da representação para o *dataset* de treinamento é determinante para o resultado final da modelagem obtida.

No caso da representação de acordes, há duas abordagens comuns elencadas por Briot, Hadjeres e Pachet (*op. cit.*):

a) *implícita* e *extensional*, com a enumeração de todas as notas (ou alturas) que compõem o acorde;
b) *explícita e intencional*, com o uso de símbolos (como cifras) para indicar explicitamente a categoria do acorde, mas não seus componentes individuais.

Em *harmonIA*, a escolha por uma representação *intencional explícita* foi determinada pela natureza e limitação dos dados coletados do site colaborativo de cifras. Utilizando cifras de acordes como *tokens* – e unidades simbólicas – para modelagem sequencial, modela-se unicamente a probabilidade de acordes em específico serem o próximo de uma determinada sucessão harmônica – levando em conta todo o contexto de dependências prévias, locais (próximas) ou remotas. Aspectos como o ritmo harmônico ou a condução de vozes não foram incorporados, tendo em vista que estes elementos não fizeram parte da

representação simbólica utilizada[39] e demandariam um conjunto de dados de treinamento obtido de outras fontes. Em todos *datasets*, houve uma intenção explícita de preservar a enarmonia, tendo em vista ser esse um fator essencial para a validação dos resultados da modelagem. Desta forma, as representações das cifras não foram simplificadas em classes de alturas, ou seja, D♭ e C♯ foram consideradas representações distintas para não distorcer os resultados da modelagem das centricidades tonais.

Nos *datasets* com cifras de artistas específicos, foram experimentadas estratégias de ampliação de dados, transpondo todos os exemplos para as 12 tonalidades (como na expansão de dados proposta em Hadjeres, Pachet e Nielsen 2016). Esta estratégia, no entanto, não representou aprimoramento significativo no treinamento destes modelos, cujos resultados inconsistentes atribuímos ao número reduzido de exemplos para treinamento – discussão aprofunda na seção 5 deste artigo. Outro experimento com os *datasets* de artistas específicos consistiu na normalização dos dados, transpondo todas músicas para uma mesma tonalidade (como em Choi 2017) para possibilitar a comparação direta entre os modelos. Novamente, a avaliação neste caso ficou prejudicada pela inconsistência dos resultados obtidos para esta categoria de modelagem.

Ponderamos que é necessário cautela com estratégias de normalização ou ampliação artificial dos dados no contexto da modelagem estilística. As três estratégias descritas anteriormente – 1) simplificação em classes de alturas; 2) transposição para as 12 tonalidades ou gama enarmônica mais completa; 3) normalização em tonalidade única – podem potencialmente resultar em afastamento do contexto original dos dados coletados, tornando a modelagem resultante menos fidedigna. Podemos considerar que há inclusive um traço estilístico relacionado aos aspectos idiomáticos de certos instrumentos musicais, resultando, por exemplo, na priorização de determinadas escolhas de alturas. Um exemplo de distorção estilística óbvia seria a transposição irrestrita (ou normalização para tonalidade única) de sucessões de acordes envolvendo o uso de cordas soltas em instrumentos de cordas como o violão e/ou guitarra, muito prevalentes na música popular como instrumentos de apoio harmônico. Estas reflexões apontam, portanto, para a necessidade de reflexão sobre os

[39] Para comparação, veja a representação dos eventos polifônicos nos corais de J.S. Bach produzida por Liang (*op. cit.*, p. 25–28) em *BachBot*, sistema de geração que também utilizada um algoritmo similar de aprendizagem computacional com redes LSTM.

objetivos da geração e, se necessário, na seleção de dados naturais em quantidade massiva para o treinamento de modelagens estilísticas satisfatórias. Fazemos a ressalva de que, em um contexto de criação musical assistida de natureza mais generalizada, estas estratégias de normalização ou ampliação artificial teriam sua validade, bem como quaisquer outras formas criativas de extrapolação dos dados musicais originais.

O *dataset* completo está disponível para consulta e uso sob a licença permissiva MIT em <https://fernandorauber.com.br/hia/data/>.[40]

4. A interface ao usuário

Como descrito anteriormente, *harmonIA* partiu de um *fork* de uma implementação mínima pré-existente para geração textual com redes LSTM (Kim, *op. cit.*). Como mecanismo de geração, esta implementação utilizava um mecanismo simples de amostragem[41] para gerar continuações para uma sequência textual qualquer. Após as primeiras tentativas de treinamento de modelos preditivos e geração automatizada com cifras de acordes, constatou-se que, para uma avaliação mínima dos produtos gerados, seria necessário criar ferramentas para visualizar o universo das probabilidades consideradas pela máquina e manipular as escolhas em cada iteração do processo preditivo.

Podemos compreender o modelo preditivo por uma rede neural multicamadas como uma cadeia de funções (Goodfellow; Bengio; Courville, *op. cit.*, p. 168) que, a partir de uma *entrada* inicial – nesse caso, qualquer sequência de acordes –, faz processamentos por meio de cadeias intermediárias que produzem, na camada de *saída*, coeficientes de ativações para cada célula individual. Estes coeficientes são então organizados em uma distribuição de probabilidades[42], representando as estimativas sobre o próximo elemento da

---

[40] Neste endereço, estão disponíveis quatro versões para cada conjunto de dados: a) dados brutos; b) dados processados para uniformização das cifras; c) dados expandidos pela transposição em 12 tons; d) dados normalizados pela transposição para tonalidade única (Dó Maior).

[41] Este mecanismo de amostragem (*sampling*) consistia em introduzir algum grau de aleatoriedade nas escolhas do modelo preditivo, evitando completar a sequência sempre com a hipótese mais provável (*greedy picking*).

[42] No caso da implementação utilizada, há uma função de ativação *softmax* na camada de saída que normaliza os índices de ativação em um conjunto probabilidades. Esta função também amplifica resultados prováveis e atenua resultados improváveis, melhorando a relação sinal-

sequência. Para visualizar esta distribuição de forma intuitiva, foi desenvolvida uma interface ao usuário que dispõe os acordes de forma proporcional à sua probabilidade de ativação. Essa interface é interativa: ao clicar em qualquer acorde, um novo acorde é acrescentado e uma nova distribuição de probabilidades é disposta, tornando o software uma espécie de (limitado) sistema de criação assistida, projetado para a experimentação direta e controle dos modelos preditivos. Neste sistema, é possível também alterar acordes prévios ou escutar qualquer um dos acordes[43] antes de sua escolha efetiva como próximo elemento da sucessão harmônica.

A Fig. 4 mostra uma captura da interface central do programa, com as probabilidades de continuação da sequência de acordes "C7M Dm7 G7 C7M Am7" no modelo da tonalidade de Dó Maior. Há quatro componentes fundamentais nesta interface: 1) o painel da esquerda apresenta botões para aumentar ou diminuir o campo de possibilidades (acordes visíveis), escutar acordes, realizar escolha aleatória, salvar (em formato MusicXML), acionar modo multimodelo, criar nova música ou fechar o programa; 2) no campo central, acordes clicáveis representam proporcionalmente as continuações; 3) o painel da direita exibe uma representação em notação musical, com possibilidade de alteração dos acordes ou mudanças em suas posições; 4) no canto inferior direito, há um elemento para seleção do modelo preditivo – agrupado por tonalidade, estilo e artista e outro botão que acessa informações sobre o modelo atualmente escolhido.

ruído na modelagem. É esta distribuição que será exposta visual e auditivamente ao usuário em *harmonIA*.

[43] Com o botão direito do mouse, o usuário alterna entre o modo de escolha e escuta dos acordes. O modo atual é indicado pelo ícone do cursor: um ponteiro para escolha de acordes e um violão estilizado para o modo de escuta.

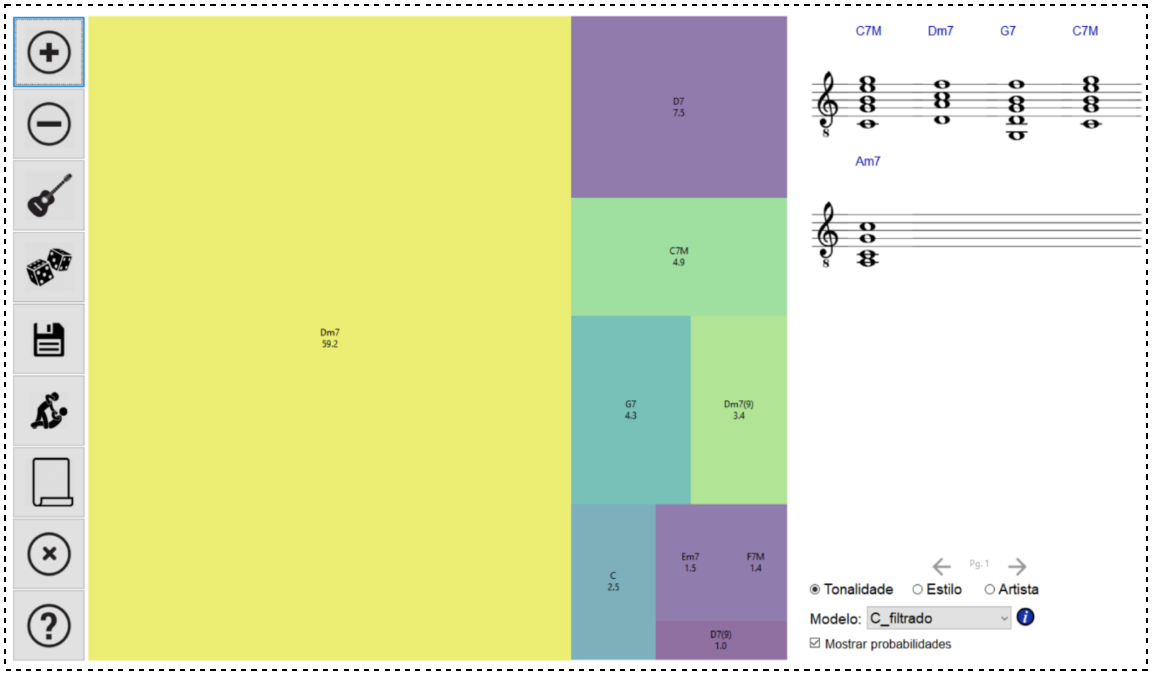

Figura 4: Interface ao usuário

Ao longo do desenvolvimento, os mecanismos de manipulação na interface foram sendo paulatinamente refinados. Os primeiros botões no painel da esquerda (vide Fig. 4) aumentam ou diminuem o campo de busca da previsão, possibilitando a visualização de continuações mais ou menos usuais, um parâmetro de controle fundamental para aumentar a diversidade na geração musical. Para tornar o processo mais aprazível, um mecanismo de geração sonora MIDI arpeja os acordes selecionados a partir de posições usuais no violão – também extraídas de forma automatizada[44]. Estas posições dos acordes e algumas variantes podem ser alternadas clicando em cada acorde[45] na partitura (Fig. 5). Esta funcionalidade limitada[46] de notação musical tem um sentido lúdico,

[44] Esta lista de posições, em notas MIDI, está disponível no arquivo *dic_acordes.py*. Foram compiladas as 4 posições mais usuais para cada cifra de acorde no violão (extraídas de um site de cifras), totalizando cerca de 20.000 posições distintas. As posições podem ser alteradas (entre as escolhas pré-determinadas) pelo usuário clicando em cima do acorde no painel de notação musical.

[45] Ao clicar em um acorde na partitura, sua posição é trocada por outra alternativa de posição usual no violão. O botão direito do mouse executa o acorde na posição escolhida e um clique com o botão direito na cifra troca o acorde.

[46] Não foi encontrada uma biblioteca adequada para notação musical em Python, portanto um *script* em Python foi criado para criar as imagens de todas as posições do dicionário de acordes utilizando o *software* aberto MuseScore como renderizador gráfico.

tendo em vista que as notas individuais que compõe cada acorde não foram um aspecto modelado[47]. Ainda assim, como a modelagem em *harmonIA* se dá no contexto da canção popular, estas posições em alguns casos podem possivelmente representar uma distribuição das notas mais estilisticamente fiel do que a obtida pelos processos de distribuição e condução de vozes usualmente abordados na literatura sobre harmonia tonal[48]. Para uma manipulação mais livre dos resultados, foi implementada também uma função de exportação em formato MusicXML, que preserva as conduções de vozes escolhidas pelo usuário.

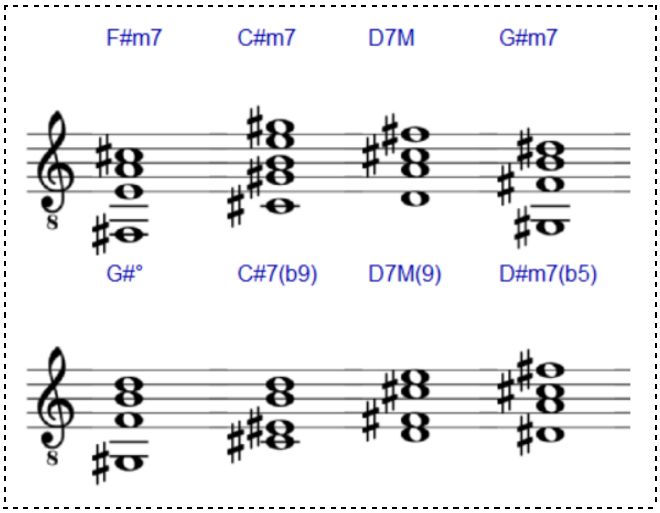

Figura 5: Visualização da notação musical em *harmonIA*

Um aspecto singular da interface construída é a possibilidade de carregar dois modelos preditivos lado-a-lado (Fig. 6), permitindo a comparação entre diferentes modelos (por exemplo, a continuação de uma sequência harmônica no contexto de duas tonalidades distintas). Neste modo multimodelo, é possível visualizar as sugestões de continuação em dois modelos simultaneamente e, ao clicar em qualquer acorde, este será acrescentado à sucessão harmônica atual.

[47] No momento, está em cogitação uma extrapolação polifônica dos dados coletados, a partir de um método original em desenvolvimento.

[48] Philip Tagg, em *Everyday Tonality*, realça este aspecto ao distinguir entre as regras de condução de vozes na harmonia tonal "clássica" e no contexto "não-clássico" (e, em especial, no *rock*), com o uso de paralelismos frequentes e estratégias harmônicas não baseadas em tríades, como os *power chords* (Tagg 2018).

Esta funcionalidade foi fundamental na avaliação subjetiva dos modelos e na experimentação com diferentes hiperparâmetros de treinamento[49].

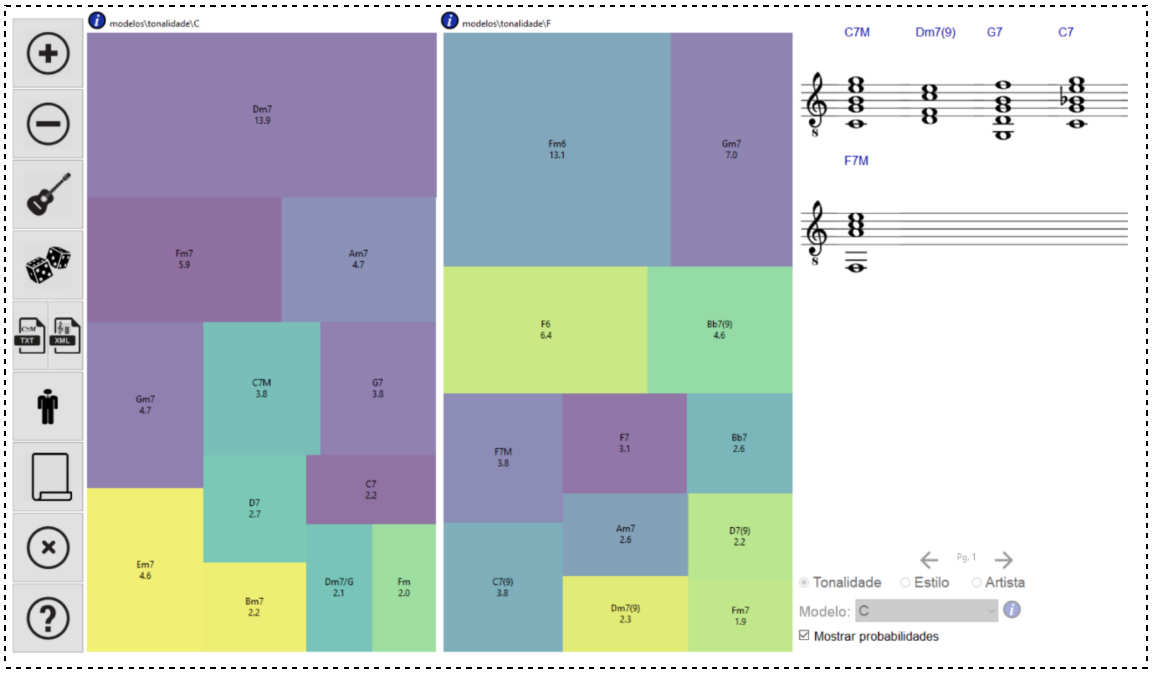

Figura 6: Dois modelos preditivos operando lado-a-lado (modo multimodelo)

Outra funcionalidade implementada para facilitar a análise dos modelos preditivos consistiu em um painel de informações com dados sobre o modelo atualmente escolhido. Esse painel fornece estatísticas sobre o *dataset* do modelo, exibe os hiperparâmetros de treinamento e permite a visualização (e escuta) das sequências de acordes mais comuns presentes nos dados brutos de treinamento, organizadas em n-gramas[50] de 4 e 8 acordes. A Fig. 7 mostra o painel de informações do modelo preditivo "*samba*":

---

[49] Este modo comparativo permitiu que variantes distintas do mesmo modelo – com hiperparâmetros de treinamento diferenciados ou diferentes mecanismos de filtragem empregados na etapa de tratamento dos dados – fossem comparados.

[50] *N-gramas* são conjuntos de elementos consecutivos. No estágio atual de desenvolvimento, essa ferramenta de análise ainda está incompleta, pois não há um agrupamento das permutações do mesmo ciclo de acordes tampouco mecanismos para lidar com a redundância (repetições da mesma sequência). Ainda assim, esta ferramenta foi útil para explorar algumas sequências de acordes comuns encontradas em cada *dataset* e para investigar possíveis enviesamentos nos modelos preditivos. Os n-gramas completos para cada modelo podem ser acessados no arquivo "*info*" contido no diretório de cada modelo.

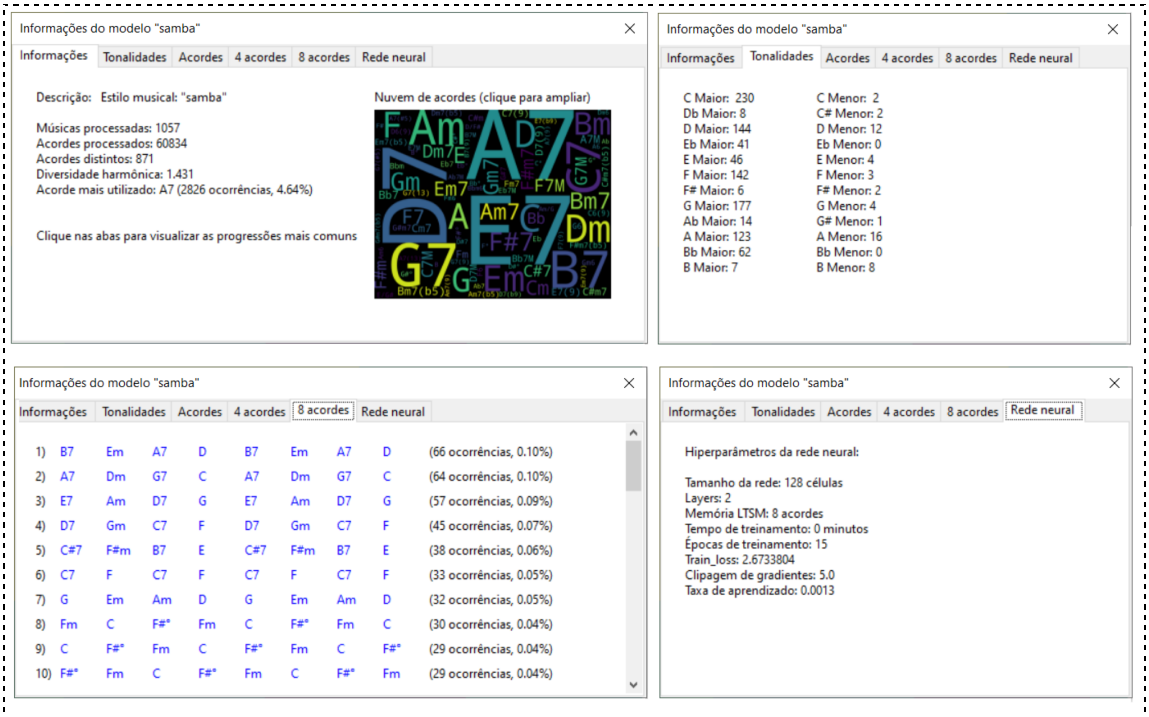

Figura 7: Painel de informações do modelo *samba*

Há, portanto, dois componentes centrais em *harmonIA*: 1) o modelo preditivo, adaptado[51] da implementação de Kim (*op. cit.*) e agora encapsulado[52] para atuar como um *back-end*[53]; 2) a interface ao usuário (*front-end*), que comanda a interação com os modelos preditivos. Por conta da separação[54] entre os dois módulos, o pesquisador cogita a possibilidade de que outras arquiteturas de

---

[51] A implementação da rede LSTM permaneceu praticamente intacta, fora mudanças pontuais nos hiperparâmetros de treinamento. Outras adaptações consistiram na remoção de partes desnecessárias para o projeto, adaptação das funções de amostragem para interação em tempo real com a interface gráfica e alterações nas rotinas de treinamento.

[52] A encapsulação é um conceito da programação orientada a objetos (OOP), denotando a separação de um objeto do restante do programa para obtenção de flexibilidade, modularidade e segurança. Transformado em um objeto, instâncias diversas do modelo preditivo podem ser executadas concorrentemente em diferentes *threads* computacionais, possibilitando a execução de dois modelos preditivos simultaneamente em *harmonIA*.

[53] Um *back-end* é o componente de um programa responsável pelo processamento efetivo dos dados, cujo funcionamento é invisível ao usuário, enquanto que o *front-end* consiste na interface ao usuário.

[54] Em *harmonIA*, cada um desses componentes roda em sua própria *thread* computacional, sincronizados por mensagens assíncronas *pub/sub*, facilitando a possível substituição do *back-end* no futuro.

aprendizado sejam incorporadas, viabilizando a visualização e operação simultânea de modelos preditivos gerados por processos diversos[55].

Além da visualização e análise dos resultados, outro aspecto que guiou o desenvolvimento da interface de criação assistida foi a perspectiva de divulgação dos resultados deste experimento a um público mais amplo[56], facilitando uma possível mobilização futura de sujeitos para avaliação externa dos resultados[57]. Visando a conveniência do usuário final, um esforço foi realizado para que o *software* pudesse ser experimentando por um público não especializado sem percalços, com o lançamento de um executável para *Windows*[58].

5. Avaliando a modelagem preditiva

Sem acesso a um conjunto de regras explícitas[59], como buscar indícios de sucesso ou fracasso na modelagem proposta em *harmonIA*? A modelagem preditiva por redes LSTM conseguiu efetivamente aprender padrões e regularidades musicalmente relevantes nos dados? Se estivéssemos lidando com um sistema meramente classificatório, aferir a capacidade de um modelo consistiria em *testar* o erro preditivo obtido ao final do treinamento, usando um conjunto de exemplos distintos daqueles utilizados no treinamento do modelo. Esse conjunto de testagem (*test set*) seria utilizado para aferir a generalização do

---

[55] Por exemplo, modelos preditivos obtidos com Cadeias de Markov – para comparação dos resultados em *datasets* com número limitado – e outras alternativas de arquiteturas de aprendizagem profunda.

[56] Ao revisar a literatura e buscar projetos similares de geração musical, a configuração de ambientes de desenvolvimento e compilação de outros projetos foi uma tarefa não trivial e muitas vezes frustrante.

[57] Yang e Lerch (2020) apontam a dificuldade em mobilizar avaliadores externos e construir testes de escuta adequados.

[58] O executável foi criado com o pacote *PyInstaller*, que incorpora em um único arquivo todo o interpretador Python e demais dependências do projeto. Desta forma, não há necessidade de configuração de um ambiente Python e instalação de dependências para executar o programa.

[59] No caso de redes LSTM, há esforços para tentar visualizar o funcionamento interno, como as dinâmicas das camadas intermediárias [*hidden layers*] em busca da extração das "regras" que condicionam a ativação de certos elementos – por exemplo, em Karpathy (2015) e Strobelt *et al*. (2018) ou, no contexto da geração musical, em Sturm (2018). Tais tentativas, no entanto, esbarram na complexidade inerente das representações aprendidas: ainda não há solução satisfatória para este complexo problema.

modelo, ou seja, sua capacidade efetiva de previsão em dados *inéditos* (Goodfellow; Bengio; Courville, *op. cit.*, p. 104). Contudo, no contexto de geração musical por modelagem, há em muitos casos uma amplitude maior de respostas (ou expectativas) contextualmente "corretas". Como avaliar, então, a plausibilidade dos resultados para além da métrica do erro preditivo?

Na literatura consultada, uma estratégia comum[60] para avaliação subjetiva consistiu em avaliar os resultados a partir do *feedback* de ouvintes[61], um tipo de *Teste de Turing* cujo objetivo é aferir se os participantes distinguiram os resultados obtidos geração automatizada da produção humana. Esta estratégia foi adotada, por exemplo, nos sistemas de geração *DeepBach* de Hadjeres, Pachet e Nielsen (*op. cit.*) e *folk-rnn* de Sturm, Santos e Korshunova (*op. cit.*), que modelaram, respectivamente, harmonização a 4 vozes ao estilo de J. S. Bach e a melodias do folclore irlandês. Este método avaliativo, no entanto, pode ser bastante inconsistente[62] e dificilmente avalia a efetiva relevância do modelo para o usuário final – o músico[63].

Embora a avaliação externa seja uma etapa importante, neste artigo enfoco em uma etapa que entendo como anterior – pouco abordada na literatura consultada – na qual um *expert* no campo de conhecimento relacionado ao modelo construído[64], aprofunda o exame dos resultados a partir de seu

---

[60] Veja Siphocly, El-Horatby e Salem (*op. cit.*), Yang e Lerch (*op. cit.)*, Carvonali e Rodà (2020) e Ji, Luo e Yang (*op. cit.*) para uma discussão aprofundada sobre métodos de avaliação – objetivos e subjetivos – na geração musical.

[61] Quantos e quem são estes ouvintes, quais seus níveis de *expertise* e aculturação são aspectos a serem cuidadosamente considerados nessas estratégias segundo Yang e Lerch (*op. cit.*).

[62] Yang e Lerch (*op.* cit.) criticam a fragilidade, inconsistência e falta de rigor das avaliações com tarefas de escuta musical inspiradas no modelo de Turing, apontando que os resultados reportados, em sua ampla maior, devem ser compreendidos somente como uma avaliação preliminar, sem validade científica.

[63] Cientes das inadequações dos métodos usados para avaliar sistemas de geração musical, um interessante experimento foi realizado por uma equipe multidisciplinar unindo pesquisadores, desenvolvedores e músicos. Esta equipe experimentou fluxos de criação musical assistida utilizando diversas ferramentas de geração musical com redes neurais, produzindo reflexões sobre o potencial e relevância destas ferramentas em um contexto criativo (Sturm *et al*. 2019).

[64] Drew Conway qualifica a ciência de dados como a interação a intersecção entre três habilidades: *Hacking Skills, Math* & *Statistics Knowledge* e *Substantive Expert*ise (Conway 2019). Para este autor, pela abrangência de cada campo, projetos desta natureza só são bem sucedidos quando as três competências são igualmente satisfeitas (seja por um pesquisador individual ou por uma equipe multidisciplinar).

conhecimento e familiaridade prévia com os dados modelados. Relembremos a premissa inicial da modelagem proposta em *harmonIA*: um modelo deve ser capaz de prever plausivelmente o acorde subsequente em sequências de acordes inéditas, que não fizeram parte dos dados de treinamento – ou seja, ser capaz de generalizar o aprendizado para novos contextos. No entanto, o que vem a ser uma continuação plausível dentro dos limites da representação modelada (sucessões de acordes enquanto cifras cordais)?

Para esboçarmos alguns parâmetros para esta plausibilidade, é preciso considerar que nem todos contextos harmônicos possuem o mesmo nível de *incerteza*: há passagens com implicações – ou expectativas – harmônicas convencionalizadas e outras de continuação mais incerta[65]. Derivamos desta premissa um primeiro critério possível de validação: estimulando os modelo preditivos com sucessões de acordes com fortes implicações harmônicas (menor entropia) no contexto da harmonia tonal – por exemplo, resoluções de acordes com função dominante, progressões harmônicas rumo à tônica e expectativas decorrentes da condução das vozes[66] –, serão as probabilidades resultantes condizentes com as continuações esperadas segundo construtos teórico-analíticos de reconhecida validade?

Há alguns fatores de complicação aqui, em especial a *heterogeneidade* do *dataset*. Por exemplo, no *corpus* composto pelas canções populares classificadas em Dó Maior (cerca de 10.000 canções), além de exemplos com erros na classificação da tonalidade[67], há muitas sucessões de acordes com comportamentos fora da funcionalidade tonal, ou cujos comportamentos

---

[65] Na música tonal, há a recorrência de elementos fortemente implicativos, como nas fórmulas cadências que sinalizam o fechamento de um elemento morfológico. Estas implicações mediam o equilíbrio entre a comunicabilidade de uma linguagem musical – através da recorrência de fórmulas padronizadas que criam consistência estilística – e a diversidade de seus elementos.

[66] Como explorado anteriormente, a condução das vozes não é diretamente modelada, porém padrões de continuidade entre diferentes inversões fazem parte de estratégias de condução de vozes.

[67] Os dados coletados para a tonalidade de Dó Maior foram os que apresentaram maior erro na classificação, provavelmente por ser a configuração *default* para tonalidade quando não há classificação da tonalidade por parte do usuário no site dos quais os dados foram coletados.

apresentam certo hibridismo entre a não-funcionalidade e a funcionalidade[68]. A análise das sucessões harmônicas por *n-gramas* neste modelo mostra o ciclo de acordes "C –G – Am – F" (I – V – vi – IV) como a sequência harmônica mais comum no *dataset* bruto da tonalidade de Dó Maior[69], bem como um número elevado de sequências "C – G – F – C" (I – V – IV – I), trazendo indícios de que comportamentos fora da funcionalidade tradicional[70] podem ter introduzido algum ruído na modelagem de harmonias funcionais de forte viés implicacional. Apesar de ser um possível fator complicador para a modelagem, a heterogeneidade não é indesejável: o aspecto mais distinto da arquitetura de aprendizagem com redes LSTM é justamente sua capacidade de modelagem das dependências prévias em elementos sequenciais, seja regularidades locais quanto associações entre elementos remotos (Goodfellow; Bengio; Courville, *op. cit.*, p. 411). Em tese, o modelo aprendido será capaz de diferenciar em alguma medida contextos harmônicos funcionais de contexto não-funcionais pelo entorno dos elementos. Hadjeres (2018) pondera que a escolha por *corpus* mais ou menos heterogêneos é um aspecto com consequências significativas para a geração musical com aprendizagem profunda:

> Eu acredito que o *trade-off* entre o tamanho do *dataset* e sua coerência seja uma questão de vital importância ao construir modelos de geração por aprendizagem profunda. Se o *dataset* é muito heterogêneo, um bom modelo gerativo deverá ser capaz de distinguir as diferentes subcategorias e generalizar satisfatoriamente. Porém, se só existirem diferenças sutis entre as subcategorias, é importante avaliar se o "modelo médio" pode produzir resultados musicalmente interessantes (Hadjeres 2018, p. 24, tradução nossa)[71].

---

[68] Veja por exemplo a classificação proposta por Walter Everett sobre os diferentes comportamentos tonais presentes no *rock* e seus hibridismos com aspectos da funcionalidade harmônica (Everett 2004).

[69] Esta sucessão de acordes também predominou na maior parte das demais tonalidades.

[70] Philip Tagg (2018) aborda alguns tipos específicos destas sucessões harmônicas, como os ciclos de acordes – *loops* ou *vamps* na terminologia proposta pelo autor –, esclarecendo que estas estratégias harmônicas resultam em sugestões ambivalentes de tonalidade, bem como empregam condução de vozes por paralelismo que neutralizam as tendências funcionais de certos graus melódicos.

[71] "I believe that this trade-off between the size of the datasets and their coherence is one of the major issue when building deep generative models. If the dataset is very heterogeneous, a good generative model should be able to distinguish the different subcategories and manage to generalize well. On the contrary, if there are only slight differences between subcategories, it is

Para simplificar a busca por indícios de sucesso no aprendizado de comportamentos implicacionais na harmonia tonal e, ao mesmo tempo, comparar o potencial de aprendizado em *datasets* com níveis distintos de heterogeneidade, um *corpus* de treinamento composto unicamente por sucessões de acordes com nítidos comportamentos funcionais foi compilado. A intenção deste *dataset* foi modelar um idioma harmônico homogêneo e bem-definido, cujos parâmetros implicativos são facilmente reconhecíveis. O ponto de partida para este modelo foram as canções em Dó Maior do modelo estilístico *"samba"*, um *dataset* cuja visualização das sucessões de acordes por *n-gramas* revelou indícios de comportamentos harmônicos bastante unificados, condizentes com a funcionalidade harmônica da tradição tonal. Para aprimorar a consistência do *dataset*, erros de classificação nas tonalidades foram manualmente corrigidos e exemplos com comportamentos indesejados foram excluídos. Além disso, outros exemplos representativos de comportamentos funcionais do *dataset* global foram selecionados de forma criteriosa e acrescentados ao conjunto de treinamento deste modelo. Por fim, este *dataset* foi transposto, resultando em modelos (idênticos) para as 12 tonalidades maiores, ou seja, uma espécie de *benchmark* para comportamentos funcionais usuais na música tonal – sem, contudo, ser capaz de produzir muita variedade na geração tendo em vista o conjunto de treinamento mais limitado[72].

O Quadro 1 exibe uma tentativa de validação dos modelos de tonalidades, apresentando as previsões realizadas a partir de sucessões harmônicas "inéditas" – que não fizeram parte dos dados de treinamento – envolvendo contextos implicacionais de nítida funcionalidade harmônica. Neste quadro, são mostradas as previsões obtidas em dois modelos distintos para cada tonalidade: 1) no modelo restritamente homogêneo descrito no parágrafo anterior, emulando comportamentos puros de funcionalidade tonal e 2) no modelo de tonalidade altamente heterogêneo, composto por todos os exemplos coletados em determinada tonalidade. Como o *output* do modelo preditivo consiste em somente *um* acorde de continuação, eventualmente foram inclusas nas colunas

---

important to know if the "averaged model" can produce musically-interesting results" (Hadjeres 2018, p. 23)

[72] Este *dataset* de treinamento contém somente 260 exemplos de sucessões de acordes (cada exemplo consiste em uma música popular completa), enquanto que os modelos de tonalidades mais abrangentes contém entre 2.000 e 10.000 exemplos.

de previsões uma consequência implicacional subsequente, ou seja, uma continuação de alta probabilidade gerada pelo mesmo modelo preditivo *após* o acorde sugerido. Estas implicações subsequentes estão indicadas no Quadro 1 por setas entre parênteses (→), seguidas pelo acorde de continuação sugerido. É importante ressaltar a diferença significativa na abrangência dos *datasets* de treinamento nos dois tipos de modelos de tonalidades demonstrados no Quadro 1: enquanto que o modelo restritamente homogêneo de comportamentos funcionais foi treinado a partir de 260 sequências de acordes – cada sequência é uma canção popular completa –, os modelos de tonalidades heterogêneos foram treinados com 2.000 a 10.000 sequências.

Sucessão harmônica → continuação original	Previsão (modelo homogêneo)	Previsão (modelo heterogêneo)
Gm7 C7 F7M/A A♭° Gm7 C7sus4 → C7 F7M *Contexto implicativo*: Resolução de acorde dominante Padrão de condução de voz *Tonalidade*: Fá Maior	F7M 22,2% C7 16% (→ **F7M**) F6(9) 5,4% Am7 5,2% F 4,6%	C7 58,8% (→ **F7M**) C7(9) 9,1% Gm7 4,4% F 3,8% Am7 2,6%
F D7 Gm7 C7 Fm A♭7 D♭7 → C7 Fm *Contexto implicativo*: Acorde de sexta aumentada *Tonalidade*: Fá Maior	C7 40,9% (→ **F** ou **Fm**) Am7 4,2% 3,9% C 2,9% C7(9) 2,6% C7(13)	C7 9,3% (→ **F7M** ou **F**) B♭m6 G7 8% B♭m 4,2% D♭7 3,2% 3,2%
G A7 D7 G G7 C C♯° → D7 G *Contexto implicativo*: Acorde diminuto como dominante secundária *Tonalidade*: Sol Maior	G 78% Am 6,5% C 2,6% Bm 1,9% D7 1,2%	D7 46% (→ **G**) G 20,9% D 5,4% C 4,6% A7 4,4%
F Dm7 Gm7 C7 F B♭7 → A7 Dm *Contexto implicativo*: Acorde de sexta aumentada resolvendo em dominante secundária *Tonalidade*: Fá Maior	B♭ 9,3% Gm7 8,5% D7 8% B♭7M 7, 6% A7 5,8%	A7 18,3% C7 17,9% Gm7 13,5% Dm7 11,3%

C G7/D C/E E7/B Am/C F C/E D7 D7/F♯ → G7 *Contexto implicativo*: Dominante secundária — *Tonalidade*: Dó maior	G7 12,6% (→ C) G 10,2% E7 9,7% Am 3,4% Em 3,2%	G7 39,4% (→ C) G 10,6% F/G 9,4% C/E 4,4% Gsus4 3,5% (→ G7)
F7M F6 F7M F6 D♭7(9) C7(9) F7M F♯° → Gm7 *Contexto implicativo*: Acorde diminuto como dominante secundária — *Tonalidade*: Fá Maior	Gm7 27,7% (→ C7) G7(9) 11,3% (→ C7) Gm7(9) 8,1% (→ C7) Gm 5,9% (→ C7) C7 5,6% (→ F7M) G7 3,2% (→ C7)	Gm7 57% (→ C7(9)) F7M 3,4% B♭7M 3,1% Gm7(9) 2,8% Gm7 2,7%
F7M F6 F7M F6 D♭7(9) C7(9) F7M F♯° Gm7 Am7(♭5) → D7(♭9) Gm7 *Contexto implicativo*: Acorde ii^{ø7} secundário — *Tonalidade*: Fá Maior	B♭7M 14,9% D7 9,3% (→ Gm) Dm7 7,7% Dm 5% Gm7 3,2%	D7(♭9) 34,5% (→ Gm7) D7(9) 22,5% (→ Gm7) D7 16,6% (→ Gm7) Dm 3,2% (→ Gm7)

Quadro 1: Previsão de continuidades geradas pelos modelos de tonalidades.

A comparação entre os dois modelos corrobora a observação anterior de Hadjeres (*op. cit.*, p. 8) sobre o valor da heterogeneidade no *dataset*. Ainda que o modelo mais homogêneo e consistente tenha conseguido modelar grande parte[73] dos contextos implicativos, o modelo mais heterogêneo apresentou uma contextualização superior das dependências prévias, resultando em previsões melhores e alternativas bastante plausíveis. Um exemplo bastante evidente desta diferença pode ser percebida na sucessão "F Dm7 Gm7 C7 F B♭7". Este é um exemplo desafiador para a previsão por conta da necessidade de contextualização prévia, que idealmente irá evitar indicar E♭ como resolução provável e interpretar B♭7, no contexto da tonalidade de Fá Maior, como um acorde de sexta aumentada que resolve na dominante secundária do sexto grau (V7/vi ou V/vi). Neste caso, o modelo homogêneo exposto a menos exemplos de treinamento não conseguiu chegar a uma previsão satisfatória, enquanto que o modelo mais heterogêneo forneceu a resposta contextualmente desejada. Resultado semelhante foi obtido na sucessão "F7M F6 F7M F6 D♭7(9) C7(9) F7M

[73] As previsões com erros óbvios foram resultado das limitações do *dataset* de treinamento. Sem uma quantidade mínima de exemplos representativos, certos comportamentos não foram aprendidos satisfatoriamente.

F♯° Gm7 Am7(♭5)", cuja continuidade foi melhor contextualizada pelo modelo heterogêneo. Neste caso, o modelo preditivo heterogêneo indicou o acorde D7(♭9) seguido por Gm7 como continuação mais provável, antevendo uma possível tonicização do segundo grau de Fá Maior.

A seguir, avaliamos em mais detalhes as capacidades dos modelos preditivos treinados em contextualizar os resultados a partir dos elementos prévios. Para demonstrar o impacto destas dependências (relacionamentos entre informações prévias), utilizaremos como exemplo a sucessão de acordes "F D7 Gm7 C7 Fm A♭7 D♭7", retirada[74] de um excerto do choro "Ainda me Recordo", de Pixinguinha e Benedito Lacerda (Fig. 7).

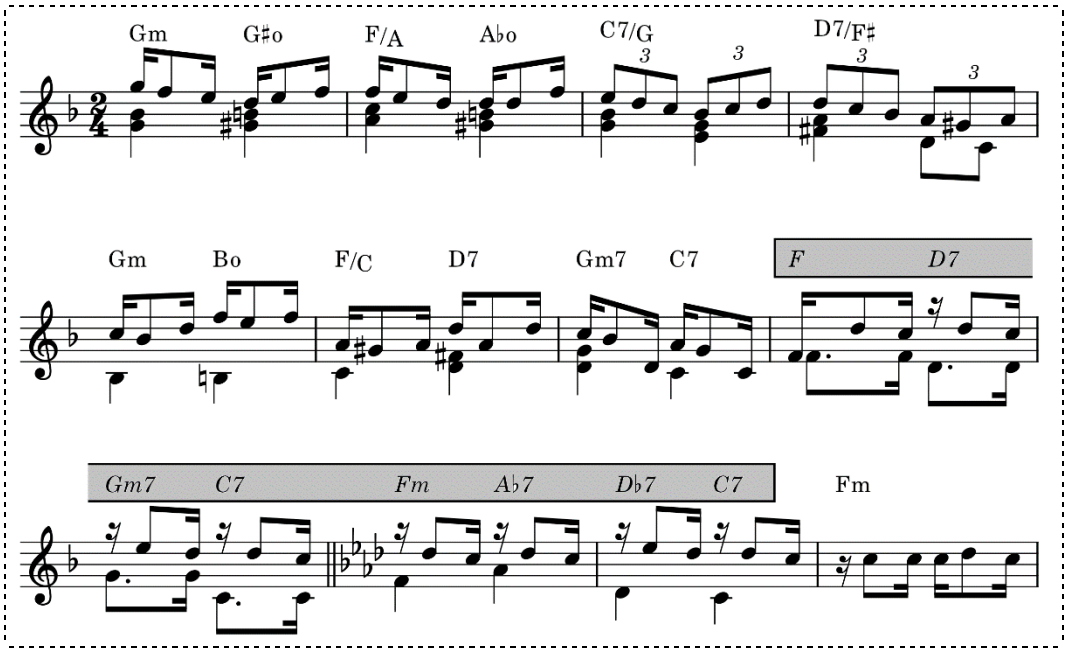

Figura 7: Excerto de "Ainda me Recordo"[75] (Pixinguinha e Benedito Lacerda)

Neste choro, a sucessão em destaque representa o final da primeira seção (em Fá maior) e transição para a próxima seção em Fá Menor. A continuidade imediata desta sucessão já foi avaliada no Quadro 1, e nesta sucessão o contexto implicacional avaliado consistiu em determinar se os modelos da tonalidade de Fá Maior seriam capazes de compreender o acorde D♭7 como um acorde de sexta aumentada, sugerindo uma resolução na dominante C7. Neste caso, como a outra

[74] Esta sucessão de acordes não esteve presente no conjunto de treinamento, portanto se tratam de dados inéditos.

[75] Transcrição de Carrasqueira (1997). A sucessão harmônica utilizada como estímulo de entrada no Quadro 2 está em realce.

possibilidade implicativa – resolução em G♭ – não é um acorde diatônico e representa distanciamento tonal significativo, a resolução para C7 se torna uma implicação de alta probabilidade – uma *regularidade* – que foi modelada sem dificuldades. Porém, será o modelo "heterogêneo" de Fá Maior capaz de reconhecer a mudança de modo (de Fá maior para Fá Menor) em curso, envolvendo uma associação entre elementos mais distanciados[76]? E de que forma os elementos prévios influenciam esta ponderação?

O Quadro 2 explora algumas variações desta sucessão de acordes, manipulando o estímulo de entrada - os elementos em realce na primeira coluna representam alterações na sucessão harmônica - para experimentar a influência dos elementos prévios na previsão, em especial o acréscimo ou exclusão de acordes associados à tonalidade de Fá Menor. Assim como no Quadro 1, setas entre parênteses (→) representam consequências implicativas subsequentes de alta probabilidade, indicando um próximo acorde após a continuação sugerida. Visando investigar concomitantemente a dimensão da heterogeneidade, o quadro também compara a previsão entre os modelos de tonalidade (heterogêneos) e o modelo treinado com o *dataset* global "12tons", com 52.303 exemplos de treinamento (contra 4.610 no modelo da tonalidade de Fá Maior). Ao utilizar o modelo "12tons" para previsão, estamos testando também se este modelo extremamente heterogêneo é capaz de discernir, pelo contexto das dependências prévias, as implicações tonais corretas sem ter sido treinado em uma tonalidade específica.

A manipulação do estímulo de entrada na Quadro 2 demonstra uma consistente influência dos elementos prévios na modelagem: mais acordes e comportamentos associados à tonalidade de Fá Menor resultam em menor incerteza na previsão da resolução de C7 para o acorde de Fá Menor em vez de Fá Maior. Algumas das variações propostas simularam contextos com implicações ambíguas, característicos da mistura modal. Na música tonal, tais ambiguidades[77] e misturas geralmente precedem modulações, bifurcando as implicações harmônicas prévias pelo acréscimo de incerteza. No exemplo em

[76] Aqui temos um caso de operação simbólica diretamente relacionado aos Experimentos 6a/6b em Hochreiter e Schmidhuber (1997), artigo que introduz a arquitetura LSTM.

[77] Como na ideia do *mehrdeutigkeiten* ("ambiguidade" ou "múltiplos significados") presente nas teorias harmônicas austro-alemãs (Freitas, 2010, p. 520–534; Freitas, 2019).

questão, o modelo preditivo parece ter lidado de forma satisfatória com o impacto dos acordes prévios na consolidação de novas implicações – a sugestão de mudança de modo. Destaco, assim como na Quadro 1, o impacto da quantidade de exemplos nos *datasets* de treinamento: na modelagem das dependências prévias, o modelo "12tons" – com um conjunto de treinamento dez vezes maior que o modelo da tonalidade de Fá Maior – demonstrou uma capacidade mais refinada em distinguir a resolução esperada nos modos maior ou menor. O modelo "12tons" também apontou hipóteses alternativas mais plausíveis, demonstrando de modo geral uma melhor compreensão da influência de elementos passados na previsão do elemento subsequente.

Sucessão harmônica	Previsão (modelo "Fá Maior")	Previsão (modelo "12 tons")
F D7 Gm7 C7 Fm A♭7 D♭7 C7 (*sucessão original, cuja continuidade é o acorde de Fm*)	F 29% Fm 13,2% C7 4,4% (→ **F**) Fm7 4,4%	Fm 32,3% Fm7 19,1% F 13,4% F7 7% C7 4,4% (→ **Fm**)
F D7 Gm7 C7 **F** A♭7 D♭7 C7 *(variante 2, com alteração de acorde)*	F 74,9% C7 3,7% (→ **F**) B♭ 2,9% F7 2,7%	F 57,7% F7 6,4% C7 5,9% (→ **F**) Fm 4.7%
F D7 Gm7 C7 **F Dm** D♭7 C7 *(variante 2, com alteração de 2 acordes)*	F 80,9% B ♭ 3,5% C7 2,3% F7 2,1%	F 71,4% Dm 3,5% Gm 3,4% F7 3,2%
F D7 Gm7 C7 Fm ~~A♭7~~ D♭7 C7 *(variante 3, com supressão de acorde)*	F 21,6% Fm 9% F7 5,8% Fm7 5,8%	Fm 36,4% Fm7 11,1% C7 10,2% F 8,9% F7 2,4%
Fm D7 Gm7 C7 Fm A♭7 D♭7 C7 *(variante 4, com alteração de acorde)*	F 23,3% Fm7 8,7% Fm 7,8% F7 7,7%	Fm 32,6% C7 12,8% F 6,6% A7 4,7% (→ **D7**)

Gm B° F/C D7 Gm7 C7 F D7 Gm7 C7 Fm A♭7 D♭7 C7 *(variante 5, com acréscimo de acordes para contextualização prévia)*	F 23,3% Fm 20,8% C7 4,1% (→ **F**) Fm7 4%	Fm7 23,3% Fm 22,1% F 8,8% C7 6,6% (→ **Fm**) Gm 6,4% (→ **C7** → **F**)
Gm B° F/C D7 Gm7 C7 Fm D7 Gm7 C7 Fm A♭7 D♭7 C7 *(variante 5, com acréscimo de acordes para contextualização prévia e alteração)*	Fm 22% F 14,3% Fm7 5,9% F6(9) 4,1% E♭7 2,9% (→ **D7**)	Fm 31% Fm7 22,2% Gm 8% C7 7,6% (→ **Fm**) F 3,6%

Quadro 2: A influência das dependências prévias na previsão de continuidades

É necessário cautela na interpretação desse resultado: é possível que os resultados sejam consequência de um viés indesejado – como alguma predileção aprendida ao longo do treinamento no modelo "12tons". Para descartar esta possibilidade e evidenciar se o modelo "12tons" foi capaz de aprender implicações similares em diferentes tonalidades – a partir do seu conjunto extremamente heterogêneo de treinamento – o Quadro 3 apresenta as previsões resultantes da mesma sequência de acordes transposta em diversas tonalidades.

Sucessão harmônica	Previsão	Sucessão harmônica (variante)	Previsão
F D7 Gm7 C7 Fm A♭7 D♭7 C7	Fm 32,3% Fm7 19,1% F 13,4% F7 7%	F D7 Gm7 C7 F A♭7 D♭7 C7	F 45,8% Fm 10% F7 7,3% C7 5,3%
G E7 Am7 D7 Gm B♭7 E♭7 D7	Gm 36% G7 15,8% G 10,8% B♭7 4,1%	G E7 Am7 D7 G B♭7 E♭7 D7	G 40,8% G7 10,5% A7 4,4% E♭7 3,7%
A F♯7 Bm7 E7 Am C7 F7 E7	Am 29,1% A 17,4% A7 12,9% Am7 8,8%	A F♯7 Bm7 E7 A C7 F7 E7	A 34,5% A7 8,4% B7 7,1% F7 6%

E♭ C7 Fm7 B♭7 E♭m G♭7 B7(*) B♭7 (*) Cb7 não fez parte do vocabulário aprendido, portanto B7 foi utilizado como enarmonia.	E♭m	17,2%	E♭ C7 Fm7 B♭7 E♭ G♭7 B7 B♭7	E♭	44,7%
	E♭	13,3%		B♭7	12,7%
	E♭m7	7,5%		E♭m	7,6%
	C7	6,9%		C7	4,6%

Quadro 3: Previsões (modelo *12tons*) para a mesma sucessão em tonalidades distintas

O Quadro 3 fornece indícios de que o modelo "12tons" conseguiu aprender diferentes subcategorias de contextos implicativos pelo conjunto das dependências prévias – ou seja, o entorno harmônico –, conseguindo modelar comportamentos de diferentes tonalidades de forma satisfatória. Neste caso, é importante frisar que nenhum rótulo de tonalidades foi fornecido no treinamento deste modelo, portanto todos esses contextos tonais foram aprendidos *sem supervisão* (Goodfellow; Bengio; Courville, *op. cit.*, p. 146), demonstrando a impactante capacidade do aprendizado computacional para este tipo de modelagem harmônica.

A partir dos resultados apresentados, podemos esboçar algumas avaliações – em caráter preliminar:

1) O modelo preditivo com redes LSTM foi capaz de aprender, sem supervisão, alguns comportamentos implicativos conhecidos na harmonia tonal de sentido funcional;
2) além de implicações locais[78], o modelo preditivo aprendeu relações entre elementos mais remotos (dependências de longo-prazo). Esta capacidade foi testada ao limite no modelo "12 tons", treinado com todas as 52.303 sucessões de acordes coletadas e que demonstrou boa capacidade de generalização;
3) em virtude do ponto anterior, consideramos provável que, além dos comportamentos conhecidos, outros comportamentos menos óbvios possam ter sido modelados com algum grau de consistência;

[78] Estas implicações mais locais, que dependem do passado mais imediatamente recente, também poderiam ser modeladas de forma satisfatória com abordagens menos complexas, como as Cadeias de Markov.

4) os *datasets* heterogêneos, com um número maior de exemplos apresentaram resultados mais contextualmente consistentes e variados, corroborando a necessidade de um número extremamente elevado de exemplos de treinamento.

Embora estes resultados preliminares sejam promissores e consistentes com a capacidade teórica das redes LSTM em modelar dependências de longo prazo, é necessário cautela e maior aprofundamento nesta avaliação.

Destacamos a seguir alguns aspectos problemáticos observados:

1) Houve bastante sensibilidade aos hiperparâmetros de treinamento. Mudanças pontuais na estrutura da rede resultaram em previsões significativamente distintas em certos modelos. Este aspecto inspira certa cautela em relação à modelagem preditiva, porém possui menor consequência para a geração musical assistida[79];
2) o número extremamente elevado de elementos necessários para uma modelagem do contexto de dependências prévias – mesmo levando em conta a representação unidimensional adotada – evidencia a necessidade de procedimentos de coleta e tratamento automatizados. Para garantir uma consistência mínima nos dados, foram necessários diversos mecanismos de controle e filtragem, um aspecto que demandou esforço considerável neste projeto. A demanda por *datasets* massivos sinaliza também a complexidade da modelagem envolvendo a interação orgânica entre diversos parâmetros musicais;
3) a modelagem do estilo harmônico de artistas específicos não atingiu um nível de generalização minimamente satisfatório, portanto os resultados sequer foram considerados nessa avaliação[80]. Estratégias

[79] Para avaliação da modelagem preditiva, consideramos a previsão em contextos harmônicos de implicações claras. A eficácia desta previsão, no entanto, é um fator menos relevante para geração musical assistida, na qual uma plausibilidade mínima da continuação é o único critério. Em busca de aprimorar a avaliação da modelagem preditiva, métricas mais refinadas teriam que ser desenvolvidas para delinear parâmetros claros de desvio dos resultados pretendidos.

[80] Para *datasets* reduzidos, abordagens probabilísticas com cadeias de Markov parecem ser mais viáveis.

simples de expansão dos dados, como a transposição, não foram efetivas neste caso. No caso de *datasets* menores, é imperativa a necessidade de revisão e tratamento manual para assegurar uma maior consistência interna, objetivo que não foi atingido com os processos automatizados utilizados;

4) os modelos preditivos treinados nas sucessões harmônicas classificadas por estilos musicais não foram objeto de avaliação, tendo em vista a falta de comportamentos suficientemente distintos entre si na representação simbólica adotada;
5) no caso de sistemas interativos, dados de entrada fornecidos pelo usuário precisam também passar por tratamentos de uniformização e normalização, em virtude da especificidade das representações simbólicas aprendidas pela máquina. Construir estes mecanismos e abordar *edge cases* é mais um dos diversos desafios a ser considerado na geração que envolva manipulação da camada de entrada da rede neural.

Por uma limitação de escopo deste artigo, não abordaremos a avaliação sobre a plausibilidade da modelagem em contextos harmônicos menos convencionalmente implicativos. A este respeito, sugerimos como possíveis critérios de avaliação construtos como a consistência harmônica[81] e a proximidade na condução das vozes, partindo do princípio que movimentações com algum grau de proximidade serão mais plausíveis que movimentações envolvendo maior distanciamento[82].

6. Considerações finais

Neste artigo, apresentamos um *software* para interação com modelos preditivos conexionistas, configurando um ambiente (limitado) de

[81] Dmitri Tymockzo conceitua a consistência harmônica como a "tendência das harmonias em uma passagem musical serem estruturalmente semelhantes", ou seja, apresentarem formações intervalares e cardinalidades similares (Tymockzo 2011, p. 6).

[82] O construto da proximidade e eficiência da condução das vozes é abordado em profundidade em "*A Geometry of Music"* (Tymockzo, *op. cit.*).

geração musical assistida. Esta iniciativa se inscreve dentro de um esforço de musicologia empírica (Cook 2014), explorando a aprendizagem computacional como ferramenta para descoberta de novas possibilidades teórico-musicais baseadas em embasamentos empíricos inferidos de dados musicais concretos. A natureza multifacetada da aprendizagem computacional demanda o encontro de competências diversas, e aquelas específicas ao teorista-analista tem um lugar essencial na avaliação dos resultados da modelagem e geração musical com paradigmas conexionistas. O teorista-analista pode não só contribuir com a avaliação dos resultados – enquanto usuário final destas tecnologias – mas também auxiliar e orientar as etapas de coleta, tratamento dos dados e treinamento dos modelos.

Em relação a construção dos *datasets* de treinamento, o autor frisa a importância do equilíbrio entre aspectos de homogeneidade (comportamentos relacionados) e heterogeneidade (diversidade de exemplos). Neste experimento, foi nítido o desempenho superior dos modelos treinados em *datasets* massivos, compostos por *milhões* de elementos individuais. Ainda assim, modelos mais reduzidos e homogêneos conseguiram modelar comportamentos específicos, embora com menor capacidade de contextualização. Novamente, o julgamento do teorista-analista neste caso é fundamental para atestar a relevância e alcance dos modelos criados – aspecto pouco aprofundado na literatura consultada – e auxiliar na criação de estratégias de tratamento dos dados, como mecanismos de filtragem, normalização e expansão. Em virtude da dificuldade da construção de *datasets* para treinamento, o autor enfatiza, em consonância com os paradigmas da Ciência Aberta[83], a necessidade de divulgação dos dados de treinamento em formatos agnósticos.

A interface de geração musical interativa proposta ainda precisa de aprimoramentos, porém propiciou uma investigação inicial da modelagem preditiva que não teria sido possível sem os mecanismos de controle e análise implementados. Possíveis aplicações da modelagem preditiva com interfaces interativas no campo da pedagogia musical são um aspecto para investigações futuras, explorando a interação entre modelos empíricos e construtos teóricos consolidados ou a modelagem em repertórios menos sistematizados pela literatura teórico-analítica.

---

[83] Veja Albagli, Clinio e Raychtock (2014).

Na avaliação da modelagem preditiva, esboçamos um possível critério de avaliação da capacidade de generalização da modelagem preditiva, buscando sucessões com implicações harmônicas convencionais na harmonia tonal de sentido funcional. No entanto, uma avaliação mais consistente para a geração musical deve considerar diversos outros parâmetros[84] e abranger também contextos menos implicativos. Mesmo trabalhando com uma única dimensão representativa, a modelagem preditiva aqui apresentada mostrou indícios de ter aprendido padrões associativos musicalmente relevantes, todavia, esse aspecto ainda demanda avaliações mais aprofundadas. Ainda que exista grande potencial no uso da aprendizagem computacional profunda para modelagem preditiva, ressaltamos a extrema complexidade envolvida em extrair e validar regras e comportamentos aprendidos pelo modelo preditivo. No entanto, o propósito da propósito da criação musical assistida, para o qual não há necessidade de extração de regras simbólicas, parece extremamente promissor e certamente será um ponto de investigação futura.

Como uma direção de continuidade futura deste trabalho, o autor pretende investigar a possibilidade de extrapolação polifônica do *dataset* coletada, usando as sucessões de acordes e posições ao violão coletadas como ponto de partida para uma heterogeneidade polifônica artificialmente construída, usando algoritmos de condução automatizada de vozes, visando especificamente um contexto de criação musical assistida e não a modelagem estilística. O autor entende que esta seja uma estratégia viável para lidar com a dificuldade de coleta de dados musicais em quantidade massiva para o treinamento de redes neurais artificiais.

O código-fonte, dados brutos coletados e o executável para sistemas *Windows* de *harmonIA* podem ser encontrados em <https://github.com/frauber84/harmonIA>.

Referências

1. Albagli, Sarita; Clinio, Anne; Raychtock, Sabryna. Ciência Aberta: correntes interpretativas e tipos de ação. 2014. *Linnc em Revista*, v. 10, n. 2.

[84] Vide o *survey* de critérios avaliativos de Yang e Lerch (*op. cit.*).

2. Bayle, Yann. Deep Learning for Music (DL4M), disponível em <https://github.com/ybayle/awesome-deep-learning-music>. Acesso em 20 de fevereiro de 2019.
3. Briot, Jean-Pierre; Hadjeres, Gaëtan; Pachet, François. 2020. *Deep Learning Techniques for Music Generation*. Suíça: Spinger.
4. Carvonali, Fillipo; Rodà, Antonio. 2020. Computational Creativity and Music Generation systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence*, v. 3.
5. Carrasqueira, Maria José. 1997. *O Melhor de Pixinguinha*: melodias e cifras. São Paulo: Irmãos Vitale.
6. Conway, Drew. 2010. *The Data Science Venn Diagram*. [S. l.]. Disponível em: <https://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>. Acesso em: 24 jan. 2019.
7. Cook, Nicholas. 2014. Computational and Comparative Musicology. In: Clarke, Eric; Cook, Nicholas. *Empirical Musicology: Aims, Methods, Prospects*. Oxford: Oxford University Press.
8. Choi, Keunwoo. *LSTM source code to generate jazz chord progressions*. Disponível em <https://github.com/keunwoochoi/lstm_real_book>. Acesso em 30 de maio de 2019.
9. Choi, Keunwoo, Fazekas, George, Sandler, Mark. 2016. Text-based LSTM networks for Automatic Music Composition. *1st Conference on Computer Simulation of Musical Creativity (CSMC 16)*.
10. Everett, Walter. Making Sense of Rock's Tonality System. 2004. *Music Theory Online, Society for Music Theory*, v. 10, n. 10. Disponível em: <https://www.mtosmt.org/issues/mto.04.10.4/mto.04.10.4.w_everett.html.> Acesso em: 30 de março de 2022.
11. Freitas, Sérgio Paulo Ribeiro de. 2010. *Que acordo ponho aqui?* Harmonia, práticas teóricas e o estudo de planos tonais em música popular. Tese (Doutorado) - Universidade Estadual de Campinas. Campinas – SP.
12. _____. 2019. Ambiguidade: uma palavra-chave na teoria tonal. *Revista Vórtex*, v. 7, n. 2, p. 1–31.
13. Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron. 2016. *Deep Learning*. Boston: MIT Press.
14. Hadjeres, Gaëtan. 2018. *Interactive deep generative models for symbolic music*. Tese (Doutorado) - Sorbonne Université, Paris, 2018.

15. Hadjeres, Gaëtan; Pachet, François; Nielsen, Frank. 2017. DeepBach: a Steerable Model for Bach Chorales Generation. *Proceedings of the 34th International Conference on Machine Learning*.

16. Hochreiter, Sepp; Schmidhuber, Jürgen. 1997. Long Short-Term Memory. *Neural Computation*, v. 9, n. 12, p. 1735–1780.

17. Ji, Shulei; Luo, Jing; Yang, Xinyu. 2020. *A Comprehensive Survey on Deep Music Generation: Multi-level Representations, Algorithms, Evaluations, and Future Directions*. Preprint, submetido em novembro de 2020, <https://arxiv.org/abs/2011.06801>.

18. Karpathy, Andrej. 2015. *The Unreasonable Effectiveness of Recurrent Neural Networks*. [s.l.]. Disponível em: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>. Acesso em: 15 de abril de 2021.

19. Kim, Sung. 2016. *Multi-layer Recurrent Neural Networks (LSTM, RNN) for word-level language models in Python using TensorFlow*. Disponível em <https://github.com/hunkim/word-rnn-tensorflow>. Acessado em 25 de fevereiro de 2019.

20. Liang, Feynman. 2016. *BachBot - Automatic composition in the style of Bach Chorales*: Developing, analysing and evaluating a deep LSTM Model for musical style. Tese (Mestrado) – University of Cambridge, Cambridge, 2016.

21. Micollis, Ana *et al*. 2021. Composição algorítmica de progressões harmônicas ao estilo de Antonio Carlos Jobim através de processos markovianos. *Musica Theorica*, v. 6, n. 1, p. 218–238.

22. Siphocly, Nermin; El-horatby, El-Sayed; Salem, Abdel-Badeeh. 2021. Applications of Computational Intelligence in Music Composition. *International Journal of Intelligent Comnputing and Information Science*, v. 2, n. 1, p. 59–67. Disponível em: <https://ijicis.journals.ekb.eg/article_156586_a45730bfbcec74f268b8e6fd2d1f769d.pdf>.

23. Strobelt, Hendrik *et al*. 2018. LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks. *IEEE Transactions on Visualization and Computer Graphics*, v. 24, p. 667–676.

24. Sturm, Bob L. *et al*. 2019. Machine learning research that matters for music creation: A case study**.** *Journal of New Music Research*, v. *48*, n. 1, p. 36–55.

25. Sturm, Bob. L.; Santos, João Felipe; Korshunova, Iryna. 2015. *Folk Music Style Modelling by Recurrent Neural Networks with Long Short Term-Story Units*.

Abstracts of the 16th Internacional Society for Music Information Retrieval Conference.

26. Sturm, Bob, L. 2018. *What do these 5,599,881 parameters mean? An analysis of a specific LSTM music transcription model, starting with the 70,281 parameters of its softmax layer*. International Conference on Computational Creativity.
27. Tymockzo, Dmitri. 2011. *A Geometry of Music*. Oxford: Oxford University Press.
28. Tagg, Philip. 2018. *Everyday Tonality II*. New York: The Mass Media Music Scholars' Press.
29. Yang, Li-Chia e Lerch, Alexander. 2020. On The Evaluation of generative models in music. *Neural Computing & Applications*, v. 32, n. 9, p. 4773–4784.