

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CURSO DE ENGENHARIA DE COMPUTAÇÃO

LUCAS FRANCESCO PICCIONI COSTA

**HARMONIZAÇÃO MUSICAL AUTOMÁTICA BASEADA EM REDES
NEURAIS ARTIFICIAIS**

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO
2019

LUCAS FRANCESCO PICCIONI COSTA

**HARMONIZAÇÃO MUSICAL AUTOMÁTICA BASEADA EM REDES
NEURAIS ARTIFICIAIS**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Computação da Universidade Tecnológica Federal do Paraná - UTFPR Campus Toledo, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Andrés Eduardo Coca Salazar
Universidade Tecnológica Federal do Paraná

TOLEDO
2019



Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Toledo
Coordenação do Curso de Engenharia de Computação



TERMO DE APROVAÇÃO

Título do Trabalho de Conclusão de Curso Nº 06

Harmonização Musical Automática Baseada em Redes Neurais Artificiais

por

Lucas Francesco Piccioni Costa

Esse Trabalho de Conclusão de Curso foi apresentado às **11h15 do dia 27 de novembro de 2019** como **requisito parcial** para a obtenção do título de **Bacharel em Engenharia de Computação**. Após deliberação da Banca Examinadora, composta pelos professores abaixo assinados, o trabalho foi considerado **APROVADO**.

Prof. Dr. Andrés Eduardo Coca Salazar
Orientador - UTFPR-TD

Prof. Dr. Gustavo H. Paetzold
UTFPR-TD

Prof. Dr. Jefferson G. Martins
UTFPR-TD

Prof. Dr. Ricardo T. Oliveira
UTFPR-TD

O termo de aprovação assinado encontra-se na coordenação do curso

Toledo, 27 de novembro de 2019

RESUMO

COSTA, Lucas. Harmonização Musical Automática Baseada em Redes Neurais Artificiais. 2019. 54 f. Trabalho de Conclusão de Curso – Curso de Engenharia de Computação, Universidade Tecnológica Federal do Paraná. Toledo, 2019.

Em termos musicais, harmonia define-se como a arte de combinar diversas notas musicais simultaneamente de forma a criar um som coerente aos ouvidos humanos, com objetivo de acompanhar, emoldurar e preencher. Todavia, a elaboração de uma harmonia para determinada melodia não é uma tarefa trivial, uma vez que requer conhecimento e experiência musical, além de intenso estudo da área de teoria musical, algo que demanda tempo, podendo até levar anos para alcançar habilidades razoáveis. Sendo assim, faz-se benéfico e proveitoso o desenvolvimento de um sistema capaz de harmonizar automaticamente melodias. Esforços e tentativas vêm sendo realizadas na solução do problema de harmonias, conforme o método proposto por Koops, Magalhães e Haas (2013), o qual gera acordes para cada nota de uma melodia dada e seleciona a melhor sequência, combinando-a à melodia de entrada em um único arquivo de saída. No presente projeto, foi proposta a construção de uma ferramenta para a harmonização de melodias predefinidas usando técnicas de inteligência artificial para obter resultados que sejam realistas, coerentes com as regras da teoria da harmonia e que apresentem diversidade. Dessarte, uma melodia definida com dados musicais simbólicos e devidamente processada serve de entrada para uma rede neural artificial que gera para esta acordes, tendo sido treinada com base em um banco de dados de músicas contendo melodia e harmonia. A saída do sistema é a sequência de acordes a ser aplicada à melodia. Em virtude da complexidade de obtenção e análise de resultados, testes quantitativos e qualitativos foram realizados, utilizando desde medidas de performance calculadas a partir de matrizes de confusão até pesquisa de enquete *online* com exposição da opinião de participantes voluntários anônimos. Ao final, pode-se concluir que o sistema foi capaz de gerar harmonias alinhadas com os objetivos.

Palavras-chave: Teoria Musical. Sistemas Inteligentes. Avaliação de Performance.

ABSTRACT

COSTA, Lucas. Automatic Musical Harmonization Based on Artificial Neural Networks. 2019. 54 f. Trabalho de Conclusão de Curso – Curso de Engenharia de Computação, Universidade Tecnológica Federal do Paraná. Toledo, 2019.

In musical terms, harmony is defined as the art of combining several musical notes simultaneously to create a coherent sound to human ears, with the purpose of accompanying, framing and filling. However, the elaboration of a harmony for a particular melody is not a trivial task, since it requires music knowledge and experience, as well as intense study of music theory, which takes time and may even take years to reach reasonable skills. Thus, the development of a system capable of automatically harmonizing melodies is beneficial. Efforts and attempts have been made in solving the harmonization problem, as shown in the method proposed by Koops, Magalhães e Haas (2013), which generates chords for each note of a given melody and selects the best sequence by matching it to the input melody in one. single output file. In the present project, it was proposed to build a tool for the harmonization of predefined melodies, using artificial intelligence techniques with results that are realistic, consistent with harmony theory rules and in order to present diversity. Thus, a melody defined with symbolic and properly processed musical data serves as input to an artificial neural network generating for this chords, having been trained based on a music database containing melody and harmony. The system output is the chord sequence to be applied to the melody. Due to the complexity of obtaining and analyzing results, quantitative and qualitative tests were performed, using from performance measures calculated from confusion matrices to online polling with anonymous volunteer participants opinion. In the end, it can be concluded that the system was able to generate harmonies aligned with the objectives.

Keywords: Music Theory. Intelligent Systems. Performance Evaluation.

LISTA DE FIGURAS

Figura 1 – Quadro comparativo da divisão proporcional dos valores das figuras rítmicas desde a semibreve até a semifusa.	6
Figura 2 – As sete notas musicais representadas na pauta usando a clave de sol e os seus respectivos nomes e intervalos de T e S.	7
Figura 3 – Os sete acordes musicais disponíveis a partir da escala de Dó maior e suas respectivas cifras e graus.	9
Figura 4 – Tríades dispostas no círculo de terças, rotuladas por funções harmônicas. . .	10
Figura 5 – Representação de um neurônio com n entradas (x_i) e pesos (w_i), e uma saída (y) com limiar θ	14
Figura 6 – Um perceptron com n neurônios na camada de entrada e um neurônio na camada de saída.	15
Figura 7 – Quatro diferentes funções de ativação usadas nas RNAs, sendo elas: (a) função linear, (b) função de rampa, (c) função passo e (d) função sigmoidal.	15
Figura 8 – Rede Neural Recorrente em forma de laço e desenrolada para entradas (x_t) e saídas (h_t) sequenciais.	16
Figura 9 – Estrutura da unidade interna A de uma (a) RNR simples e de uma (a) LSTM com suas respectivas camadas, funções de ativação e operações internas. . .	17
Figura 10 – Demonstrativo visual dos tipos de redes neurais citadas no texto, sendo elas: perceptron, <i>feedforward</i> (FF), redes neurais recorrente (RNR) e convolucional (RNC).	17
Figura 11 – Diagrama de blocos ilustrativo sobre a metodologia proposta para a harmonização automática de melodias.	18
Figura 12 – Trecho da música <i>Amazing Grace</i> , de John Newton, em notação musical. . .	19
Figura 13 – Formato dos vetores de (a) notas e (b) acordes codificados segundo o método <i>one-hot</i> (HARRIS; HARRIS, 2013).	22
Figura 14 – Visão geral do sistema proposto com processamento da melodia em vetores de notas servindo de entrada para a RNR, cuja saída é um vetor de acordes que posteriormente é decodificado.	23
Figura 15 – Modelagem interna genérica da RNA utilizada, para um conjunto de q vetores de notas de um compasso.	24
Figura 16 – Gráfico de acurácia de teste (barra) e de treino (linha superior) para valores diferentes de lotes e neurônios para RNR_2 e Se_4	29
Figura 17 – Valores de erro de treino e teste calculados usando a função de custo de entropia cruzada (Equação (4)) para todos os modelos da primeira fase de testes.	30
Figura 18 – Valores de erro de treino e teste com alteração do valor de <i>dropout</i> para 0.5.	31

Figura 19 – Acurácia de teste (barra) e treino (linha superior) dos nove melhores modelos selecionados da segunda fase de testes com <i>dropout</i> em 0.5.	31
Figura 20 – Desenvolvimento da acurácia de treino e de teste (em %) para os melhores modelos de (a) RNR ₁ , (b) RNR ₂ e (c) RNR ₃ escolhidos da terceira fase de testes com 150 épocas de treinamento.	32
Figura 21 – Trecho de oito compassos da música <i>December, 1963 (Oh What a Night)</i> de <i>Frankie Valli & The Four Seasons</i> com harmonização original e gerada por cada um dos 3 melhores modelos encontrados.	34
Figura 22 – Matrizes de confusão para os três modelos propostos: (a) RNR ₁ , (b) RNR ₂ e (c) RNR ₃ , com porcentagem de acertos referenciadas por cores.	35
Figura 23 – F_1 -score de classes individuais de acordes para cada um dos modelos propostos.	36
Figura 24 – Resultados da enquete <i>online</i> com a porcentagem de escolha para cada música e a média de escolhas feita por cada grupo: experiente e não experiente, onde barras inferiores indicam a preferência pela harmonia automática e barras superiores, harmonia original.	37
Figura A.1 – Tela inicial do formulário de pesquisa <i>online</i> com apresentação e primeira pergunta.	45
Figura A.2 – Segunda tela do formulário de pesquisa com disposição dos trechos de música audíveis para seleção do usuário.	46
Figura A.3 – Tela final do formulário <i>online</i> indicando a conclusão da pesquisa.	46
Figura B.1 – Partitura com cifra da música <i>Alabama Waltz</i> de Hank Williams analisada na Seção 6.4.2.	47
Figura C.1 – Carta de confirmação recebida pela submissão do artigo à revista.	48

LISTA DE TABELAS

Tabela 1 – Principais escalas musicais, estrutura de tons e semitons, e exemplo de construção começando em dó.	7
Tabela 2 – Nomes dos intervalos, forma de representação, quantidade de semitons que contêm e exemplo de aplicação com relação à nota Dó.	8
Tabela 3 – Modos possíveis para as tríades, a regra para sua composição e exemplo montado a partir da nota Dó.	8
Tabela 4 – Exemplificação da construção da matriz de confusão binária.	12
Tabela 5 – Equações para cálculo das taxas de performance sensibilidade (S), especificidade (E), precisão (P) e F_β -score a partir da matriz de confusão nos casos simples e multiclasse.	13
Tabela 6 – Conteúdo do trecho da música <i>Amazing Grace</i> , de John Newton, da Figura 12, dentro do banco de dados <i>CSV Leadsheets</i>	19
Tabela 7 – Padrão dos valores de duração das notas do banco de dados.	20
Tabela 8 – Classes de acordes consideradas para o processamento do banco de dados. .	21
Tabela 9 – Quantidade máxima de notas por compasso para cada forma de seleção como definido na Subseção 5.2.2.	28
Tabela 10 – Configurações finais selecionadas empiricamente para cada modelo de RNR testado.	33
Tabela 11 – Macro-média das taxas de performance (M) e valores de acurácia para os modelos propostos.	34
Tabela 12 – Valores de média (μ) e desvio padrão (σ) para a preferência sobre músicas harmonizadas automaticamente para os dois grupos analisados: 1) sem e 2) com experiência musical.)	38

LISTA DE ABREVIATURAS E SIGLAS

AG	Algorítmo Genético
BLSTM	<i>Bidirectional Long Short-Term Memory</i>
CSV	<i>Comma Separated Values</i>
DFF	<i>Deep Feedforward</i>
FFN	<i>Feedforward Networks</i>
IA	Inteligência Artificial
LSTM	<i>Long Short-Term Memory</i>
MIDI	<i>Musical Instrument Digital Interface</i>
MCP	McCulloch e Pitts
MIR	<i>Music Information Retrieval</i>
MLP	<i>Multilayer Perceptron</i>
RNA	Rede Neural Artificial
RNC	Rede Neural Convolucional
RNR	Rede Neural Recorrente
R&B	<i>Rhythm and Blues</i>

LISTA DE SÍMBOLOS

T	Distância de tom
S	Distância de semiton
Tm	Distância de um tom e meio
S	Sensibilidade ou taxa de verdadeiros positivos
E	Especificidade ou taxa de verdadeiros negativos
P	Precisão ou valor preditivo positivo
F_{β} -score	Medida de desempenho F com precisão e sensibilidade ponderadas por β .
VP	Verdadeiro positivo
FN	Falso negativo
FP	Falso positivo
VN	Verdadeiro negativo
M	Macro-média de taxas de performance
x_n	Entradas de um neurônio
w_n	Pesos entre as conexões de um neurônio
y	Saída de um neurônio
θ	Limiar da função de ativação
α	Constante real
γ	Indicador de limite superior ou inferior
Ps	Parâmetro de determinação da suavidade
d	Duração de uma nota musical
Fc	Fórmula de compasso
Se_i	Formas de seleção dos conjuntos de treino

SUMÁRIO

1	INTRODUÇÃO	1
2	OBJETIVOS	3
2.1	OBJETIVO GERAL	3
2.2	OBJETIVOS ESPECÍFICOS	3
3	JUSTIFICATIVA	4
4	REFERENCIAL TEÓRICO	5
4.1	TEORIA GERAL DA MÚSICA	5
4.1.1	Teoria Harmônica Tonal Funcional	7
4.2	ESTADO DA ARTE SOBRE A HARMONIZAÇÃO AUTOMÁTICA DE MELODIAS	11
4.3	MEDIDAS DE DESEMPENHO	12
4.4	REDES NEURAIS ARTIFICIAIS	13
5	MATERIAIS E MÉTODOS	18
5.1	DETALHAMENTO DO BANCO DE DADOS UTILIZADO	18
5.2	PROCESSAMENTO E PREPARAÇÃO DAS INFORMAÇÕES DO BANCO DE DADOS	19
5.2.1	Padronização de Informações Melódico-Harmônicas	19
5.2.2	Preparação e Codificação dos Conjuntos de Treino	21
5.3	PREPARAÇÃO DA REDE NEURAL ARTIFICIAL	22
5.3.1	Modelagem e Configuração da RNA	22
5.3.1.1	Camadas Usadas	23
5.3.2	Descrição da Etapa de Treino da RNA	25
5.3.3	Predição de Novos Valores	26
5.4	TESTES QUANTITATIVOS	26
5.5	TESTES QUALITATIVOS	26
5.5.1	Enquete <i>Online</i>	26
5.5.2	Análise Técnica da Produção Musical do Sistema	26
5.6	REFERENCIAL EXPERIMENTAL	27
6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	28
6.1	SELEÇÃO DE INFORMAÇÕES DO BANCO DE DADOS	28
6.2	TREINO DOS MODELOS DE RNA	28
6.3	RESULTADOS QUANTITATIVOS	33

6.4	RESULTADOS QUALITATIVOS	35
6.4.1	Interpretação das Respostas da Enquete <i>Online</i>	36
6.4.2	Resultado da Análise Técnica da Produção Musical do Sistema	37
7	CONCLUSÃO	39
	REFERÊNCIAS	40
	APÊNDICES	44
	APÊNDICE A TELAS DO FORMULÁRIO DE PESQUISA <i>ONLINE</i>	45
	APÊNDICE B PARTITURAS E CIFRAS DA MÚSICA <i>ALABAMA WALTZ</i> DE HANK WILLIAMS	47
	APÊNDICE C CARTA DE SUBMISSÃO E ARTIGO SUBMETIDO À REVISTA <i>JOURNAL OF MATHEMATICS AND MUSIC</i>	48

1 INTRODUÇÃO

Entende-se harmonia como a disposição bem-ordenada entre as partes de um todo, significando concórdia, ordem, proporção ou simetria. Ainda, segundo Bueno (2007), harmonia é a consonância ou sucessão agradável de sons.

Tendo por alvo a teoria da música e somando as definições previamente encontradas, pode-se entender o desenvolvimento de harmonia como a arte de encadear diversos conjuntos de frequências sonoras ou alturas que, por sua vez, quando reproduzidas simultaneamente formam os chamados acordes, de forma a criar um som coerente e de sentido aos ouvidos humanos, comumente atrelada a uma melodia ou linha musical, servindo de propósito a acompanhar e emoldurar (GUEST, 2010).

Não obstante, construir a harmonia que servirá de acompanhamento para uma melodia é um trabalho que requer experiência em assuntos de música, envolvendo conhecimento aprofundado sobre teoria musical, mais especificamente sobre teoria de harmonia (CHEDIAK, 1986). Observando a harmonia do ponto de vista de um processo, verifica-se harmonização como a busca de uma harmonia para determinada melodia, sendo que, para esta, várias opções existem, cuja quantidade é geralmente limitada por conjuntos de regras e convenções estéticas (NOUGUÉ, 2011).

Segundo Makris, Kayrdis e Sioutas (2013), a harmonização é extensão e aplicação natural da análise harmônica. A pesquisa dos passos que constituem tal processo, bem como uma maneira de os automatizar, apresenta uma contribuição valiosa para o estudo da música, inteligência artificial e o campo da Recuperação de Informação Musical (MIR, *Music Information Retrieval*) (KOOPS; MAGALHÃES; HAAS, 2013).

Para o problema de harmonização com objetivo de acompanhamento, definida como harmonia funcional e por vezes denotada como popular, a abordagem aplicada é a de gerar acordes preocupando-se com a função que cada acorde representa durante o desenvolvimento harmônico, com estruturação flexível principalmente no que diz respeito às regras de condução de vozes, sendo que não necessariamente deve acompanhar as notas da melodia (SANTOS, 2012), diferentemente da harmonia tradicional. Esta, por sua vez, tem por tarefa a harmonização de quatro partes onde constrói-se uma sequência de acordes condizente com a melodia ou as vozes (MAKRIS; KAYRDIS; SIOUTAS, 2013).

Nesse sentido, diversas tentativas de solução automatizada vêm sendo realizadas, com algoritmos frequentemente focados em métodos evolutivos usualmente ligados a Algoritmos Genéticos (FREITAS; GUIMARÃES; RUELA, 2011), ou sistemas baseados em regras (EB-CIOGLU, 1988). Por sua vez, Wiggins et al. (1998) apresentaram uma discussão sobre o uso de Algoritmos Genéticos para dois aspectos, dos quais um deles é a geração de harmonias a partir de uma melodia predefinida, tendo por interesse a simulação de comportamento humano na atividade composicional, além da obtenção de resultados musicais. Por outro lado, Koops,

Magalhães e Haas (2013) definiram um sistema baseado em regras existentes de harmonia com foco em aspectos e detalhes de teoria musical. Por fim, Phon-Amnuaisuk e Wiggins (1999) realizaram uma comparação de desempenho entre o uso de sistemas de regras e Algoritmos Genéticos, trazendo como conclusão que a qualidade do resultado depende da quantidade de conhecimento que o sistema possui sobre o tema.

Ao ter contato com estes estudos, é possível verificar que métodos generativos recentes, como Redes Neurais e *Deep Learning*, também foram empregados na tentativa de solução deste problema. Os trabalhos de Huang e Wu (2016), Dong et al. (2017) e Liu e Yang (2018) são exemplos que usam tais metodologias, porém com o objetivo de gerar músicas inteiras, desde a melodia e a harmonia, até o arranjo para a execução.

Neste trabalho, foi desenvolvida uma ferramenta com foco na solução singular do problema de harmonização de melodias musicais usando técnicas de inteligência artificial, mais especificamente de aprendizado de máquina com uso de Redes Neurais Artificiais (RNAs), aplicando Redes Neurais Recorrentes (RNR) do tipo *Long Short-Term Memory* (LSTM). Um banco de dados chamado *CSV Leadsheets Database* (LIM; RHYU; LEE, 2017) contendo músicas compostas de melodia e harmonia forneceu dados de treino para uma RNA. Após a fase de treino, essa rede foi apta para gerar harmonias tendo por referência uma melodia, sendo a entrada do sistema. A nova harmonia foi gerada pelo sistema com base no aprendizado adquirido das músicas anteriormente analisadas e a saída do sistema foi uma harmonia para a melodia. Medidas estatísticas e de performance foram calculadas sobre os resultados para descrever a qualidade da produção, bem como testes subjetivos auditivos envolvendo participantes humanos na resposta de uma enquete *online* para então concluir sobre o nível de desempenho do sistema.

Este trabalho está organizado da seguinte forma: os Capítulos 2 e 3, respectivamente, estabelecem os objetivos geral e específicos, e então é justificada a concepção destes. No Capítulo 4, encontra-se a revisão da literatura existente, e no Capítulo 5 são abordados os materiais e métodos que foram utilizados no desenvolvimento deste trabalho, abrangendo aspectos da teoria musical e de harmonia, métodos estatísticos para realização de testes e obtenção de resultados e aspectos teóricos de inteligência artificial com foco em RNAs. Resultados são expostos no Capítulo 6 e no Capítulo 7 estão as conclusões. Por fim, são apresentadas as referências usadas ao longo deste trabalho.

2 OBJETIVOS

Para a resolução do problema deste estudo, objetivos a serem alcançados foram definidos, divididos em um objetivo principal e três objetivos específicos.

2.1 OBJETIVO GERAL

Desenvolver um sistema para harmonização de melodias musicais automaticamente por meio de algoritmos de Redes Neurais, levando em conta a análise de harmonias existentes de músicas populares.

2.2 OBJETIVOS ESPECÍFICOS

- Produzir um sistema capaz de gerar automaticamente harmonias que sejam realistas;
- Alcançar resultados coerentes com as regras da teoria harmônica tonal;
- Conceber harmonias que apresentem diversidade composicional.

3 JUSTIFICATIVA

Ao iniciar os estudos de música, iniciantes se deparam frequentemente com dificuldades em tarefas de harmonização, dada a necessidade de conhecimento específico e experiência. Um sistema que fosse capaz de realizar essa tarefa seria de grande valia no auxílio dos primeiros trabalhos destes estudantes. Além disso, tal sistema poderia revelar novidades de ideias a músicos profissionais e experientes, concedendo aportes de criatividade na elaboração de um trabalho musical. Sendo assim, músicos iniciantes e profissionais, compositores, arranjadores, estudantes de música e entusiastas seriam os principais favorecidos.

Dessarte, o desenvolvimento de tal sistema também é importante para o campo da engenharia de computação, uma vez que oferece a oportunidade da descoberta de novas aplicações, metodologias, análises e utilidades nas áreas de IA, bem como para a teoria musical, expandindo a teoria para abranger aspectos do mundo digital.

4 REFERENCIAL TEÓRICO

Neste capítulo, é feita uma breve introdução à teoria da música com maior foco em harmonia, bem como levantamento do estado da arte do tema. Também são providas explicações sobre medidas de desempenho aplicadas à análise de resultados e métodos de inteligência artificial que foram usados.

4.1 TEORIA GERAL DA MÚSICA

A música é constituída de três elementos principais, sendo eles (MED, 1996):

- **Melodia:** sons, alturas ou notas, organizadas de forma sucessiva, geralmente responsáveis pela caracterização da música, sendo por vezes o elemento mais marcante da obra;
- **Harmonia:** conjuntos de notas dispostas simultaneamente que emolduram e dão sentido à melodia;
- **Ritmo:** ordenação e organização temporal dos sons da música, tanto melódicos como harmônicos.

As notas pelas quais se dão qualquer música correspondem com sons organizados em sistemas de afinação, sendo o Sistema Temperado o mais comumente aplicado nas músicas ocidentais (NOUGUÉ, 2011). Os nomes pelos quais se conhecem estas notas são, respectivamente: Dó, Ré, Mi, Fá, Sol, Lá, Si. Estas podem ser representadas utilizando notação musical em um espaço denominado pauta ou pentagrama, um conjunto de cinco linhas.

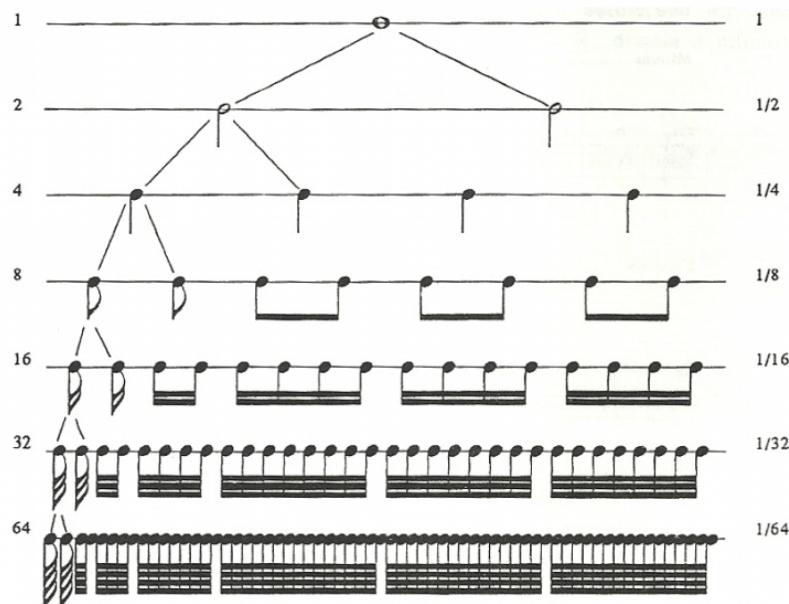
As diferentes formas de uma figura musical são necessárias para a representação de duração de uma nota, fazendo parte da característica rítmica da música e as localizando no tempo. Para organizar as durações na pauta, os tempos são agrupados em espaços denominados compassos, os quais seguem como regra uma fórmula de compasso que é indicada no início da música. A fórmula de compasso é exposta com formato de fração, onde o numerador indica a quantidade de tempos de um compasso e o denominador indica a figura que tem valor de um tempo (CHEDIAK, 1986). Sendo assim, a duração que cada figura rítmica representa é descrita a seguir, para as quatro primeiras figuras (MED, 1996):

- **Semibreve (.)**: figura de maior duração possível, sendo que as demais figuras são derivações desta;
- **Mínima (♩)**: tem o valor de metade da semibreve, ou seja, quando duas mínimas são somadas, equivalem a uma semibreve;
- **Semínima (♪)**: vale a metade da mínima; quatro semínimas equivalem a uma semibreve, e uma mínima tem duração igual a soma de duas semínimas;
- **Colcheia (♪)**: é caracterizada por uma bandeirola (colchete), e vale a metade da semínima, seguindo a lógica constituída.

As demais formas e figuras podem ser resumidas pela Figura 1, que mostra um quadro compa-

rativo da quantidade de figuras que representam a anterior, a começar pela semibreve. Pode-se notar então que cada figura tem a metade da duração da sua superior.

Figura 1 – Quadro comparativo da divisão proporcional dos valores das figuras rítmicas desde a semibreve até a semifusa.



Fonte: Med (1996)

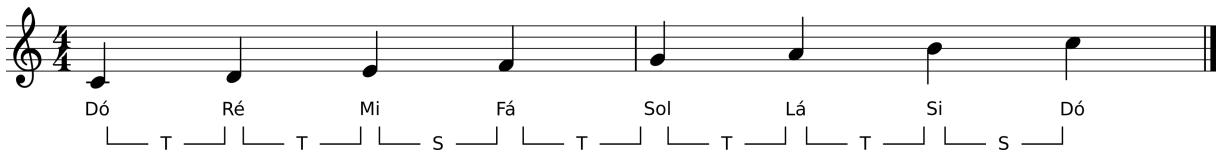
Para referenciar o nome de cada nota na pauta, é preciso ainda outro elemento, a clave, havendo para esta três representações: clave de sol ($\text{\(\textcircled{G}\)}$), clave de fá ($\text{\(\textcircled{A}\)}$) e clave de dó ($\text{\(\textcircled{B}\)}$). Na Figura 2, as notas estão referenciadas pela clave de sol. É possível perceber que a clave dá indicativo de a nota sol estar na segunda linha de baixo para cima, e a partir dela é possível nomear as outras notas usando espaços e linhas (MED, 1996).

As escalas musicais são sequências ordenadas de notas, de acordo com regras de distância, ou intervalos entre notas. A distância entre notas musicais é medida em termos de tons (T) e semitons (S), sendo que um semitom, segundo Med (1996) é definido como "o menor intervalo adotado entre duas notas na música ocidental (no sistema temperado de 12 notas)". Também é útil definir que a soma de um semitom com um tom equivalem a um tom e meio (Tm). São muitas as variações possíveis para organizar as notas em escalas (COCA; ZHAO, 2017), das quais três são principais: cromática, maior e menor natural, esta última dvida em harmônica e melódica (MED, 1996). A Tabela 1 contém informações sobre a construção em intervalos de T, S e Tm destas escalas, bem como um exemplo.

A escala cromática, em especial, contém todas as 12 notas existentes no Sistema Temperado, sendo elas (usando notação sustenida¹): Dó, Dó♯, Ré, Ré♯, Mi, Fá, Fá♯, Sol, Sol♯,

¹Um sustenido (♯) é classificado como um acidente musical, elevando a nota por ele referenciada em um semitom, como por exemplo: Dó♯ fica entre Dó e Ré. Outra forma de se referir a um acidente é utilizando um bemol

Figura 2 – As sete notas musicais representadas na pauta usando a clave de sol e os seus respectivos nomes e intervalos de T e S.



Fonte: Autoria Própria. Adaptado de Med (1996)

Lá, Lá[#] e Si. Por serem compostas de todas as notas, não há definição de tonalidade.

A escala maior se baseia no uso de sete notas diatônicas. Na Figura 2 está ilustrada a escala maior de Dó, com intervalos entre notas indicados. Importante entender que, na mudança da nota inicial, as distâncias entre notas são mantidas para a construção da escala. Na verdade, são as distâncias entre as notas que de fato caracterizam a forma da escala.

A escala menor possui três formas de construção básica. A primeira, escala menor natural, é semelhante à escala maior se iniciada a partir da terça menor descendente da escala maior. Pode-se ver para o exemplo referente da Tabela 1, que as notas da escala menor natural são iguais às da maior, muda apenas o ponto de início. Tendo por base a escala menor natural, se aumentada com um sustenido a sétima nota, de forma a criar uma sétima maior, tem-se a menor harmônica. Por fim, a menor melódica é como a menor harmônica, porém com a sexta nota também aumentada na forma ascendente mas sem essa alteração na forma descendente.

Tabela 1 – Principais escalas musicais, estrutura de tons e semitons, e exemplo de construção começando em dó.

Escala	Estrutura de intervalos	Exemplo de construção
Maior	T T S T T T S	Dó Ré Mi Fá Sol Lá Si Dó
Menor natural	T S T T S T T	Lá Si Dó Ré Mi Fá Sol Lá
Menor harmônica	T S T T S Tm S	Lá Si Dó Ré Mi Fá Sol [#]
Menor melódica ascendente	T S T T T T S	Lá Si Dó Ré Mi Fá [#] Sol [#]

Fonte: Autoria Própria. Adaptado de Chediak (1986)

4.1.1 Teoria Harmônica Tonal Funcional

Segundo Schoenberg (1978), "harmonia é o estudo de sons simultâneos (acordes) e de como eles podem ser unidos em relação a seus valores arquitetônicos, melódicos e rítmicos; e seu significado, e seu peso relativo um ao outro". O estudo da harmonia estabelece regras precisas e conceitos importantes, dos quais alguns serão apresentados a seguir.

Uma forma comum de construir acordes é fazendo uso de tríades, consistindo de três notas tocadas ao mesmo tempo (acordes de três notas), geralmente feitas da sobreposição de

(>, que possui lógica oposta, diminuindo um semiton. Ainda, o bequadro (♯) é um sinal que indica a anulação de um acidente (sustenido ou bemol).

notas em intervalos de terças. Na Tabela 2 são apresentados os nomes dos intervalos dentro do espaço de uma oitava justa (quando as notas começam a se repetir) e a quantidade de semitonos que os representam, bem como a nota que está com essa distância relativa à nota Dó, como exemplo. Um intervalo de terça pode ser maior ou menor, sendo que o primeiro representa uma distância de três semitonos de uma nota de partida, e o segundo, quatro. Por exemplo, de Dó até Mi, há uma distância de quatro semitonos e consequentemente um intervalo de terça maior. Já entre Dó e Ré \sharp , um intervalo de três semitonos, uma terça menor. Dependendo da qualidade dos intervalos de terças entre as notas de uma tríade, poderá chamar o acorde de maior, menor, aumentado ou diminuto. A Tabela 3 mostra as possibilidades de modos de acordes disponíveis e a definição dos mesmos para a tônica² (1^a) e suas respectivas terça (3^a) e quinta (5^a), ambas podendo ser maiores (M) ou menores (m), e a 5^a também aumentada (A) ou diminuta (D).

Tabela 2 – Nomes dos intervalos, forma de representação, quantidade de semitonos que contêm e exemplo de aplicação com relação à nota Dó.

Nome do Intervalo	Representação	Quantidade de Semitonos	Nota relativa de Dó
Segunda menor	2 ^a m	1	Dó \sharp
Segunda maior	2 ^a M	2	Ré
Terça menor	3 ^a m	3	Mi \flat
Terça maior	3 ^a M	4	Mi
Quarta justa	4 ^a J	5	Fá
Quinta diminuta	5 ^a D	6	Sol \flat
Quinta justa	5 ^a J	7	Sol
Quinta aumentada	5 ^a A	8	Sol \sharp
Sexta maior	6 ^a M	9	Lá
Sétima menor	7 ^a m	10	Sí \flat
Sétima maior	7 ^a M	11	Sí
Oitava justa	8 ^a J	12	Dó

Fonte: Autoria Própria. Adaptado de Kaizer (2017)

Tabela 3 – Modos possíveis para as tríades, a regra para sua composição e exemplo montado a partir da nota Dó.

Modo	Composição	Exemplo a partir de Dó
Aumentado	1 ^a 3 ^a M 5 ^a A	Dó Mi Sol \sharp
Maior	1 ^a 3 ^a M 5 ^a J	Dó Mi Sol
Menor	1 ^a 3 ^a m 5 ^a J	Dó Mi \flat Sol
Diminuto	1 ^a 3 ^a m 5 ^a D	Dó Mi \flat Sol \flat

Fonte: Autoria Própria. Adaptado de Kaizer (2017)

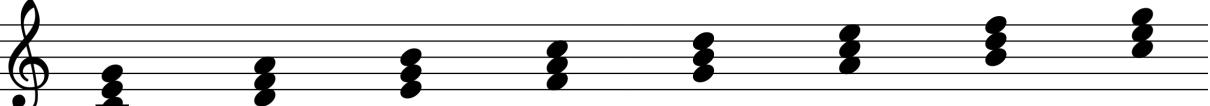
Para se referir ao grupo de acordes que pode ser executado a partir de uma escala, é usada a expressão campo harmônico. Dentro deste, estão contidos sobretudo acordes maiores e

²A nota fundamental de um acorde é chamada de tônica, a qual também é responsável por dar nome ao acorde.

menores que constituem uma relação funcional entre si. A construção de um campo harmônico se dá de forma a obter um acorde para cada nota da escala selecionada, usando a sobreposição de terças e mantendo as notas dos acordes dentro da mesma escala. Por exemplo, para a escala de Dó maior (Figura 2), para cada nota haverá uma tríade. A primeira, sobre a nota Dó, será com as notas Dó, Mi e Sol; para a segunda (Ré), as notas Ré, Fá e Lá, e assim por diante. Os acordes formados a partir da escala maior podem ser representados como na Figura 3. Em uma música elaborada desta maneira e sem modulação, os acordes montados estarão dentro deste campo harmônico e cada acorde terá uma função diferente (KAIZER, 2017).

Outra forma comum para a representação de acordes é a utilização de letras (ou cifras) ao invés do nome das notas, como também ilustrado na Figura 3. Para este sistema, troca-se a nota lá pela letra A (maiúscula), e segue-se sequencialmente até a letra G (Sol). Isso facilita a cifragem, e permite outras convenções, como a representação de acordes menores (letra “m” minúscula). Os números romanos, por sua vez, servem para indicação dos graus dos acordes, generalizando a representação.

Figura 3 – Os sete acordes musicais disponíveis a partir da escala de Dó maior e suas respectivas cifras e graus.

	Dó maior C I	Ré menor Dm II	Mi menor Em III	Fá maior F IV	Sol maior G V	Lá menor Am VI	Si diminuto Bm(b5)7 VII	Dó maior C I
--	--------------------	----------------------	-----------------------	---------------------	---------------------	----------------------	-------------------------------	--------------------

Fonte: Autoria Própria. Adaptado de Chediak (1986)

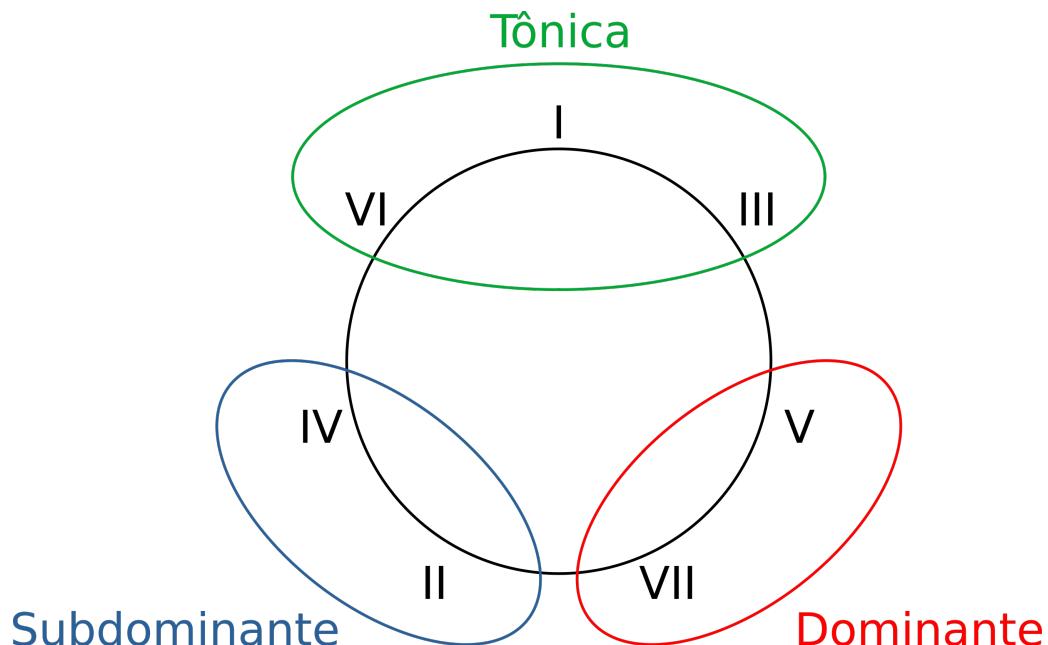
Entende-se por função, em termos de harmonia, a característica de um acorde e seu valor em relação aos demais, no contexto da obra, que é determinado pela somatória relacional de todos acordes com um centro tonal, por onde se estabelece o conceito de tonalidade. A tonalidade é definida, então, pelo conjunto de categorias ou funções tradicionalmente denominadas de tônica, dominante e subdominante, respectivamente. Cada uma pode ter suas funções definidas como (KOELLREUTER, 1978; CHEDIAK, 1986)

- **Tônica:** concede sensação de repouso, tendo funcionalidade de início ou finalização. Sua única representante é o grau I do campo harmônico;
- **Dominante:** dá o sentido de movimento de aproximação, uma vez que gera tensão e necessidade de resolução, geralmente provida pela tônica. É o grau V do campo harmônico;
- **Subdominante:** também relacionado ao movimento, porém de afastamento, um meio termo entre tônica e dominante, sendo representada pelo grau IV.

Para os demais acordes (ou graus) presentes no campo harmônico, suas funções são herdadas das funções principais, chamadas funções secundárias, e estes acordes são então

denominados de relativos. Segundo (KOELLREUTER, 1978): “os acordes (relativos) tem o significado harmônico daquela tônica, subdominante ou dominante, da qual são vizinhos de terça”, ou seja, são acordes diatônicos cujas fundamentais estão a uma distância de terça da função principal. Em linhas gerais, aplica-se o ilustrado na Figura 4 (SHAFFER; HUGHES; MOSELEY, 2014), onde os graus III e VI são relativos da tônica, VII da dominante e II da subdominante. Isso quer dizer que estes graus compartilham da mesma função principal, em nível menos intenso, e podem funcionar como substitutos.

Figura 4 – Tríades dispostas no círculo de terças, rotuladas por funções harmônicas.



Fonte: Shaffer, Hughes e Moseley (2014)

Em especial para o grau V, a adição de uma quarta nota ao acorde, uma terça acima da última, chamada de sétima (podendo ser maior ou, como geralmente, menor), é uma prática comumente empregada. O acorde é então representado como V^7 . Exemplificando, para o acorde de G 7 , as notas que o compõem são: Sol, Si, Ré, Fá. Essa nova nota adiciona um intervalo específico entre a segunda e quarta nota do acorde (Si e Fá) chamado de tritono, “um intervalo de três tons (quarta aumentada ou quinta diminuta) e que representa o som preparatório do acorde dominante” (FARIA, 1991).

Respeitando as funções de cada grau, estabelece-se o movimento padrão dos acordes dentro da música: a função dominante é comumente precedida pela tônica, dando a resolução da dissonância; a função subdominante tem maior liberdade, entretanto sua sensação de movimento normalmente leva à dominante; a função tônica por sua vez é geralmente aplicada no início ou final de uma música ou frase musical, provendo a ideia de conclusão (CHEDIAK, 1986).

A harmonia funcional, portanto, tem por objetivo estudar o funcionamento destas funções, bem como estabelecer regras e conceitos para a sua correta utilização dentro da música. A partir do entendimento das funções é possível harmonizar músicas de forma coerente, que

façam sentido aos ouvintes. Sendo também regras que, quando respeitadas, indicam a qualidade da obra.

4.2 ESTADO DA ARTE SOBRE A HARMONIZAÇÃO AUTOMÁTICA DE MELODIAS

Há décadas, computadores vêm sendo utilizados para a tarefa de composição musical. Com foco na composição automática de música, pode-se entender a harmonização automática de melodias como uma extensão da composição algorítmica, a qual consiste na aplicação de algoritmos de inteligência artificial, especialmente de aprendizado de máquina, durante o processo de concepção de uma obra musical (MAKRIS; KAYRDIS; SIOUTAS, 2013). A harmonização automática possibilita, então, um processo de composição assistida. As técnicas mais empregadas para alcançar tal feito são algoritmos baseados em regras, AGs ou RNAs, nas suas mais diversas formas, com o uso de informações estatísticas e descriptivas extraídas de conjuntos de treinamento, sendo estes normalmente feitos de obras musicais existentes, representadas de maneira simbólica, por exemplo, através do protocolo MIDI³.

Sistemas de composição algorítmica baseados em regras abordam o campo da representação de conhecimento, diferente de outros que se valem geralmente de técnicas de aprendizado de máquina. O conjunto de regras definido é considerado como todo o conhecimento que o algoritmo tem para a produção, semelhante a um estudante que aprende profundamente a teoria, mas não possui experiência. Não obstante, tais sistemas geram bons resultados, como é o caso do trabalho de Ebcioğlu (1988) com o desenvolvimento do sistema CHORAL. Neste sistema, a base de conhecimento algorítmica buscava representar principalmente as regras de composição de harmonias usadas nos corais de Bach. Ao final, o sistema era capaz de produzir níveis aceitáveis de competência, próximas de um "talento estudante de música que tenha estudado os corais de Bach"(EBCIOĞLU, 1988).

Recentemente, Koops, Magalhães e Haas (2013) propuseram um sistema automático para harmonização chamado FHARM, baseado em um modelo de harmonia funcional. O seu diferencial está na garantia de que ao menos regras básicas de harmonia sejam seguidas no resultado final. A partir de dois experimentos, o primeiro com resultados analisados por especialistas em harmonia e o segundo com comparação da harmonia gerada para determinada melodia com a harmonia elaborada por um estudante de música, foi possível confirmar que o sistema gera harmonias realísticas.

Algoritmos Genéticos baseiam-se na ideia do processo evolutivo por seleção natural. Por causa disso, esses algoritmos têm sido amplamente utilizados para a tarefa de harmonização automática. Freitas, Guimarães e Ruela (2011), por exemplo, em seu trabalho, definem duas funções de avaliação (*fitness*), uma de simplicidade e outra de dissonância. Em conjunto, estas caracterizam um processo multiobjetivo que revelou flexibilidade para aplicação ou não de todas as regras definidas, dando uma característica de criatividade ao sistema, ao apresentar

³Protocolo cujo propósito é prover um padrão eficiente para transmissão e recepção de informações musicais, como notas, durações e intensidades, na forma de dados eletrônicos (ASSOCIATION, 2014).

soluções factíveis, algo inalcançável por um sistema somente baseado em regras. Uma importante conclusão a que chegaram foi de que

"... atualmente, a maior parte das condições de avaliação de uma solução consideram apenas o acorde seguinte, no máximo. Seria interessante imaginar sistemas que pudessem analisar os compassos como um todo antes de aplicar operadores genéticos"(FREITAS; GUIMARÃES; RUELA, 2011).

Nesse sentido, (LIM; LEE, 2017) apresentam uma possibilidade de composição de harmonias usando Redes *Bidirectional Long Short-Term Memory* (BLSTM) e treinadas com base em dados simbólicos de harmonias de músicas reais. Tais redes são capazes de refletir sobre o contexto musical nos dois sentidos temporais, tendo em mente que acordes são formados com base na ordem de seleção de notas, tanto à frente quanto passadas. Sendo assim, cada compasso de música era analisado e tinha, então, um acorde gerado correspondente. Comparado com resultados de outros métodos convencionais, ouvintes preferiram as harmonias geradas por este modelo.

4.3 MEDIDAS DE DESEMPENHO

Para poder descrever o desempenho dos resultados gerados de forma qualitativa é preciso usar medidas calculadas a partir das notas e acordes da harmonia. Quanto a tarefa de harmonização a quatro vozes, percebe-se uma quantidade considerável e interessante de medidas caracterizadoras em trabalhos que fazem uso de AGs (DONNELLY; SHEPPARD, 2011). Entretanto, para o problema de harmonização automática de acompanhamento, a preferência é pelo cálculo de acurácia e análise de medidas de performance geral para o sistema final (LIM; LEE, 2017).

A medida de acurácia é padrão e melhor usada para avaliar a igualdade das gerações. Um detalhamento maior pode ser feito ao se analisar a matriz de confusão, também chamada de matriz de erro, que relaciona as previsões e valores verdadeiros em uma tabela onde as linhas são instâncias de uma classe predita, e colunas são instâncias de classes verdadeiras, como ilustrado na Tabela 4 (SOKOLOVA; LAPALME, 2009). Desta matriz, podem ser calculadas diversas outras medidas de desempenho, sendo elas: sensibilidade (S , taxa de verdadeiros positivos), especificidade (E , taxa de verdadeiros negativos) e precisão (P , valor preditivo positivo), dentre outras. Há ainda a medida F_{β} -score, que combina precisão e sensibilidade, sendo β um valor real positivo. É comum definir $\beta = 1$, para que haja ponderamento igual entre as taxas combinadas.

Tabela 4 – Exemplificação da construção da matriz de confusão binária.

Valores Reais	Predição Positiva		Predição negativa
	Verdadeiro Positivo (VP)	Falso Positivo (FP)	
Positivo	Verdadeiro Positivo (VP)	Falso Negativo (FN)	
Negativo	Falso Positivo (FP)	Verdadeiro Negativo (VN)	

Fonte: Autoria Própria. Adaptado de Sokolova e Lapalme (2009)

Para um problema multiclasse, o qual se encaixa a classificação de vários acordes diferentes, a quantidade de taxas é calculada individualmente para cada classe (COCA, 2015). Além disso, a qualidade geral da classificação pode ser verificada pela média das medidas calculadas para cada uma das classes, tratando-as com igualdade, método chamado de macro-média (M) (SOKOLOVA; LAPALME, 2009). Estas equações são definidas na Tabela 5 para l igual a quantidade de classes e usando a macromedia as taxas são denotadas como " $taxa_M$ ".

Tabela 5 – Equações para cálculo das taxas de performance sensibilidade (S), especificidade (E), precisão (P) e F_β -score a partir da matriz de confusão nos casos simples e multiclasse.

Medida	Definição Simples	Macro-Média (M)
S	$\frac{VP}{VP+FN}$	$\frac{\sum_{i=1}^l \frac{VP_i}{VP_i+FN_i}}{l}$
E	$\frac{VN}{VN+FP}$	$\frac{\sum_{i=1}^l \frac{VN_i}{FP_i+VN_i}}{l}$
P	$\frac{VP}{VP+FP}$	$\frac{\sum_{i=1}^l \frac{VP_i}{VP_i+FP_i}}{l}$
F_β -score	$\frac{(\beta^2+1)PS}{(\beta^2P)+S}$	$\frac{(\beta^2+1)P_M S_M}{(\beta^2 P_M) + S_M}$

Fonte: Autoria Própria. Adaptado de Sokolova e Lapalme (2009)

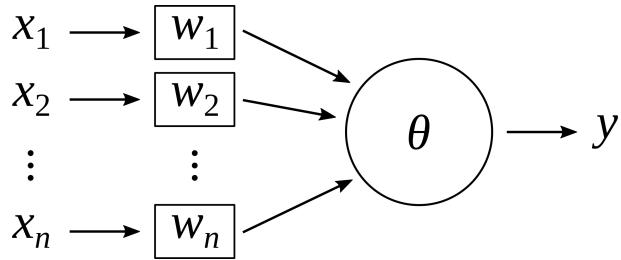
4.4 REDES NEURAIS ARTIFICIAIS

A Inteligência Artificial (IA) é o ramo científico que estuda como desenvolver sistemas inteligentes (ANGLEY; LAIRD, 2006), podendo ser definida sob perspectivas de comportamento e de processo de pensamento, na tentativa de fazer com que máquinas atuem como humanos. São quatro principais estratégias para o estudo de IA, sendo elas: processamento de linguagem natural, raciocínio automatizado, representação de conhecimento e aprendizado de máquina. As duas últimas são normalmente aplicadas quando o objetivo é gerar informações, como no caso da harmonização automática de melodias. Uma técnica popular para elaborar sistemas inteligentes são as RNAs, as quais são descritas a seguir.

Segundo Braga (2000), define-se por RNA: "sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas", sendo que estes nodos são modelos baseados na estrutura funcional de um neurônio biológico. O modelo McCulloch e Pitts (MCP) (MCCULLOCH; PITTS, 1943) é uma descrição matemática do funcionamento de um neurônio, ilustrado na Figura 5. Neste, existem n entradas x_1, x_2, \dots, x_n com os respectivos pesos w_1, w_2, \dots, w_n de valores positivos ou negativos, para uma saída y . O efeito de uma sinapse⁴ i é dado por $x_i w_i$, sendo que os pesos determinam o grau de consideração que o neurônio dará a um sinal de entrada.

⁴A sinapse controla o fluxo de informação entre os nodos de uma rede neural.

Figura 5 – Representação de um neurônio com n entradas (x_i) e pesos (w_i), e uma saída (y) com limiar θ .



Fonte: Autoria Própria. Adaptado de Braga (2000)

A saída de um neurônio é inicialmente um valor Booleano. Para que esta seja verdadeira, é preciso que a soma dos valores das sinapses ultrapasse um certo limiar θ , como descrito pela equação:

$$\sum_{i=1}^n x_i w_i \geq \theta. \quad (1)$$

Essa equação é chamada função de ativação, e determina que a saída do neurônio será ativa se o resultado do somatório ultrapassar o limiar θ . Assim sendo, um neurônio sozinho se torna um classificador linear, ou seja, classificando apenas dados que sejam separáveis linearmente.

A união de dois neurônios para as mesmas entradas, e cujas saídas servem de entrada para um terceiro neurônio, constituem-se o que se chamam perceptrons, e à disposição de neurônios concorrentes é dado o nome de camada, como mostrado na Figura 6. Neste modelo, existem apenas camadas de entrada e uma camada de saída.

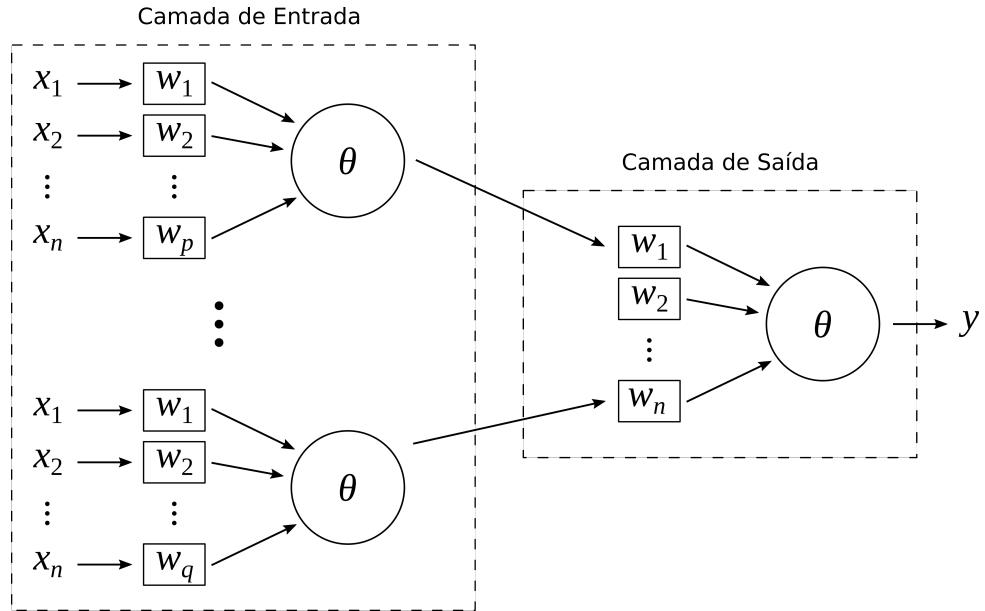
Aumentando a complexidade, realizando conexões de perceptrons em camadas, define-se o termo perceptron multicamadas (MLP, *Multilayer Perceptron*). Este modelo possui uma ou mais camadas de neurônios intermediárias ("escondidas"), e também são conhecidas por FFN (*Feedforward Networks*, que significa algo como "alimentar para frente"). Esse formato retira a limitação de uma única função de ativação, e amplia o leque de possibilidades de modelagem de redes neurais (BRAGA, 2000).

Tendo a função de ativação como a responsável pela resposta de um neurônio, outros modelos foram criados, permitindo outras respostas não Booleanas. É possível exemplificar esse aspecto mostrando outras quatro funções de ativação, exibidas na Figura 7. A função linear (Figura 7(a)) é definida como $y = \alpha x$, para α igual a um número real arbitrário. Se for restrinuida, torna-se uma função de rampa (Figura 7(b)), tendo valores de limite máximo e mínimo. Semelhante à função Booleana, a função passo (Figura 7(c)) produz dois valores, $+\gamma$ e $-\gamma$ para diferentes valores de x . Finalmente, a função sigmoidal (Figura 7(d)) caracteriza-se como uma função semilinear e limitada, definida por:

$$y = \frac{1}{1 + e^{-x/Ps}}, \quad (2)$$

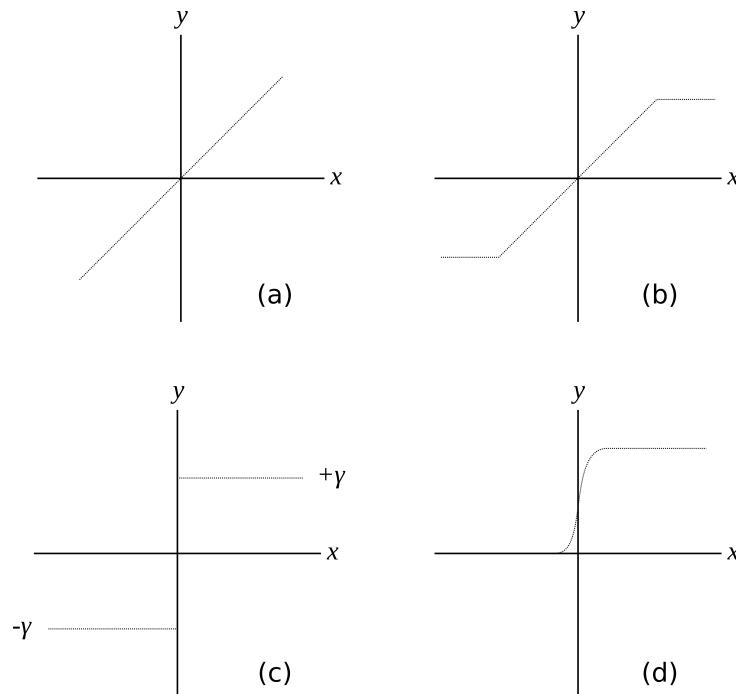
com Ps sendo o parâmetro de determinação da suavidade da curva (BRAGA, 2000).

Figura 6 – Um perceptron com n neurônios na camada de entrada e um neurônio na camada de saída.



Fonte: Autoria Própria. Adaptado de Jeronymo (2018)

Figura 7 – Quatro diferentes funções de ativação usadas nas RNAs, sendo elas: (a) função linear, (b) função de rampa, (c) função passo e (d) função sigmoidal.



Fonte: Autoria Própria. Adaptado de Braga (2000)

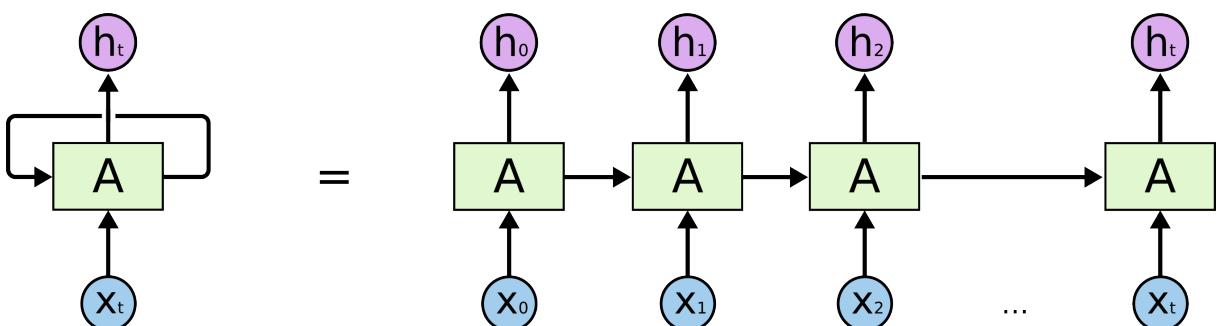
Quando há a existência de duas ou mais camadas escondidas, já é possível classificar o modelo como sendo DFF (*Deep Feedforward*). Nestas redes, o termo *feedforward* significa que a informação de entrada não revisita neurônios ou camadas passadas (GOODFELLOW; BENGIO;

COURVILLE, 2016).

As RNAs são modelos que atendem a estratégia de aprendizado de máquina quando retêm informação. O objetivo primário de tais modelos é poder, após uma fase de treino, realizar previsões e encontrar padrões em grandes quantidades de dados. Um exemplo disso são as redes neurais convolucionais (RNC), um exemplo de rede *feedforward* com alta capacidade descritiva (LIU et al., 2017). Estas, por sua vez, são construídas com uso de células convolucionais, que passam para a próxima camada de células o resultado de um *kernel* de convolução multiplicado para um intervalo relativo de células da camada anterior (VEEN, 2017). O *kernel* usado para a operação de convolução com os dados de entrada é, usualmente, uma matriz ou vetor pequeno de números (LUDWIG, 2017).

Quando redes *feedforward* têm sua funcionalidade estendida a conexões de *feedback*, ou seja, onde a informação pode revisitar neurônios de camadas passadas, uma outra classe de RNAs pode ser introduzida. São as redes neurais recorrentes (RNR), que apresentam ainda a faceta de gerar novos dados, graças à habilidade do modelo de processar dados em sequência, sendo que a produção desta também pode ser uma sequência. Isso se dá pelo fato de que sua estrutura funciona em ciclos, cujo estado interno é atualizado a cada novo dado da entrada. A Figura 8 representa esse comportamento, onde a parte da rede *A* recebe uma entrada x_t e dá a saída h_t . Esta pode ser pensada como se fosse várias cópias de si mesma para entradas sequenciais ($0 \dots t$), levando uma mensagem para seu *A* sucessor. Com isso, tais redes podem a cada passo receber um dado para análise, atualizar seu estado interno e realizar previsões, gerando novos dados baseados na sequência previamente analisada (SUTSKEVER; MARTENS; HINTON, 2011).

Figura 8 – Rede Neural Recorrente em forma de laço e desenrolada para entradas (x_t) e saídas (h_t) sequenciais.

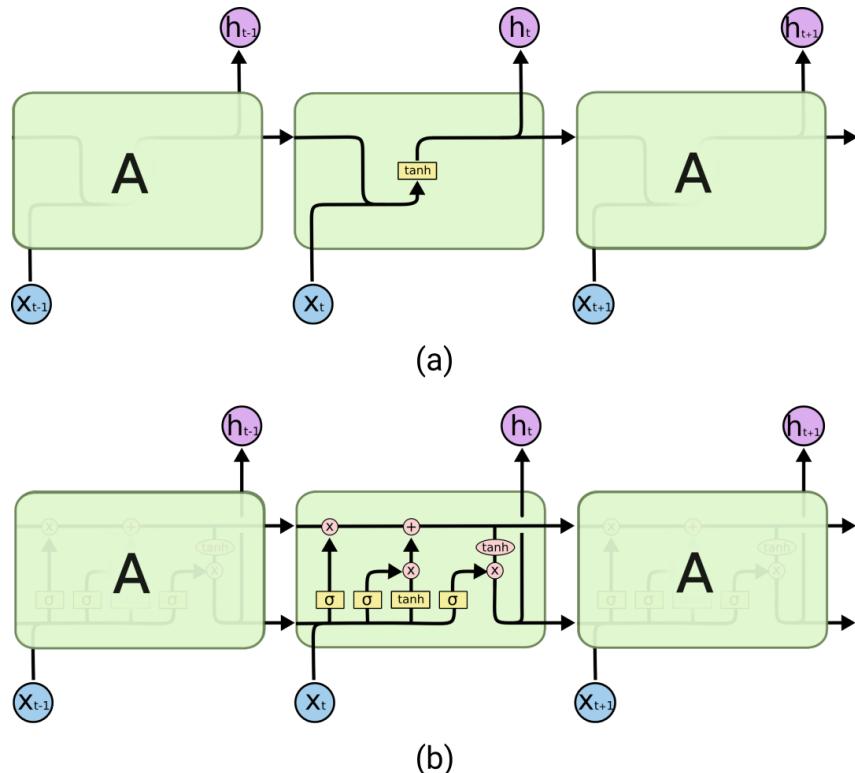


Fonte: Olah (2015)

Um tipo especial de RNR chamado de *Long Short-Term Memory* (LSTM) é capaz de realizar o aprendizado de dependências à longo termo, sendo esta uma das dificuldades da RNR tradicional (HOCHREITER; SCHMIDHUBER, 1997). Isso acontece devido a uma diferença da estrutura interna da unidade de LSTM, como pode-se ver na Figura 9. A Figura 9(a) demonstra como é a unidade padrão da RNR, contendo apenas uma camada ou função de ativação, diferentemente do que se pode ver na Figura 9(b): a unidade de LSTM possui

quatro camadas com diferentes funções de ativação e operações entre as saídas, aumentando sua complexidade (OLAH, 2015).

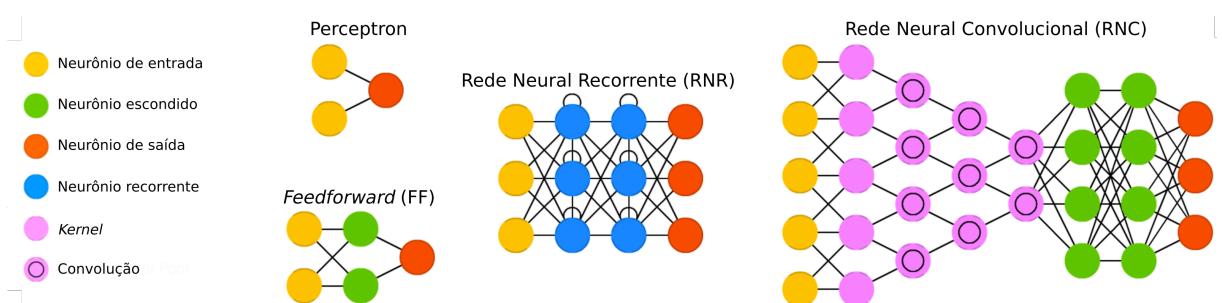
Figura 9 – Estrutura da unidade interna A de uma (a) RNR simples e de uma (a) LSTM com suas respectivas camadas, funções de ativação e operações internas.



Fonte: Olah (2015)

Em suma, é possível ilustrar em termos gerais os comportamentos destas redes através da Figura 10 (VEEN; LEIJNEN, 2019).

Figura 10 – Demonstrativo visual dos tipos de redes neurais citadas no texto, sendo elas: perceptron, *feedforward* (FF), redes neurais recorrente (RNR) e convolucional (RNC).

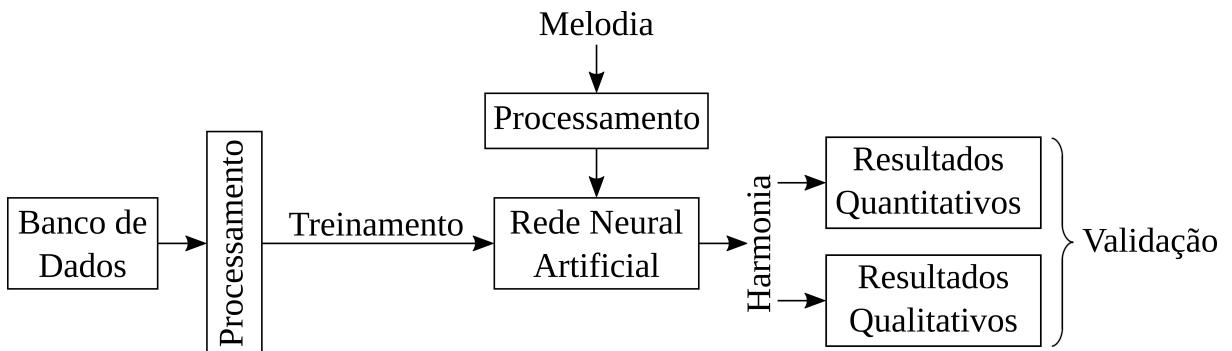


Fonte: Veen e Leijnen (2019)

5 MATERIAIS E MÉTODOS

O método proposto para este trabalho pode ser identificado pelo diagrama de blocos da Figura 11. A partir de um banco de dados de músicas codificadas em dados simbólicos, contendo tanto a melodia como a harmonia, uma RNA será treinada e posteriormente utilizada para gerar novas harmonias condicionadas por uma melodia de entrada, com o objetivo de gerar um acorde por compasso. Esta, por sua vez, será primeiro processada, para então servir de entrada a uma rede neural treinada com base em um banco de dados também processado. O resultado será uma harmonia, a qual será avaliada qualitativa e quantitativamente. A seguir serão descritos com detalhe cada um dos blocos e etapas do diagrama.

Figura 11 – Diagrama de blocos ilustrativo sobre a metodologia proposta para a harmonização automática de melodias.



Fonte: Autoria Própria

5.1 DETALHAMENTO DO BANCO DE DADOS UTILIZADO

Para aplicação da RNA, são necessários dados para efetuar o treinamento. Estes dados consistem de músicas representadas simbolicamente, preferencialmente em formatos como MIDI (ASSOCIATION, 2014), *MusicXML* (GOOD, 2001) ou *Humdrum* (HURON, 2002), devido ao fácil processamento das informações contidas.

Diversos bancos de dados estão disponíveis na Internet, livres de custo e úteis para pesquisas relacionadas à música geralmente voltadas à MIR. Dentre eles, *The NES Music Database* (DONAHUE; MAO; MCAULEY, 2018), "The Largest MIDI Collection on the Internet" (MAN, 2015) e Lakh MIDI (RAFFEL, 2011; RAFFEL, 2016).

Para este trabalho, foi selecionado o banco de dados *CSV Leadsheets Database* (LIM; RHYU; LEE, 2017), que contém 2252 músicas dos gêneros *rock*, *pop*, *country*, *jazz*, *folk*, *R&B*, canções infantis, etc., já processadas no formato CSV, cujos arquivos contêm organizadas informações sobre tempo, compasso, tônica e modo do acorde tocado no momento, nota da melodia, sua duração, entre outras informações para cada música, tornando-se interessante para

pesquisas em harmonia. Um exemplo de parte do conteúdo de uma das músicas deste banco de dados, representada em notação musical na Figura 12, pode ser visto na Tabela 6, para um trecho da música *Amazing Grace* de John Newton.

Figura 12 – Trecho da música *Amazing Grace*, de John Newton, em notação musical.



Fonte: Autoria Própria. Adaptado de John Newton

Tabela 6 – Conteúdo do trecho da música *Amazing Grace*, da Figura 12, dentro do banco de dados CSV *Leadsheet*.

Tempo	Compasso	Tônica do acorde	Tipo do acorde	Nota da melodia	Duração da nota
3/4	1	Sem acorde	Sem acorde	Pausa	8,0
3/4	1	Sem acorde	Sem acorde	Ré	4,0
3/4	2	Sol	Maior	Sol	8,0
3/4	2	Sol	Maior	Si	2,0
3/4	2	Sol	Maior	Sol	2,0
3/4	3	Sol	Maior com 7 ^a	Si	8,0
3/4	3	Sol	Maior com 7 ^a	Lá	4,0
3/4	4	Dó	Maior	Sol	8,0
3/4	4	Dó	Maior	Mi	4,0
...

Fonte: Autoria Própria

5.2 PROCESSAMENTO E PREPARAÇÃO DAS INFORMAÇÕES DO BANCO DE DADOS

Com objetivo de facilitar a posterior fase de treinamento da RNA, é preciso primeiro processar os dados de forma a deixá-los padronizados. Depois, para servir de entrada ao sistema, essas informações processadas precisam ser codificadas para que a rede possa continuar a execução. Portanto, o processamento dos dados do banco acontece em duas etapas: padronização de dados de treino e preparação e codificação dos mesmos, que serão explicadas a seguir.

5.2.1 Padronização de Informações Melódico-Harmônicas

A padronização acontece em termos de:

- **Tonalidade:** as melodias e harmonias são transpostas para uma tonalidade comum, para que exista consistência de dados e comparabilidade. Neste caso, todas as músicas do banco

de dados foram transpostas para a tonalidade de Dó maior; isso garante um conjunto de notas esperadas (escala de Dó maior) e define um campo harmônico comum.

- **Duração das notas:** no banco de dados, a duração de cada nota está padronizada de acordo com os valores exibidos na Tabela 7, sendo que a lógica de duração segue as divisões como exposto na Figura 1. Dado o fato de que cada música contém a própria fórmula de compasso, é preciso que haja normalização, pois para cada compasso de músicas com andamentos diferentes, a somatória das durações das notas de cada compasso teriam valores diferentes. Para balancear a soma de durações, a duração de cada nota é multiplicada pelo inverso da fórmula de compasso corrente.

Tabela 7 – Padrão dos valores de duração das notas do banco de dados.

Figura rítmica	Duração
Semibreve (.)	16.0
Mínima (J)	8.0
Semínima (J)	4.0
Colcheia (J)	2.0
...	...

Fonte: Autoria Própria

Sendo d a duração de uma nota musical, o desejado é que $\sum d = 16$ independente da fórmula de compasso da música. Para atender este requisito, sendo F_c a fórmula de compasso, é preciso fazer:

$$d \times F_c^{-1} = d \times \frac{\text{Tipo da figura rítmica}}{\text{Quantidade de tempos da figura}}. \quad (3)$$

Dessa forma, os compassos sempre terão duração total igual. Por fim, para cada compasso da melodia de entrada, haverá um vetor de notas normalizadas no tempo e transpostas.

- **Tipos de acorde:** o problema dos acordes é a grande quantidade de possibilidades para sua formação. Portanto, foi feita uma simplificação, mantendo apenas acordes maiores, maiores com sétima e menores para todas as possibilidades cromáticas, usando somente notação com sustenidos, o que resulta em um total de 36 classes de acordes, expostas na Tabela 8.

Tabela 8 – Classes de acordes consideradas para o processamento do banco de dados.

Maiores	Menores	Com 7 ^a
C	Cm	C ⁷
C♯	C♯m	C♯ ⁷
D	Dm	D ⁷
D♯	D♯m	D♯ ⁷
E	Em	E ⁷
F	Fm	F ⁷
F♯	F♯m	F♯ ⁷
G	Gm	G ⁷
G♯	G♯m	G♯ ⁷
A	Am	A ⁷
A♯	A♯m	A♯ ⁷
B	Bm	B ⁷

Fonte: Autoria Própria

Ao final da padronização, pode-se concluir que há a simplificação das informações para somente o essencial para o presente objetivo, sendo que os dados finais são apenas compasso, acorde, nota e duração, ou seja, o suficiente para representar as músicas no tempo e a altura das notas.

5.2.2 Preparação e Codificação dos Conjuntos de Treino

Tendo em mente a necessidade de realização de testes para determinar melhor a performance do harmonizador automático, cinco formas diferentes de selecionar os dados foram preparados para realização da etapa de treino da RNA. Esses conjuntos levam em conta diferentes aspectos da relação entre melodia e harmonia das músicas. Entretanto, têm todos em comum a consideração de somente o primeiro acorde de cada compasso de música, pois é o objetivo: gerar um acorde por compasso. São as formas de seleção (Se_i) descritas a seguir:

- **Seleção 1 (Se_1):** somente a primeira nota de cada compasso, para observar se essa relação já é suficiente para determinar a harmonia;
- **Seleção 2 (Se_2):** a primeira nota e a nota do meio de cada compasso, pois determinam geralmente os momentos mais fortes, podendo dar bom indicativo do acorde corrente;
- **Seleção 3 (Se_3):** o conjunto de todas as notas da primeira metade do compasso considerando as pausas;
- **Seleção 4 (Se_4):** todas as notas do compasso, também considerando as pausas;
- **Seleção 5 (Se_5):** as notas do compasso descartando as pausas, dando importância somente aos sons.

Os conjuntos são então codificados para servirem de entrada à RNA. Tanto notas como acordes são codificados em vetores usando o método *one-hot* (HARRIS; HARRIS, 2013). Dessa forma, um vetor de 13 posições representa as notas (12 notas musicais possíveis mais uma

posição para identificar uma pausa), e outro vetor de 36 posições indica o acorde (12 classes de acordes maiores, mais 12 maiores com sétima e 12 menores). A Figura 13 oferece uma ilustração para a explicação feita, para a codificação da nota Ré e do acorde de Sol menor.

Figura 13 – Formato dos vetores de (a) notas e (b) acordes codificados segundo o método *one-hot* (HARRIS; HARRIS, 2013).

	C	C#	D	D#	E	F	F#	G	G#	A	A#	B	Pausa
<i>nota</i> =	0	0	1	0	0	0	0	0	0	0	0	0	0
(a)													
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B	Maior
<i>acorde</i> =	0	0	0	0	0	0	0	0	0	0	0	0	
...	0	0	0	0	0	0	0	1	0	0	0	0	menor
...	0	0	0	0	0	0	0	0	0	0	0	0	7ª
(b)													

Fonte: Autoria Própria

5.3 PREPARAÇÃO DA REDE NEURAL ARTIFICIAL

Tendo por entrada os vetores codificados de notas, estas são levadas à RNA, que será responsável por dar acordes também codificados na saída. O processo de preparação dos dados, a entrada e a saída da RNA podem ser simplificados pelo ilustrado na Figura 14.

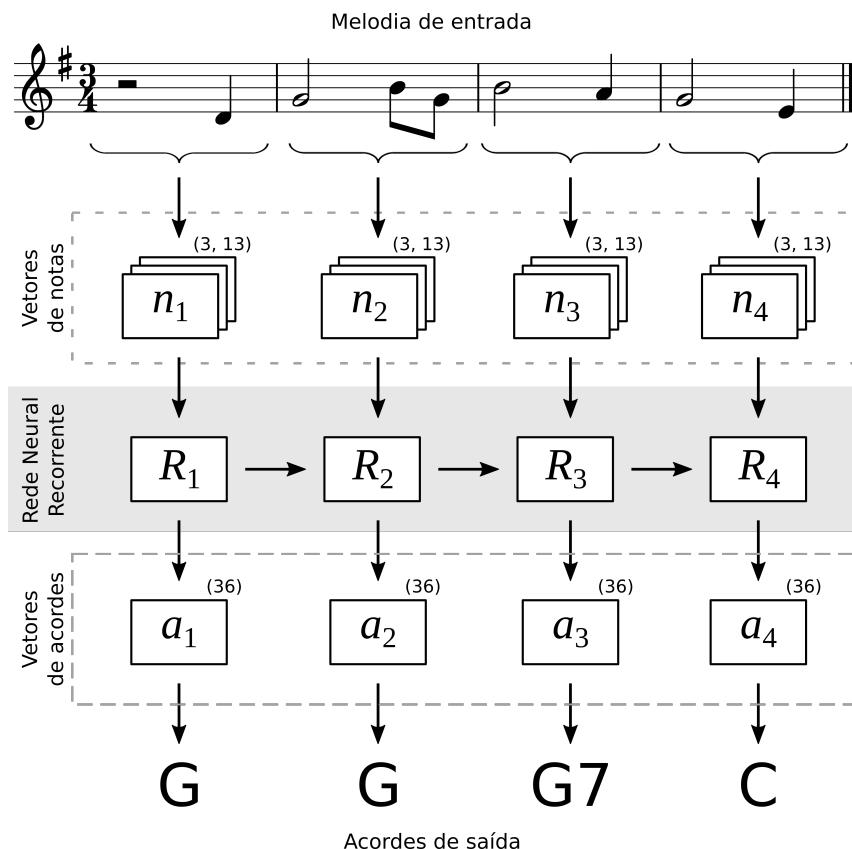
Cada nota é transformada em um vetor codificado, portanto cada compasso terá vários vetores de notas. Para o caso ilustrado, o segundo compasso detém a maior quantidade de notas (três), portanto para que a entrada seja equilibrada, cada compasso produz três vetores de notas. Nos demais compassos, onde não existe a mesma quantidade de notas, é adicionado um vetor vazio. Então os quatro compassos ilustrados são processados em quatro conjuntos de vetores de notas n , sendo que cada conjunto possui três vetores de notas de tamanho 13. A RNR então recebe a cada iteração R estes conjuntos de vetores, e realiza a retropropagação (indicada pelas setas horizontais). Também, a cada iteração, é produzido um vetor de acorde a de tamanho 36, referente ao compasso de entrada, que pode ser finalmente decodificado para um nome conhecido, caracterizando os acordes de saída.

A seguir, será explicada a forma de modelagem e configuração da RNA, as camadas usadas na sua constituição, bem como a forma de treino e predição do sistema.

5.3.1 Modelagem e Configuração da RNA

Para este trabalho, três formas de RNA foram analisadas, todas pertencentes a categoria de RNR, sendo elas: RNR simples (RNR_1), LSTM (RNR_2) e LSTM de duas camadas (RNR_3). A

Figura 14 – Visão geral do sistema proposto com processamento da melodia em vetores de notas servindo de entrada para a RNR, cuja saída é um vetor de acordes que posteriormente é decodificado.



Fonte: Autoria Própria. Adaptado de Lim e Lee (2017)

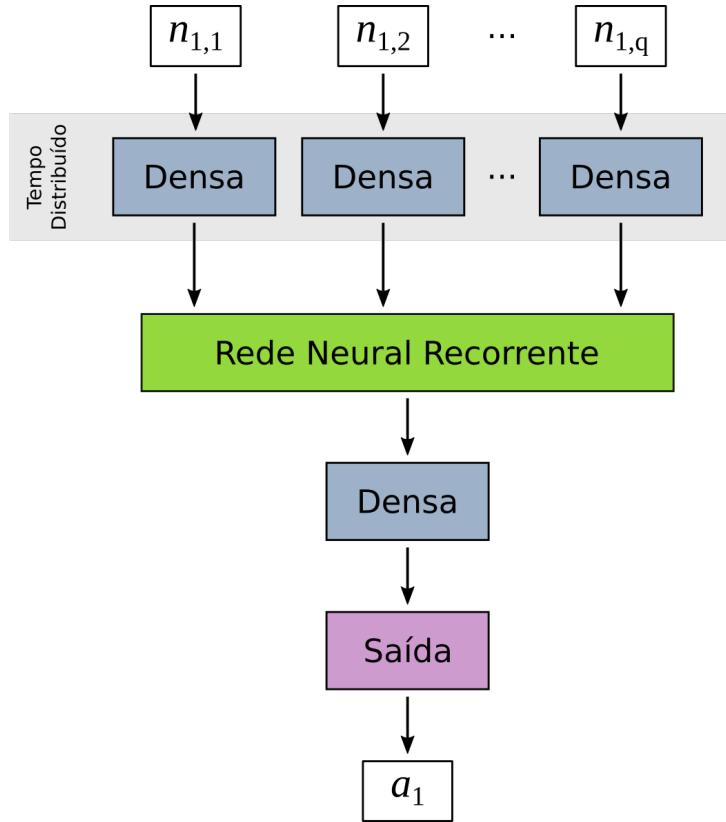
construção conta na entrada com uma camada de tempo distribuído densa tridimensional, cujas dimensões são os compassos, a quantidade de notas considerada por compasso e o tamanho do vetor de nota codificado, seguido da camada de RNR, tendo por fim duas camadas densas, a primeira com o mesmo número de unidades que a RNR, e a última com 36 neurônios de saída, em concordância com o vetor de acorde codificado. Entre as camadas escondidas, há a passagem por *dropout*¹. A Figura 15 exemplifica a construção dos modelos de maneira genérica. Cada camada será explicada a seguir.

5.3.1.1 Camadas Usadas

Diferentes tipos de camadas, tanto de entrada, saída e escondidas, foram usadas para a construção dos modelos propostos. Isso é necessário devido a complexidade e características do problema abordado. Ao todo, foram usadas camadas do tipo densa, de tempo distribuído e recorrentes, esta última caracterizadora do tipo da rede.

¹Usada para regularização e prevenção de sobre-ajuste (ou sobre-treino, do inglês *overfitting*) (SRIVASTAVA et al., 2014), *dropout* (traduzido como "deixar fora" ou "cair fora") define uma quantidade randômica de dimensões cuja saída deve ser zero, com a probabilidade desejada.

Figura 15 – Modelagem interna genérica da RNA utilizada, para um conjunto de q vetores de notas de um compasso.



Fonte: Autoria Própria

A camada densa é uma camada padrão, totalmente conectada e regular. Sua função principal é conectar neurônios entre camadas e possibilitar a adição de pesos ao sistema, além de ser usada também para alterar a dimensionalidade da saída (KERAS, 2015).

A camada de Tempo distribuído funciona como um invólucro para aplicar uma camada (densa) a cada parte de uma entrada que considera divisão temporal (KERAS, 2015). É o caso analisado, a considerar que a entrada é um compasso de música, de notas sequenciais, dependendo do conjunto de dados processado. Isso não se aplica ao conjunto que considera apenas a primeira nota do compasso. Neste caso, essa camada funciona apenas como uma simples camada densa.

A configuração dos modelos para o processo de aprendizado é feita aplicando como função de custo entropia cruzada categórica, otimizados pelo algoritmo Adam (nome derivado de *Adaptive Moment Estimation*) (KINGMA; BA, 2014), tendo por métrica principal de observação a acurácia. Esses detalhes serão explicados a seguir:

- **Função de Custo:** dentre diversas funções de custo estabelecidas, a de entropia cruzada foi aplicada neste trabalho. A definição da função de entropia cruzada EC , para $a = \sigma(z)$ que é a função sigmoid, e z representa a soma de pesos e entradas de um neurônio mais a

função de ativação, é como segue:

$$EC = -\frac{1}{n} \sum_x [y \ln(a) + (1 - y) \ln(1 - a)], \quad (4)$$

sendo que n representa o número total de dados de treino, x é o dado de entrada e y é a saída esperada. Segundo Academy (2019), "a entropia cruzada descreve a perda entre duas distribuições de probabilidade". Essa função é preferida pois a entropia cruzada é positiva e tende a zero conforme a computação da saída desejada é melhorada, além de evitar o problema de desaceleração de aprendizado, que é quando determinado neurônio tem dificuldade de aprender o comportamento correto quando este retorna uma saída muito errada (ACADEMY, 2019).

- **Otimizador:** entendendo a tarefa de aprendizado de máquina como um problema de otimização, faz-se necessária a seleção de um algoritmo de otimização, sendo que é aplicado tradicionalmente um algoritmo de descida de gradiente estocástico pela sua flexibilidade e velocidade de operação (FACURE, 2017).

Um método popular na comunidade é o Adam, um método para otimização estocástica eficiente que requer pouca memória, e combina características de outros dois métodos recentes: AdaGrad e RMSProp. Algumas das vantagens de se utilizar Adam para a solução do problema de condicionamento da RNA para harmonização automática são que seus tamanhos de degrau (ou lote) são limitados pelos hiperparâmetros definidos no modelo, e não requer um objetivo estacionário (KINGMA; BA, 2014).

5.3.2 Descrição da Etapa de Treino da RNA

O treinamento da RNA é a fase que antecede a geração automática de informações, e é onde o sistema é condicionado, onde de fato aprende a realizar previsões. Para que isso ocorra, o banco de dados precisa ser dividido em dois grupos, um para treino e outro para teste. O primeiro conjunto de dados é aplicado à rede que, com base nas entradas e saídas esperadas, calcula os pesos das conexões entre neurônios em um processo de retropropagação. Outrossim, os dados de teste, que não foram vistos pelo sistema durante o treino, são então usados para verificar a consistência das previsões para dados novos.

Além de dividir em conjuntos o banco de dados, outros parâmetros precisam ser definidos, chamados de hiperparâmetros, como por exemplo: quantidade de épocas ou passadas sobre os dados de treino, o tamanho de cada lote antes de ocorrer a atualização dos pesos das conexões, fator de *dropout* e quantidade de neurônios ou unidades em cada camada de rede. Embora não existam quantidades predefinidas ideais pelo fato da singularidade de cada problema, segundo Reimers e Gurevych (2017), algumas características causam maior impacto que outras. Ainda assim, faz-se necessária a realização de testes para encontrar uma configuração que possibilite melhores retornos.

5.3.3 Predição de Novos Valores

Tendo sido treinada, a RNA é capaz de realizar previsões sobre novos conjuntos de dados de entrada. Sendo assim, para o conjunto de dados de teste, são preditos acordes que são usados posteriormente na fase de coleta e análise de resultados, a qual é descrita na Seção 5.4.

Duas modalidades de testes foram consideradas, que resultam em respostas quantitativas e qualitativas sobre as previsões do sistema. Estes são importantes para realizar determinações e conclusões sobre a eficiência do sistema, permitindo dizer se os objetivos desejados foram atingidos ou não.

5.4 TESTES QUANTITATIVOS

A saída geral do sistema são harmonias para as melodias servidas na entrada. Para o conjunto de teste selecionado no período de processamento do banco de dados, as harmonias geradas são armazenadas separadamente para realizar levantamentos sobre os resultados obtidos e medir a taxa de desempenho do sistema. Este levantamento e medição de taxas se dá aplicando as definições encontradas na Seção 4.3, sendo o uso da matriz de confusão e suas respectivas taxas e medidas, como sensibilidade, precisão e F_β -score e usando a macro-média.

5.5 TESTES QUALITATIVOS

Os testes qualitativos foram realizados para inferir conclusões a respeito da qualidade da saída do sistema, e são importantes pois revelam questões que não podem ser facilmente explicadas por meio de medidas quantitativas. Para tanto, as duas formas propostas para levantar resultados qualitativos são descritas a seguir.

5.5.1 Enquete *Online*

Um fator interessante de se observar é como se comporta o gosto do público geral mediante os resultados de harmonização gerados pelo sistema de harmonização automática proposto. Para tanto, uma pesquisa foi feita com algumas perguntas para dois grupos: 1) pessoas que têm experiência musical, seja com composição, estudo teórico ou prático, ou ainda que sejam músicos; e 2) pessoas sem qualquer experiência técnica musical além dos seus conceitos prévios naturais, já que as pessoas só por ouvirem música têm uma ideia do que é adequado ou não de maneira intrínseca. Ambos foram levados a responder qual harmonização preferem dentre duas possibilidades: a original e a automática para cinco músicas, sem revelar aos participantes qual harmonização é qual.

5.5.2 Análise Técnica da Produção Musical do Sistema

Além de analisar como se comporta o gosto dos participantes da pesquisa, outro ponto importante é realizar uma análise mais profunda sobre a saída do sistema. Para fazer este teste,

uma obra é harmonizada automaticamente e é inspecionada por um especialista em teoria de harmonia, que dará uma nota para a harmonização do sistema. Essa nota, por se tratar de um valor real, pode ser incorporado também aos resultados quantitativos.

5.6 REFERENCIAL EXPERIMENTAL

Para conferir o cumprimento de cada um dos objetivos propostos foram definidas formas e experimentos para avaliação, conforme discussão a seguir.

Como resultado tangível, espera-se que este método esteja apto a encontrar harmonias realistas. Determinar isso não é uma tarefa trivial, uma vez que sua resposta pode ser subjetiva e relativa ao gosto do ouvinte, no caso de harmonias musicais. Portanto foram propostas formas de defini-la tanto qualitativa quanto quantitativamente. Esta última, por meio da interpretação de medidas calculadas para caracterizar os resultados, além da opinião e critério pessoal de cada indivíduo. Igualmente, métodos para atingir resultados qualitativos também foram usados, na tentativa de entender como foi a reação popular, e qual grau de semelhança com o comportamento humano foi possível obter.

Entende-se que, para que os resultados tenham boa qualidade, é necessário que estes sejam coerentes com algumas das regras de harmonia consideradas mais comuns, já que elas estabelecem não rigidez, mas um guia de padrões funcionais. Sabendo que estas regras estão, em grande parte, respeitadas nas músicas populares e clássicas (é da música que vieram as regras), e que são estas as mesmas que estão contidas no banco de dados, espera-se por parte do modelo, o aprendizado de como compor respeitando os padrões e, portanto, regras estabelecidas na teoria da música.

Uma das propostas deste projeto é que as harmonizações forneçam respostas diversificadas das harmonizações existentes, quando comparadas às respostas convencionais, em outras palavras, humanas. Isso possibilita o uso ágil da aplicação em tarefas reais e, para alguns, cotidiana. A diversidade de composição foi analisada frente a comparação dos resultados obtidos com harmonizações reais e comumente empregadas em melodias conhecidas.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Os resultados da aplicação da metodologia proposta para harmonização automática de melodias, sendo estes resultados da seleção de dados do banco, do treino dos modelos e da aplicação dos testes quantitativos e qualitativos, são descritos em detalhes nas seções subsequentes.

6.1 SELEÇÃO DE INFORMAÇÕES DO BANCO DE DADOS

Ao todo, 2249 músicas foram processadas, totalizando 342723 notas e pausas e 75087 acordes, o mesmo número de compassos, pois considera-se um acorde por compasso. Considerando as cinco formas de seleção propostas (descritas na Seção 5.2), para cada, uma quantidade diferente de notas por compasso existe, sendo que um destes compassos conta com o maior número de notas possível. Para cada forma de seleção, esses máximos são mostrados na Tabela 9.

Tabela 9 – Quantidade máxima de notas por compasso para cada forma de seleção como definido na Subseção 5.2.2.

Forma de seleção	Quantidade máxima
Se_1	1
Se_2	3
Se_3	10
Se_4	32
Se_5	32

Fonte: Autoria Própria

Para realizar as etapas de treino e teste dos modelos de RNA propostos, o conjunto de músicas foi dividido em dois, sendo que 80% foi destinado para treinamento, e os demais 20% para teste, ou seja, usando *Holdout* 80%.

6.2 TREINO DOS MODELOS DE RNA

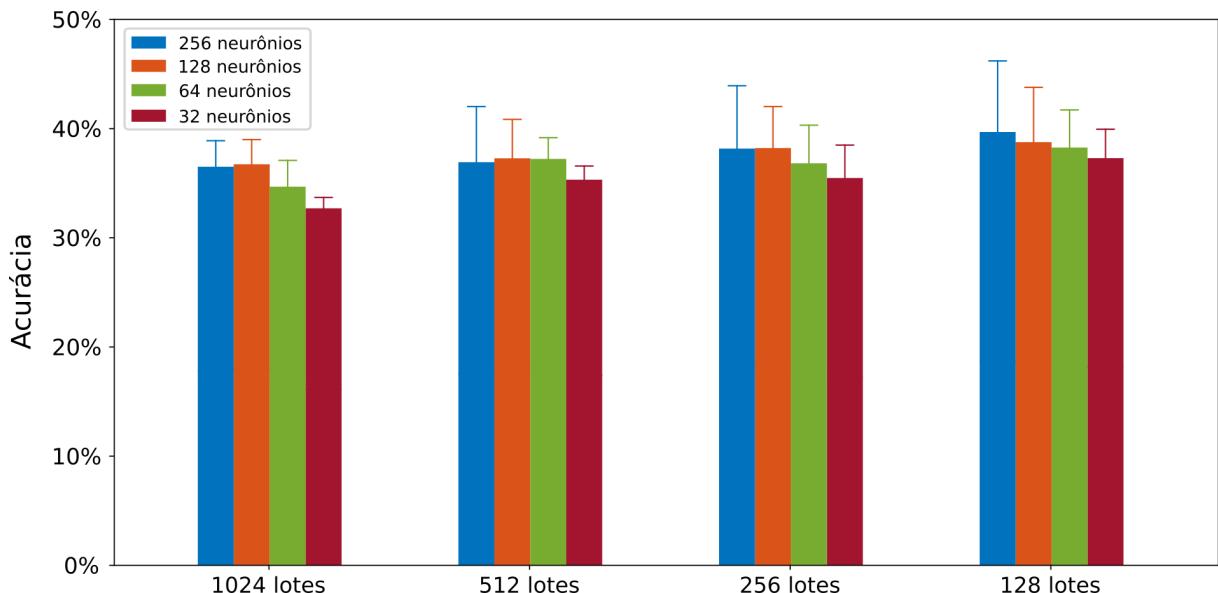
O desenvolvimento de cada modelo aconteceu de forma a encontrar empiricamente hiperparâmetros que retornassem o melhor resultado de acurácia ou taxa de acerto. Importante observar que duas acuráncias diferentes são calculadas para cada modelo, a depender do conjunto usado: acurácia sobre o conjunto de treino e sobre o conjunto de teste. A análise de ambas é importante pois ajuda a entender como está condicionado o sistema, podendo também dizer se

houve sobre-ajuste ou sub-ajuste¹, caracterizados pela discrepância entre as medidas, sendo a acurácia de treino maior que a de teste no caso de sobre-ajuste, e menor para o caso de sub-ajuste. Outro fator importante é que deve-se considerar como principal a acurácia sobre o conjunto de teste, uma vez que esta é indicativa da capacidade real que o modelo tem de processar entradas novas ou desconhecidas. A escolha do melhor modelo depende, portanto, do valor mais alto de acurácia de teste e que haja pouca diferença relativa entre acuráncias.

Foram considerados cinco hiperparâmetros variáveis durante a fase de treino dos modelos: tamanho do lote, quantidade de neurônios, épocas, taxa de *dropout* e porcentagem de dados a ser considerada dos conjuntos de treino e teste. Para os primeiros testes, a quantidade de épocas e porcentagem de dados foram reduzidas para diminuir o tempo de treino, usando 30 épocas e 10% do conjunto de dados. *Dropout* foi definido em 0.2, e valores de lote e neurônios foram variados, sendo as possibilidades, para lotes: 128, 256, 512, 1024; e para quantidade de neurônios: 32, 64, 128, 256, respectivamente.

Com o intuito de exemplificar o desenvolvimento da escolha de melhores configurações nesta primeira etapa de testes, é mostrado o gráfico de acurácia de teste e de treino do modelo RNR₂ utilizando a forma de seleção *Se*₄, para as quatro possibilidades de valores de lotes e neurônios. As barras representam a acurácia de teste para cada configuração de lotes e neurônios, e as linhas superiores os valores de acurácia de treino.

Figura 16 – Gráfico de acurácia de teste (barra) e de treino (linha superior) para valores diferentes de lotes e neurônios para RNR₂ e *Se*₄.



Fonte: Autoria Própria

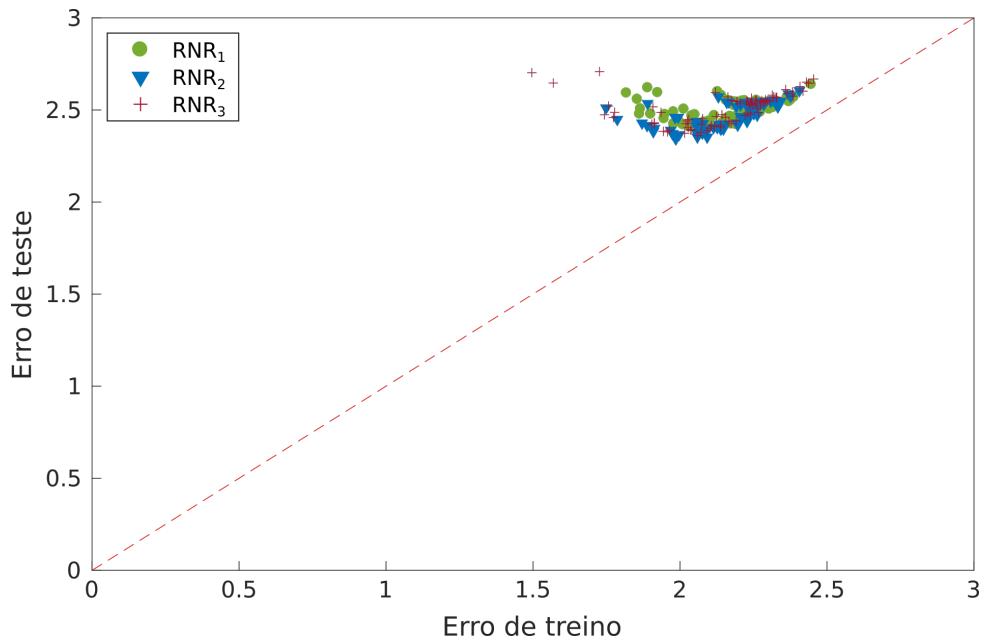
De forma geral, podem-se notar duas tendências: a diminuição da acurácia para valores menores de neurônios, e o aumento geral de acurácia para menores valores de lotes. Entre-

¹Sobre-ajuste acontece quando o sistema se torna especialista para os dados do conjunto de treino mas ineficaz mediante novas entradas. Sub-ajuste, por sua vez, é a incapacidade de aprender tanto sobre o conjunto de treino como para novos dados (RUSSELL; NORVIG, 2009).

tanto, valores menores de lotes implicam em maior distanciamento entre acurácia de teste e de treino, uma característica indesejada por denotar sobre-ajuste. Sendo assim, tem-se por configurações preferidas 512 lotes com 64 neurônios e 128 lotes com 32 neurônios, devido a pouca diferença entre acurárias e alto valor de acurácia de teste relativos aos demais, sendo seus valores respectivamente de 36.52% e 36.35%, com diferenças entre acurárias de 2.68% e 3.62%, respectivamente.

Nesta primeira etapa de treino foi possível perceber a ocorrência de sobre-ajuste generalizado, como mostrado na Figura 17, para valores de erro de treino e teste calculados aplicando a função de custo de entropia cruzada definida pela Equação (4). Espera-se que os valores de erro estejam o mais próximo possível da linha central, representativo de concordância. Portanto, a taxa de *dropout* foi alterada para 0.5 e todos os testes foram feitos novamente.

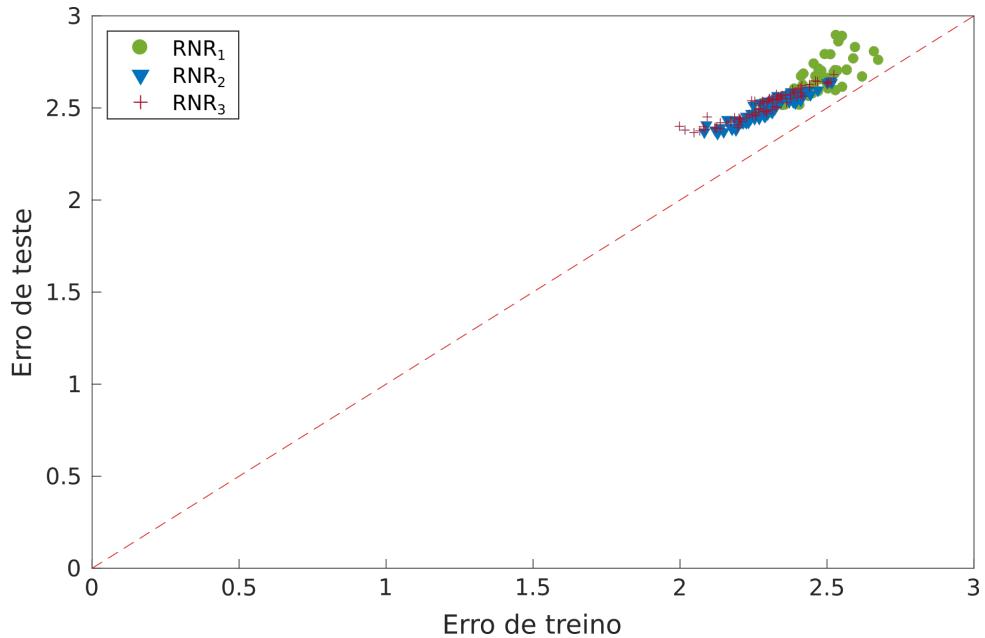
Figura 17 – Valores de erro de treino e teste calculados usando a função de custo de entropia cruzada (Equação (4)) para todos os modelos da primeira fase de testes.



Fonte: Autoria Própria

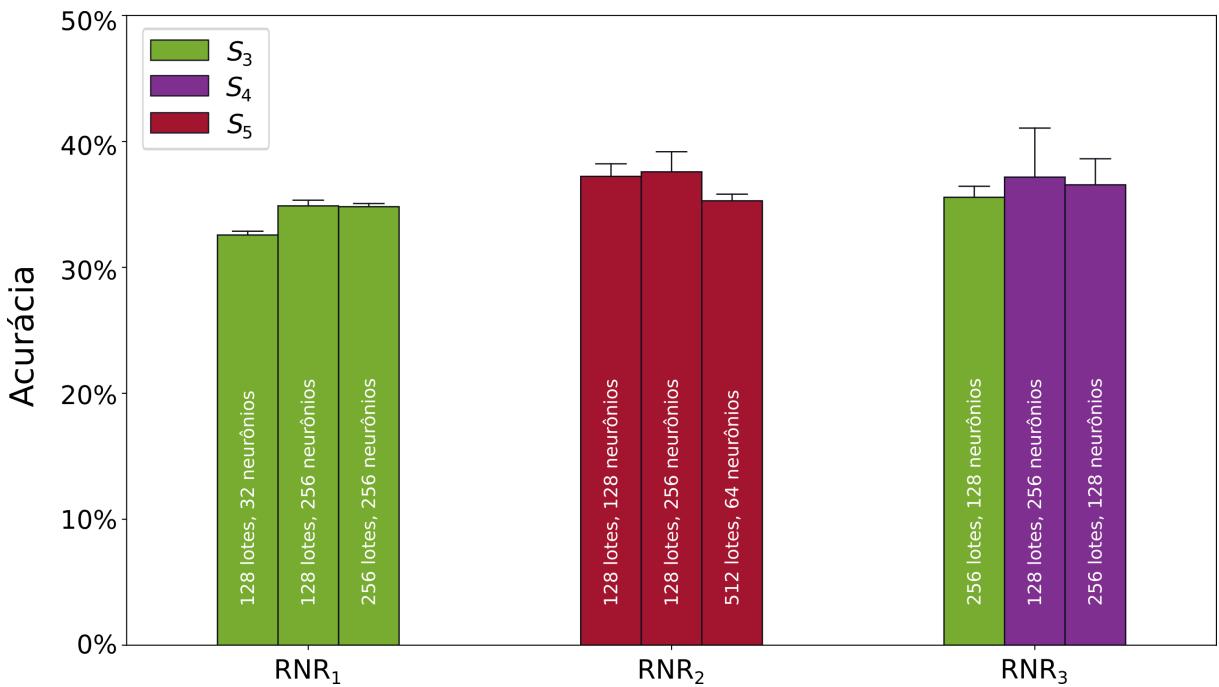
O gráfico da Figura 18 confirma a redução do sobre-ajuste, aprovando o novo valor definido para *dropout*, também notado na seleção de modelos da Figura 19, onde são exibidos os nove melhores modelos encontrados. Os dados da Figura 19 revelam que para o modelo RNR₁, houve melhor resposta ao se aplicar os dados de *Se*₃, sendo que o modelo com melhor acurácia é o de 256 neurônios e 128 lotes, com 34.74%. Já para RNR₂, o que foi escolhido como de melhores respostas foi a seleção *Se*₅, com 36.89% de acurácia para 128 lotes e 128 neurônios. Para RNR₃, houve maior dificuldade em conter o sobre-ajuste, sendo que *Se*₃ e *Se*₄ foram preferidas, e o modelo que considera 256 lotes e 128 neurônios usando *Se*₄, embora um pouco sobre-treinado, é escolhido como melhor, por retornar 35.87% de acurácia com diferença entre acuráncias de treino e de teste de 2.76%.

Figura 18 – Valores de erro de treino e teste com alteração do valor de *dropout* para 0.5.



Fonte: Autoria Própria

Figura 19 – Acurácia de teste (barra) e treino (linha superior) dos nove melhores modelos selecionados da segunda fase de testes com *dropout* em 0.5.

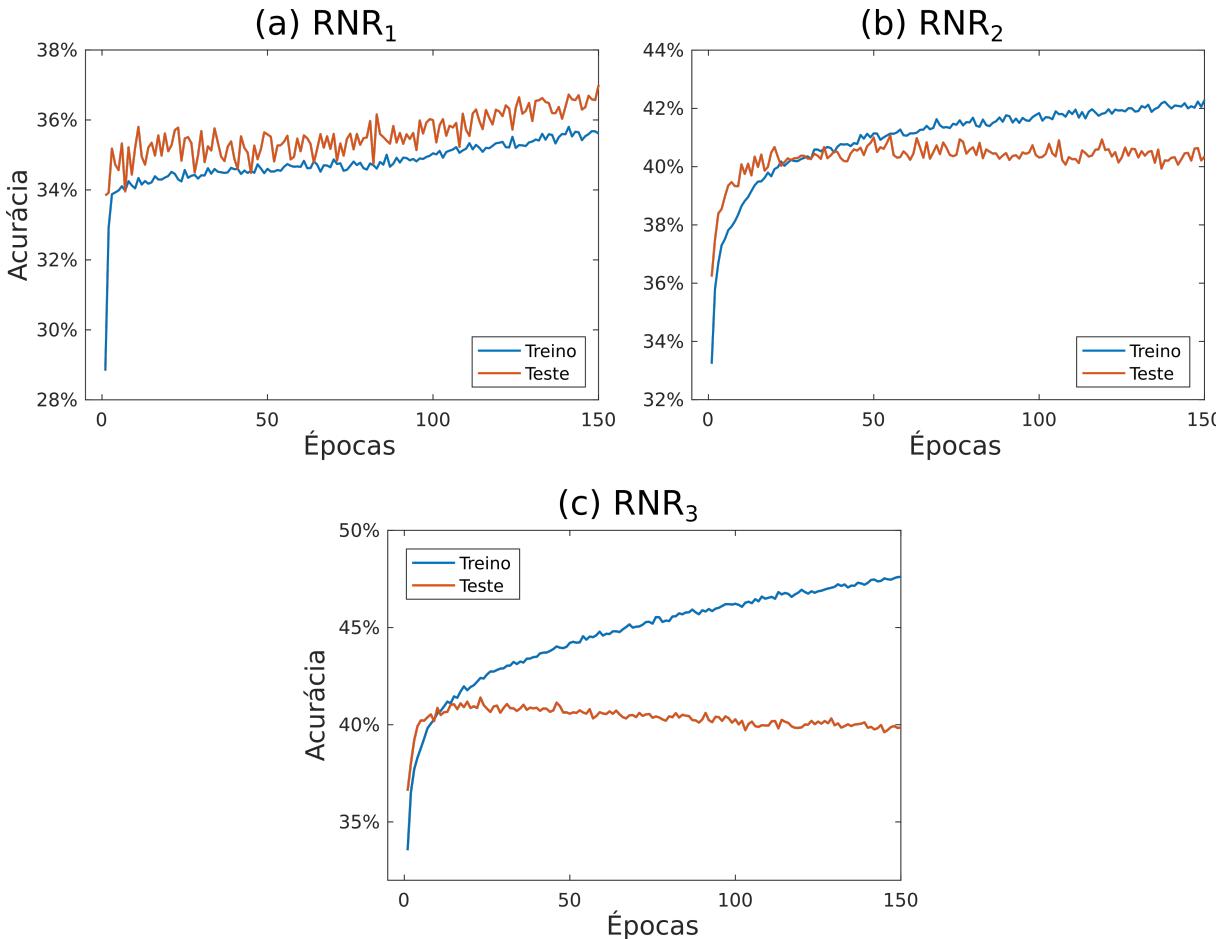


Fonte: Autoria Própria

Utilizando estes nove melhores modelos, o próximo teste envolve a total utilização dos dados processados e maior número de épocas para observar a curva de aprendizado de cada um ao longo do tempo, visando determinar o melhor valor de épocas. Para este teste, foram usadas 150 épocas de treinamento e observadas as acurácias de treino e de teste. O gráfico do melhor de

cada modelo pode ser visto na Figura 20.

Figura 20 – Desenvolvimento da acurácia de treino e de teste (em %) para os melhores modelos de (a) RNR₁, (b) RNR₂ e (c) RNR₃ escolhidos da terceira fase de testes com 150 épocas de treinamento.



Fonte: Autoria Própria

Para os três modelos, vê-se comportamento diferente e particular. O modelo RNR₁ (Figura 20(a)) demonstra sub-ajuste, abrindo espaço para adição de mais épocas, já que a tendência é de crescimento de acurácia. A curva de acurácia de teste, por sua vez, é bastante ruidosa, fato que pode ser explicado pela dificuldade que este formato de RNR tem em guardar informações de longo prazo. Isso pode ser contrastado para os dois outros casos, sendo que o último é de maior estabilidade, devido a sua complexidade. O modelo RNR₂ (Figura 20(b)) manteve bom desempenho até atingir a quantidade máxima de aproximadamente 50 épocas, sendo este o valor ideal para este hiperparâmetro. Para RNR₃ (Figura 20(c)), percebe-se rápida estabilização da acurácia de teste, tendo o melhor valor de épocas igual a 15. Esse rápido desenvolvimento pode ser explicado pelo fato de que neste teste os conjuntos de treino e teste foram completamente usados.

Após a realização destes testes de modelagem, tem-se os modelos com hiperparâmetros definidos de acordo com a Tabela 10, para a seleção de dados aplicada e valores de neurônios,

lotes, *dropout* e épocas.

Tabela 10 – Configurações finais selecionadas empiricamente para cada modelo de RNR testado.

Modelo	Seleção	Neurônios	Lotes	<i>Dropout</i>	Épocas
RNR ₁	<i>Se</i> ₃	256	256	0.5	500
RNR ₂	<i>Se</i> ₅	128	256	0.5	50
RNR ₃	<i>Se</i> ₄	256	128	0.5	15

Fonte: Autoria Própria

6.3 RESULTADOS QUANTITATIVOS

A seguir, faz-se uma análise mais profunda sobre a resposta de cada modelo com critério comparativo, para determinar melhor a escolha final para o sistema de harmonização automática. Um exemplo do resultado de uma melodia harmonizada por cada modelo, bem como a harmonia original, é mostrada na Figura 21 para o trecho da música *December, 1963 (Oh What a Night)*, de *Frankie Valli & The Four Seasons*, entre os compassos 11 e 18. Nota-se que a linha do modelo RNR₁ é a mais discrepante da original, e que RNR₃ apresenta resultado mais próximo, com a escolha de mais acordes parecidos. Entre os modelos, existe total concordância nos compassos 1, 4 e 7, embora estes sejam diferentes do original, podendo inferir uma alternativa de harmonia. Nos compassos 5 e 8, RNR₂ e RNR₃ selecionaram o mesmo acorde que o original, porém utilizando a 7^a, mostrando comportamento correto no entendimento de que estes devem ser usados em pontos de maior tensão que necessitarão de resolução posterior. Por fim, é concluído o objetivo de obter um sistema que apresente diversidade das harmonias originais, uma vez que para a mesma melodia, três harmonizações diferentes foram criadas. Além disso, a diversidade pode ser comprovada pela possibilidade de uso de 36 classes de acordes.

A anterior análise permite dizer que houveram diferenças entre a performance e compreensão de cada modelo. Uma análise mais profunda pode ser feita observando medidas, como as definidas na Seção 5.4.

A matriz de confusão dos três sistemas é ilustrada na Figura 22 para (a) RNR₁, (b) RNR₂ e (c) RNR₃. Analisando as matrizes de confusão, percebe-se o extensivo uso de acordes principais das funções tônicas (C, F e G) e em especial, o acorde dominante (G⁷). Generalizando entre modelos, alguns pontos que chamam a atenção são os de acerto na predição do acorde de E⁷ e G⁷, além da extensiva troca de G por G⁷. É interessante notar também a presença de leves linhas verticais para os acordes Dm e A♯, sugerindo a generalização e simplificação feita por parte dos modelos durante a composição. Por fim, entre todos os modelos, RNR₃ visualmente performa melhor que os demais, sendo o que mais teve acertos, com uma porcentagem de acurácia igual a 41%.

Figura 21 – Trecho de oito compassos da música *December, 1963 (Oh What a Night)* de *Frankie Valli & The Four Seasons* com harmonização original e gerada por cada um dos 3 melhores modelos encontrados.

Original:	F	F	F	Am
RNR ₁ :	C	C	Am	C
RNR ₂ :	C	F	Am	C
RNR ₃ :	C	F	F	C

Original:	G	F	F	G
RNR ₁ :	G	C	Am	C
RNR ₂ :	G7	F	Am	G7
RNR ₃ :	G7	F	Am	G7

Fonte: Autoria Própria

Para dar continuidade, é apresentado e analisado o gráfico de F_1 -score (Figura 23), ponderando igualmente precisão e sensibilidade para cada classe de acordes. Para vários acordes, percebe-se que não houveram acertos. Alguns picos são notados, para os acordes de C, F e G⁷ principalmente, como também pode ser observado nas matrizes de confusão da Figura 22. Dentre todos, vê-se maior participação do modelo RNR₃, com maior quantidade de picos e, por vezes, aparecendo sozinho, como nos acordes D e A.

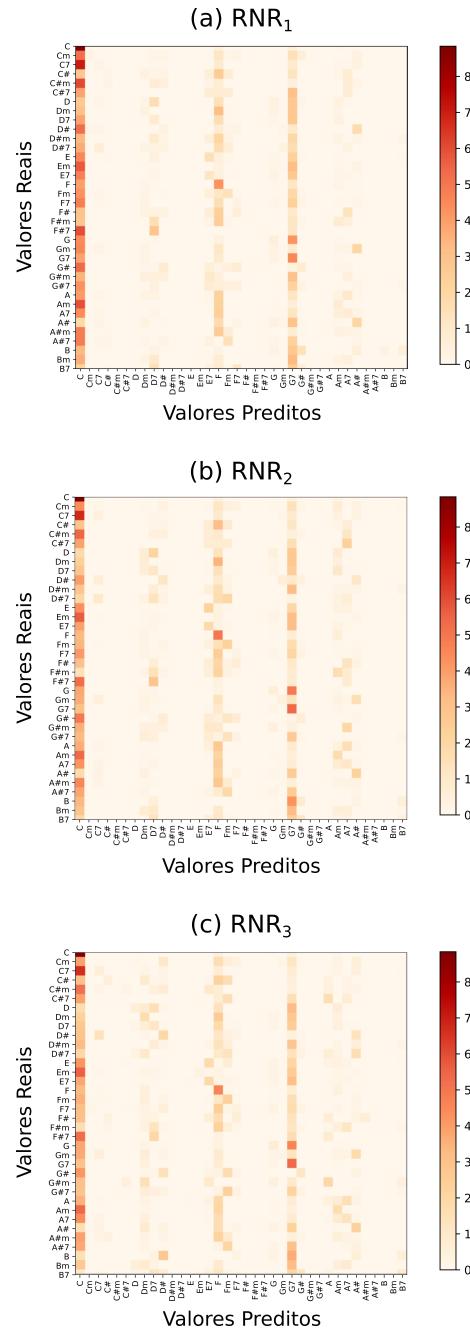
Os resultados finais das taxas S_M , E_M , F_1 -score_M e da acurácia são mostrados na Tabela 11. Embora o modelo RNR₂ apresente maior acurácia que RNR₁, este tem maior precisão que todos, ou seja, é o modelo capaz de selecionar maior quantidade de itens relevantes. Apesar disso, o modelo RNR₃ tem maior valor de sensibilidade e F_1 -score, sendo mais bem equilibrado. Além disso, também é o modelo com maior valor de acurácia, superando os demais.

Tabela 11 – Macro-média das taxas de performance (M) e valores de acurácia para os modelos propostos.

Modelo	S_M	E_M	F_1 -score _M	Acurácia
RNR ₁	18.79%	7.89%	7.91%	37.20%
RNR ₂	13.63%	9.46%	9.64%	40.12%
RNR ₃	18.51%	10.81%	11.40%	41.00%

Fonte: Autoria Própria

Figura 22 – Matrizes de confusão para os três modelos propostos: (a) RNR₁, (b) RNR₂ e (c) RNR₃, com porcentagem de acertos referenciadas por cores.

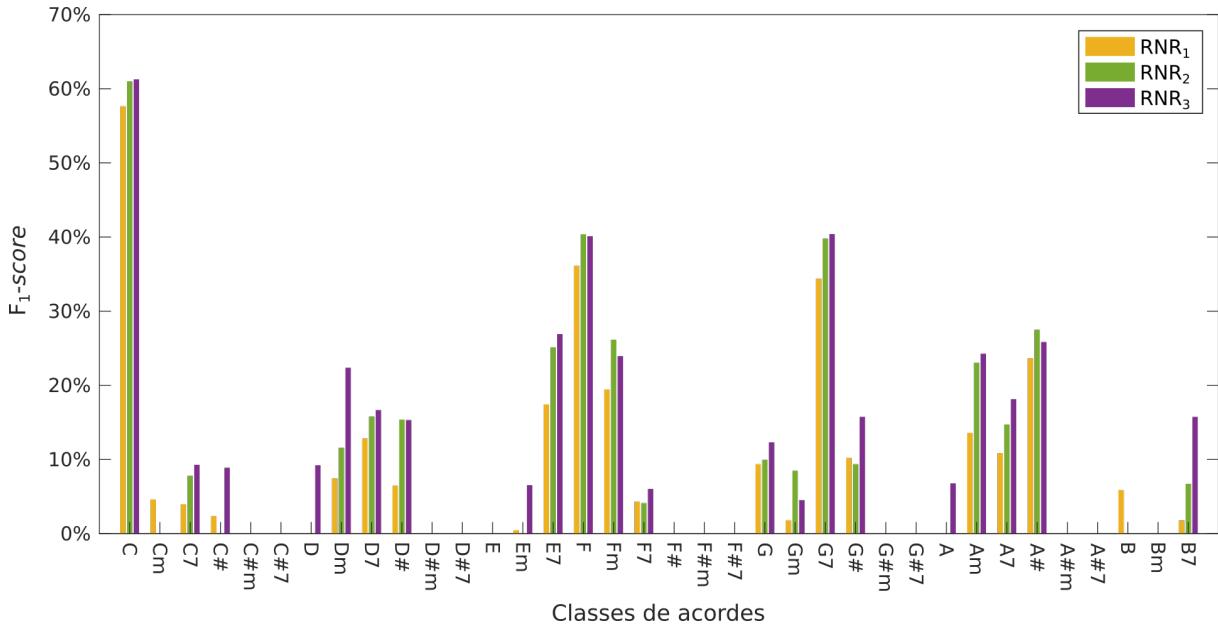


Fonte: Autoria Própria

6.4 RESULTADOS QUALITATIVOS

A pesquisa de caráter qualitativo foi desempenhada em duas frentes: enquete *online*² e a análise técnica de produções musicais resultantes do sistema. A primeira teve como objetivo

²Embora a resposta de formulário envolva a participação de seres humanos, por se tratar de uma participação anônima (sem identificação do participante) com posicionamento público de opinião, não há exigência de aprovação pelo Comitê de Ética em Pesquisa da UTFPR (www.utfpr.edu.br/comissoes/permanentes/comite-de-etica-em-pesquisa).

Figura 23 – F_1 -score de classes individuais de acordes para cada um dos modelos propostos.

Fonte: Autoria Própria

avaliar as gerações do melhor modelo analisado (RNR_3) quando comparadas com harmonizações originais e conhecidas. A segunda, por sua vez, teve como propósito saber se o sistema era capaz de respeitar regras da teoria de harmonia musical.

6.4.1 Interpretação das Respostas da Enquete *Online*

A construção do formulário contou com auxílio de ferramentas *web*, sendo elas: SoundCloud³ para armazenamento dos fragmentos de áudio, e JotForm⁴ para elaboração das perguntas. Ao todo, foram feitas seis perguntas: a primeira a respeito da experiência musical do respondente e as cinco demais sobre comparação de obras, onde o participante ouve pequenos trechos de música de no máximo 33 segundos em caráter de comparação, havendo sempre duas opções que continham a harmonização original e a automática, porém isto foi transparente ao usuário. As músicas usadas no formulário foram: *Happy Birthday* (14 segundos), *O Grande Amor* (32 segundos), *Rock of Ages* (20 segundos), *Samba Pa Ti* (33 segundos) e *Zie Ginds Komt de Stoomboot* (20 segundos), denotadas como m_1 a m_5 , respectivamente.

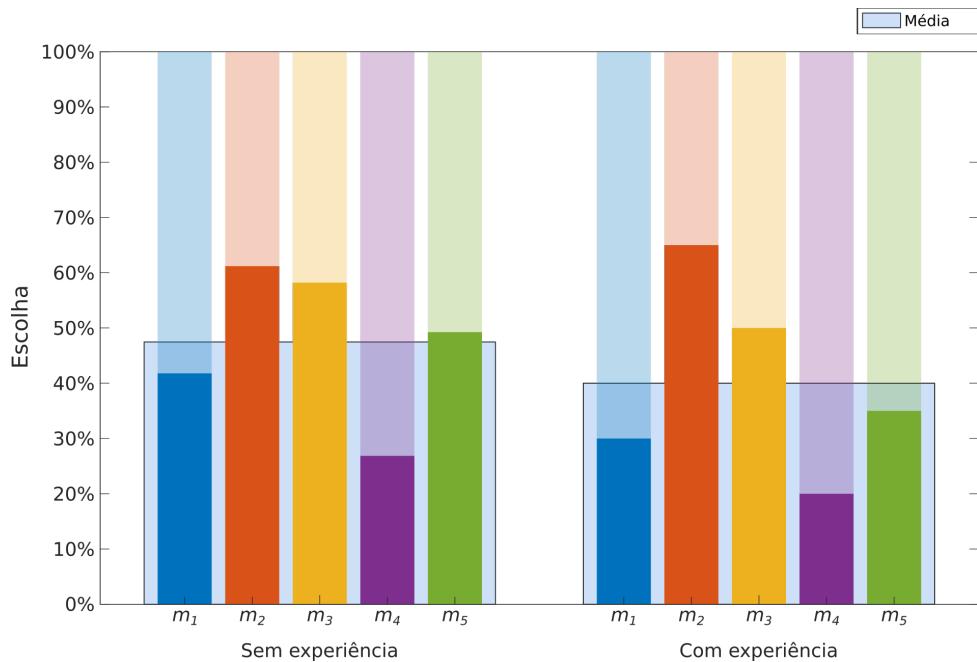
Ao todo, 87 pessoas participaram da pesquisa, das quais 20 declararam ter experiência musical diversa, e os demais 67 não. A Figura 24 mostra a porcentagem de vezes em que harmonizações automáticas foram escolhidas individualmente e a média para ambos grupos pesquisados.

A análise do gráfico revela preferências diferentes entre os grupos estudados. No entanto, existe semelhança na relação de escolhas das músicas: percebe-se a mesma ordem decrescente de

³Fragments disponíveis em: soundcloud.com/lucas-costa-398727209/sets/fragmentos-da-pesquisa.

⁴Pesquisa disponível em: form.jotformz.com/93036119173656.

Figura 24 – Resultados da enquete *online* com a porcentagem de escolha para cada música e a média de escolhas feita por cada grupo: experiente e não experiente, onde barras inferiores indicam a preferência pela harmonia automática e barras superiores, harmonia original.



Fonte: Autoria Própria

preferências para as músicas (m_2 , m_3 , m_5 , m_1 e m_4). Notadamente, os participantes preferiram a harmonização automática à original para m_2 , com 61.20% para o grupo dito sem experiência (1), e 65% para o segundo grupo (2). A música menos preferida, ou com harmonia original como predileta foi m_4 , com 26.87% e 20% para os grupos 1 e 2, respectivamente.

Valores de média e desvio padrão são mostrados na Tabela 12 para ambos grupos pesquisados. Analisando o grupo 1, pode-se afirmar que os respondentes ficaram divididos entre as harmonizações automáticas e originais, sem poder dizer com exatidão a sua preferência, diferente do grupo 2, onde é possível observar uma melhor definição pelas harmonias originais. Este fato, porém, permite assumir que o sistema foi capaz de desenvolver harmonias realistas, ocasionando eventos onde os participantes tiveram maior preferência por harmonias automáticas. Ademais, a maioria dos participantes da pesquisa teve preferência pelas harmonias originais com média geral de 43.73% de escolha pelas harmonias automáticas.

6.4.2 Resultado da Análise Técnica da Produção Musical do Sistema

A análise técnica foi realizada por um profissional especialista em teoria musical e de harmonia, o prof. Dr. Andrés Eduardo Coca Salazar⁵, que analisou a música *Alabama Waltz*, uma obra de Hank Williams, cujas partituras e cifras podem ser vistas no Apêndice B. A análise

⁵Licenciado em música pela Universidad de Caldas, Manizales, Colômbia.

Tabela 12 – Valores de média (μ) e desvio padrão (σ) para a preferência sobre músicas harmonizadas automaticamente para os dois grupos analisados: 1) sem e 2) com experiência musical.

Grupo	μ	σ
1	47.46%	13.82%
2	40.00%	17.68%
Total	43.73%	15.75%

Fonte: Autoria Própria

harmônica foi realizada desde duas perspectivas, a harmonização de notas reais e as cadências. Segue comentário na íntegra do professor sobre a obra:

"No compasso 6 há uma semínima em Lá com acorde de tônica. Esta nota não pertence ao acorde de Dó maior, além disso, a nota dura um terço do compasso e isso realça a dissonância de segunda maior criada entre a quinta do acorde e a nota da melodia. Mesmo as duas notas seguintes pertencendo ao acorde de Dó maior, este compasso deveria ser harmonizado com outro acorde ou com dois acordes diferentes. Este efeito seria minimizado se houvesse uma ligadura com a colcheia anterior, convertendo-a assim em uma nota de suspensão (nota prolongada do compasso anterior com resolução por grau conjunto descendente). Situação similar acontece nos compassos 15 e 31.

No compasso 22 encontra-se um Fá# no segundo tempo com duração de semínima, o qual não pertence ao acorde de G7 e cria uma dissonância de segunda menor com a 7ª do acorde, essa nota tem uma duração significativa e não se enquadra em nenhuma classe de nota estranha ou melódica, se assemelhando mais com uma modulação, dessa forma, este acorde não pode se considerar totalmente adequado para este compasso. Dessa forma, é possível induzir que o harmonizador automático não está preparado para este tipo de movimentos tonais, as modulações.

Com relação às cadências, a peça tem uma forma musical AA'BA', portanto 3 pontos cadenciais inter-frase e um final. Na transição entre as duas primeiras partes (A-A') encontra-se a cadência imperfeita ou semi-cadência I-V (cadência suspensiva à dominante), a qual cria a sensação de interrupção para dar continuidade à seguinte frase, que é uma variação da primeira. Nos compassos 15 a 17 deveria existir uma cadência conclusiva, porém não existe cadência alguma, simplesmente a repetição do acorde de tônica. Entre a frase B e A' (compassos 23 a 25) existe uma cadência perfeita completa IV-V7-I, a qual auxilia muito bem a transição entre essas duas frases. Novamente, no final da obra deveria existir uma cadência conclusiva perfeita, porém está ausente, o que deve-se ao fato do acorde do compasso 31 não ser o adequado, conforme explicado acima, desta forma a harmonização deixou a obra sem preparação para o final."

Ao final da análise, o professor concedeu uma nota final à obra levando em conta as duas perspectivas observadas. A nota final para a harmonização de notas reais, sem considerar a modulação, foi de 90.62, considerando três erros para os 32 compassos harmonizados; e para cadências, dois acertos de quatro pontos cadenciais existentes na música, uma nota igual a 50. Sendo pouca a quantidade de pontos cadenciais, e considerando a alta nota obtida relativa a harmonização dos compassos, é possível concluir que o sistema foi capaz de harmonizar a melodia de forma coerente com as regras da teoria harmônica tonal.

7 CONCLUSÃO

Um sistema para harmonização automática baseado em melodias musicais foi proposto, testado e apresentado. Para propor o sistema, foi feita a revisão de regras básicas musicais e teoria de harmonia funcional. O estado da arte foi levantado, revelando a importância da solução do problema de harmonização musical automática.

A metodologia proposta contou com a aplicação de banco de dados de músicas contendo harmonia e melodia, e de RNAs como sistema preditor e passível de treino para aprendizado supervisionado. Foram apresentadas diferentes formas para avaliar os resultados de maneira quantitativa e qualitativa.

O sistema final foi uma RNA do tipo recorrente que contou com uma camada dupla de LSTM com 256 neurônios, 256 lotes, *dropout* de 0.5 e 15 épocas de treino, um modelo que devido a sua complexidade foi o melhor indicado para realizar a tarefa de harmonização. A acurácia do modelo foi de 41%, um valor razoável a se considerar os resultados encontrados na literatura e se comparado com o resultado aleatório que, levando em conta as 36 classes de acordes disponíveis e usando uma distribuição uniforme, tem probabilidade de seleção de um acorde igual a 2.78%. Além disso, a utilização dessa quantidade de acordes diferentes comprova a capacidade do sistema de diversificação.

Os resultados demonstraram de maneira geral que o sistema tem na verdade capacidade de simplificação de harmonias, uma vez que regras básicas de harmonia não foram desrespeitadas durante o processo, sendo resultados coerentes como pode-se observar pela análise de uma obra real por um profissional, e também realistas graças ao exame das respostas da enquete *online*.

Sugestões para continuidade ou melhoramento deste trabalho são listadas a seguir:

- Utilização de outros bancos de dados, como por exemplo bancos de dados selecionados com exemplos de apenas um estilo musical, para observar se o sistema é capaz de fazer harmonizações que mantenham as características específicas daquele estilo;
- Extração de mais informações das melodias além de somente utilizar as notas musicais como entrada, aplicando maior representação de conhecimento e fornecendo um conjunto de medidas maior como entrada ao sistema;
- Utilizar métodos para otimização de hiperparâmetros. A escolha de hiperparâmetros é por si só um problema de busca que pode ser otimizado aplicando técnicas de validação cruzada e Estimadores da Árvore de Parzen, por exemplo;
- Desenvolvimento de uma interface para fácil utilização do sistema, ampliando seu uso customizado para a comunidade de músicos e estudantes.

Por fim, pode-se concluir que os objetivos propostos foram atingidos, uma vez que as harmonias geradas foram realistas e coerentes ao respeitar padrões convencionais da música e apresentaram diversidade frente às harmonias originais, oferecendo uma alternativa para a harmonização de melodias.

REFERÊNCIAS

- ACADEMY, D. S. **Deep Learning Book**. 2019. Disponível em: <<http://www.deeplearningbook.com.br/>>. Acesso em: 11 out. 2019. Citado na página 25.
- ASSOCIATION, M. M. **The Complete MIDI 1.0 Detailed Specification**. 3. ed. La Habra, CA: MMA, 2014. 334 p. Citado 2 vezes nas páginas 11 e 18.
- BRAGA, A. d. P. **Redes Neurais Artificiais: Teoria e aplicações**. 4. ed. Livros Téc. e Cient., 2000. 280 p. Citado 3 vezes nas páginas 13, 14 e 15.
- BUENO, F. d. S. **Silveira Bueno: minidicionário da língua portuguesa**. 2. ed. São Paulo: FTD, 2007. Citado na página 1.
- CHEDIAK, A. **Harmonia e Improvisação**. 7. ed. Rio de Janeiro: Lumiar Editora, 1986. 353 p. Citado 5 vezes nas páginas 1, 5, 7, 9 e 10.
- COCA, A. E. **Mineração de estruturas musicais e composição automática utilizando redes complexas**. Tese (Doutorado) — Universidade de São Paulo, 2015. Citado na página 13.
- COCA, A. E.; ZHAO, L. Musical Scales Recognition via Deterministic Walk in a Graph. **Brazilian Conference on Intelligent Systems**, p. 151–156, 2017. Citado na página 6.
- DONAHUE, C.; MAO, H. H.; MCAULEY, J. The NES Music Database: A multi-instrumental dataset with expressive performance attributes. **International Society for Music Information Retrieval**. 2018. Citado na página 18.
- DONG, H.-W. et al. MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. **AAAI Conference on Artificial Intelligence**, p. 34–41, 2017. Citado na página 2.
- DONNELLY, P.; SHEPPARD, J. Evolving four-part harmony using genetic algorithms. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 6625 LNCS, p. 273–282, 2011. Citado na página 12.
- EBCIOĞLU, K. An Expert Four-part Harmonizing Chorales. **Computer Music Journal**. v. 12, n. 3, p. 43–51, 1988. Citado 2 vezes nas páginas 1 e 11.
- FACURE, M. **Gradiente Descendente: Um método poderoso e flexível para otimização iterativa**. 2017. Disponível em: <<https://matheusfacure.github.io/>>. Acesso em 15 set. 2019. Citado na página 25.
- FARIA, N. **A Arte da Improvisação**. 1. ed. Lumiar Editora, 1991. Citado na página 10.
- FREITAS, A. R. R. D.; GUIMARÃES, F. G.; RUELA, A. S. Computação evolutiva multiobjetivo para harmonização de melodias. **Simpósio Brasileiro de Pesquisa Operacional**. p. 2000–2011, 2011. Citado 3 vezes nas páginas 1, 11 e 12.
- GOOD, M. MusicXML for Notation and Analysis. HEWLETT, W.; SELFRIDGE-FIELD, E. (Ed.). **The Virtual Score: Representation, Retrieval, Restoration**. 12. ed. Los Altos, CA: MIT Press, 2001. cap. 8, p. 113–124. Citado na página 18.

- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. MIT Press, 2016. Citado na página 16.
- GUEST, I. **Harmonia - Método Prático**. 1. ed. Lumiar, 2010. 164 p. Citado na página 1.
- HARRIS, D. M.; HARRIS, S. L. **Digital Design and Computer Architecture**. 2. ed. Waltham, MA: Elsevier, 2013. 721 p. Citado 3 vezes nas páginas , 21 e 22.
- HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 16.
- HUANG, A.; WU, R. Deep Learning for Music. **ArXiv** p. 1–9, 2016. Citado na página 2.
- HURON, D. Music Information Processing Using the Humdrum Toolkit: Concepts, Examples, and Lessons. **Computer Music Journal**, v. 26, n. 2, p. 11–26, 2002. Citado na página 18.
- JERONYMO, D. C. **Inteligência Computacional: Redes Neurais**. Toledo: UTFPR, 2018. Citado na página 15.
- KAIZER, W. **Curso de Harmonia**. 1. ed. Vila Velha, ES. 2017. 75 p. Citado 2 vezes nas páginas 8 e 9.
- KERAS. **Layer Wrappers**. 2015. Disponível em: <<https://keras.io/layers/wrappers/>>. Acesso em: 3 set. 2019. Citado na página 24.
- KINGMA, D.; BA, J. Adam: A Method for Stochastic Optimization. **International Conference on Learning Representations**, 2014. Citado 2 vezes nas páginas 24 e 25.
- KOELLREUTER, H.-J. **Harmonia Funcional: Introdução À Teoria Das Funções Harmônicas**. 3. ed. Campos Elíseos, São Paulo: Ricordi Brasileira S. A., 1978. 71 p. Citado 2 vezes nas páginas 9 e 10.
- KOOPS, H. V.; MAGALHÃES, J. P.; HAAS, W. B. de. A functional approach to automatic melody harmonisation. **ACM SIGPLAN workshop on functional art, music, modeling and design** p. 47-58, 2013. Citado 4 vezes nas páginas , 1, 2 e 11.
- ANGLEY, P.; LAIRD, J. Artificial Intelligence and Intelligent Systems. 2006. Citado na página 13.
- LIM, H.; LEE, K. Chord Generation From Symbolic Melody Using BLSTM Networks. **Seoul National University, Korea**, 2017. Citado 2 vezes nas páginas 12 e 23.
- LIM, H.; RHYU, S.; LEE, K. **CSV Leadsheets Database**. 2017. Disponível em: <http://marg.snu.ac.kr/chord_generation/>. Acesso em: 4 jul. 2019. Citado 2 vezes nas páginas 2 e 18.
- LIU, H.-M.; YANG, Y.-H. Lead Sheet Generation and Arrangement by Conditional Generative Adversarial Network. **IEEE International Conference on Machine Learning and Applications**. 2018. p. 722–727. Citado na página 2.
- LIU, W. et al. A survey of deep neural network architectures and their applications. **Neurocomputing**, Elsevier B.V., v. 234, p. 11–26, 2017. Citado na página 16.
- LUDWIG, J. **Image Convolution**. Portland State University, Portland, 2017. Citado na página 16.

- MAKRIS, D.; KAYRDIS, I.; SIOUTAS, S. Automatic Melodic Harmonization. **International Joint Conference on Neural Networks**. IEEE, p. 146–165. 2013. Citado 2 vezes nas páginas 1 e 11.
- MAN, M. **The Largest MIDI Collection on the Internet**. 2015. Disponível em: <https://www.reddit.com/r/WeAreTheMusicMakers/comments/3ajwe4/the_largest_midi_collection_on_the_internet/>. Acesso em: 8 jun. 2019. Citado na página 18.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943. Citado na página 13.
- MED, B. **Teoria da Música**. 4. ed. Brasília: Musimed, 1996. 420 p. Citado 3 vezes nas páginas 5, 6 e 7.
- NOUGUÉ, C. **O que é o Sistema Temperado**. 2011. Disponível em: <<http://www.aboamusica.com.br/2011/11/o-que-e-o-sistema-temperado.html>>. Acesso em: 26 jun. 2019. Citado 2 vezes nas páginas 1 e 5.
- OLAH, C. **Understanding LSTM Networks**. 2015. Disponível em: <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 18 nov. 2019. Citado 2 vezes nas páginas 16 e 17.
- PHON-AMNUAISUK, S.; WIGGINS, G. A. The Four-Part Harmonisation Problem. **Artificial Intelligence and Simulation for Behaviour Symposium on Musical Creativity**, p. 28–34, 1999. Citado na página 2.
- RAFFEL, C. **The Lakh MIDI Dataset v0.1**. 2011. Disponível em: <<https://colinraffel.com/projects/lmd/>>. Acesso em: 8 jun. 2019. Citado na página 18.
- RAFFEL, C. **Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching**. 222 p. Tese — Columbia University, 2016. Citado na página 18.
- REIMERS, N.; GUREVYCH, I. Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks. **ArXiv**, 2017. Citado na página 25.
- RUSSELL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Prentice Hall, 2009. 1152 p. Citado na página 29.
- SANTOS, A. G. **Harmonia 1**. Cânone Musical, 2012. Citado na página 1.
- SCHOENBERG, A. **Theory of Harmony**. University of California Press, 1978. Citado na página 7.
- SHAFFER, K.; HUGHES, B.; MOSELEY, B. **Open Music Theory: Harmonic Functions**. 2014. Disponível em: <<http://openmusictheory.com/harmonicFunctions.html>>. Acesso em: 3 set. 2019. Citado na página 10.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing and Management**, Elsevier Ltd, v. 45, n. 4, p. 427–437, 2009. Citado 2 vezes nas páginas 12 e 13.
- SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, v. 15, p. 1929–1958, 2014. Citado na página 23.

- SUTSKEVER, I.; MARTENS, J.; HINTON, G. E. Generating Text with Recurrent Neural Networks. **The Auk**, v. 116, n. 2, p. 528–535, 2011. Citado na página 16.
- VEEN, F. van. **Neural Network Zoo Prequel: Cells and Layers**. 2017. Disponível em: <<https://www.asimovinstitute.org/neural-network-zoo-prequel-cells-layers/>>. Acesso em: 13 jul. 2019. Citado na página 16.
- VEEN, F. van; LEIJNEN, S. **A mostly complete chart of Neural Networks**. 2019. Disponível em: <www.asimovinstitute.org/neural-network-zoo/>. Acesso em: 9 jun. 2019. Citado na página 17.
- WIGGINS, G. et al. Evolutionary Methods for Musical Composition. **International Journal of Computing Anticipatory Systems**, 1998. Citado na página 1.

APÊNDICES

APÊNDICE A – TELAS DO FORMULÁRIO DE PESQUISA *ONLINE*

Um dos testes realizados para determinar a qualidade do sistema harmonizador automático foi a pesquisa de opinião usando um formulário *online* para realização de uma enquete. Nesta, seis perguntas foram feitas, sendo a primeira sobre a experiência do participante, e as demais sobre sua opinião de escolha para cinco músicas diferentes.

Figura A.1 – Tela inicial do formulário de pesquisa *online* com apresentação e primeira pergunta.

Harmonização Musical Automática Baseada Em Redes Neurais Artificiais
Análise de Qualidade

Olá!

Seja bem vindo a página de pesquisa para o Trabalho de Conclusão de Curso de Lucas Costa, do curso de Engenharia de Computação na UTFPR - Câmpus Toledo.

Este trabalho, intitulado "Harmonização Musical Automática Baseada Em Redes Neurais Artificiais" desenvolvido durante 2019, tem por objetivo a construção de uma ferramenta para a harmonização de melodias predefinidas, usando técnicas de inteligência artificial, com boa qualidade melódico-harmônica, coerente com as regras da teoria da harmonia e de forma a apresentar diversidade.

Portanto, se possível for despender 5 minutos de seu tempo para responder este questionário de opinião, ao ouvir as músicas, escolha qual mais lhe agrada.

Em avanço, muito obrigado! Sua participação é muito importante!

Você é músico? *

Sim
 Não

Se respondeu "sim" para a pergunta anterior, por favor especifique sua experiência:

Toca algum instrumento musical, é músico, estudante de música...

Fonte: Autoria Própria

A Figura A.1 mostra a tela inicial do formulário, onde foi apresentada uma introdução e um texto explicativo da pesquisa. A seguir, a pergunta sobre a experiência do respondente, com abertura opcional de descrever qual a experiência quando selecionada a alternativa positiva.

Na próxima página (Figura A.2), foram dispostas as opções de músicas. O participante podia então ouvir dois trechos de uma música com mesma melodia, porém de harmonias diferentes, e então decidir qual era preferida. Importante notar que em momento algum é revelado ao usuário qual das músicas fora harmonizada automaticamente, tornando o teste imparcial.

Figura A.2 – Segunda tela do formulário de pesquisa com disposição dos trechos de música audíveis para seleção do usuário.

Seleção de Músicas

Essa é a primeira música. Escolha qual das versões você prefere: *

Versão 1



 Lucas Costa
Música 1, Versão 1

 SOUND CLOUD

 Compartilhar

0:14

▶ 159

Versão 2



 Lucas Costa
Música 1, Versão 2

 SOUND CLOUD

 Compartilhar

0:14

▶ 147

Essa é a segunda música. Escolha qual das versões você prefere: *

Versão 1



 Lucas Costa

 SOUND CLOUD

Fonte: Autoria Própria

Figura A.3 – Tela final do formulário *online* indicando a conclusão da pesquisa.



Fonte: Autoria Própria

Ao final da pesquisa, a tela ilustrada na Figura C.1 é mostrada, indicando a conclusão bem-sucedida da pesquisa. A pesquisa está disponível em: form.jotformz.com/93036119173656.

APÊNDICE B – PARTITURAS E CIFRAS DA MÚSICA *ALABAMA WALTZ* DE HANK WILLIAMS

A seguir, as partituras e cifras resultantes da harmonização automática realizada pelo sistema, cuja análise e discussão foi feita na Subseção 6.4.2.

Figura B.1 – Partitura com cifra da música *Alabama Waltz* de Hank Williams analisada na Seção 6.4.2.

Alabama Waltz

Hank Williams

The musical score for "Alabama Waltz" by Hank Williams is presented in six staves. The key signature is common time (indicated by a 'C'). The music consists of six measures, each starting with a quarter note. Chords are indicated above the notes:

- Measure 1: C, C, F, F
- Measure 2: C, C, G7, C
- Measure 3: C, F, F, C, C
- Measure 4: C, C, F, F, C, G7, F
- Measure 5: G7, C, C, F, F
- Measure 6: C, C, C, C

Fonte: Autoria Própria. Adaptado de Hank Williams.

APÊNDICE C – CARTA DE SUBMISSÃO E ARTIGO SUBMETIDO À REVISTA *JOURNAL OF MATHEMATICS AND MUSIC*

Um artigo com o conteúdo do trabalho resumido foi submetido para publicação na revista *Journal of Mathematics and Music*. Nas próximas páginas, a carta de confirmação de submissão e o artigo na íntegra podem ser conferidos.

Figura C.1 – Carta de confirmação recebida pela submissão do artigo à revista.

 Journal of Mathematics and Music

- [!\[\]\(b3ad15bbadad8b30523e17a79d209ab4_img.jpg\) Home](#)
- [!\[\]\(d6455b74f296db3deede18dde8db75b0_img.jpg\) Author](#)
- [!\[\]\(f8e20590fa768e54bfed291cf8b550b1_img.jpg\) Review](#)

Submission Confirmation

[!\[\]\(afcc9020b459ac6bf2ebe6cc629da5cb_img.jpg\) Print](#)

Thank you for your submission

Submitted to	Journal of Mathematics and Music
Manuscript ID	TMAM-2019-0049
Title	Automatic Musical Harmonization Based On Artificial Neural Networks
Authors	Piccioni Costa, Lucas Coca Salazar, Andrés Eduardo
Date Submitted	21-Nov-2019

Author Dashboard

Fonte: *Journal of Mathematics and Music*.

Automatic Musical Harmonization Based On Artificial Neural Networks

Lucas F. P. Costa¹, Andrés Eduardo Coca S.²

Curso de Engenharia de Computação

Universidade Tecnológica Federal do Paraná (UTFPR)

Toledo, Brazil

¹lucascosta74@gmail.com; ²andressalazar@utfpr.edu.br

Abstract—In musical terms, harmony is defined as the art of combining several musical notes simultaneously to create a coherent sound to human ears, with the purpose of accompanying, framing and filling. However, the elaboration of a harmony for a particular melody is not a trivial task, since it requires music knowledge and experience, as well as intense study of music theory, which takes time and may even take years to reach reasonable skills. Thus, the development of a system capable of automatically harmonizing melodies is beneficial. In the present paper, it was proposed to build a tool for the harmonization of predefined melodies, using artificial intelligence techniques, that are realistic, consistent with the rules of harmony theory and in order to present diversity. Thus, a melody, defined with symbolic and properly processed musical data, serves as input to an artificial neural network, generating for this chords, having been trained based on a music database containing melody and harmony. The system output is the chord sequence to be applied to the melody. Due to the complexity of obtaining and analyzing results, quantitative and qualitative tests were performed, using from performance measures calculated from confusion matrices to online polling with anonymous volunteer participants opinion. In the end, it can be concluded that the system was able to generate harmonies aligned with the objectives.

Index Terms—Musical Theory. Intelligent Systems. Performance Evaluation.

I. INTRODUCTION

Harmony is understood to be the well-ordered arrangement between the parts of a whole, meaning concord, order, proportion or symmetry. Aiming at the theory of music, one can understand the development of harmony as the art of stringing together various sets of sound frequencies or pitches which, when played simultaneously, form the so-called chords in order to create a coherent sound that has meaning to human ears, commonly attached to a melody or musical line, serving purpose to accompany and frame [1].

Nevertheless, building the harmony that will accompany a melody is a work that requires experience in music matters, involving in-depth knowledge of music theory, more specifically harmony theory [2]. Observing harmony from the point of view of a process, harmonization is seen as the pursuit of a harmony for a given melody, for which several options exist, the amount of which is generally limited by sets of aesthetic rules and conventions [3].

As described in [4], harmonization is the extension and natural application of harmonic analysis. Researching the steps that constitute such a process, as well as a way to automate

them, makes a valuable contribution to the study of music, artificial intelligence, and the field of Musical Information Retrieval (MIR) [5].

In this sense, several attempts at automated solution have been made, with algorithms often focused on evolutionary methods usually linked to Genetic Algorithms (GA). [6], [7], or rule based systems [5], [8]. On the other hand, in [9] a performance comparison between the use of rules systems and GAs was made, concluding that the quality of the result depends on the amount of knowledge that the system has about the subject.

Neural Networks and *Deep Learning* were also employed in trying to solve this problem. In [10] a possibility of composing harmonies using Bidirectional Long Short-Term Memory networks (BLSTM) is presented, trained on symbolic data from real song harmonies. Besides this, the works [11]–[13] are examples that use such methodologies, but with the objective of generating whole songs, from the melody to the arrangement for the performance.

In this work, a tool was developed for the harmonization of a melody, using artificial intelligence techniques, more specifically machine learning using Artificial Neural Networks (ANN). A database called *CSV Leadsheet Database* [14] containing songs composed of melody and harmony served to provide training data for an ANN. After the training phase, this network was able to generate harmonies by melody reference, being the system input. The new harmony was generated by the system based on the learning from the previously analyzed songs and the system output was a harmony for the melody. Statistical and performance measures were calculated on the results to describe the quality of the production, as well as subjective auditory tests involving human participants in the response to an online survey to then conclude on the system performance level.

This paper is divided into sections, of which: Sections II and III, respectively, briefly address the theory of musical functional harmony and establish the methodology used. Results are exposed in Section IV, whose conclusions define Section V.

II. FUNCTIONAL HARMONY TECHNIQUE

According to [15], "harmony is the study of simultaneous sounds (chords) and how they can be united in relation to their

architectural, melodic and rhythmic values; and its meaning, and its relative weight to each other". The study of harmony establishes precise rules and important concepts, some of which will be presented below.

A common way to build chords is to make use of triads, consisting of three notes played at the same time (three-note chords), usually made by overlapping notes at intervals of thirds. A range of thirds may be longer or shorter, with the first representing a distance of three semitones from a starting note and the second, four. For example, from C to E, there is a distance of four semitones, a major third interval. But between C and D \sharp , is an interval of three semitones, a minor third. Depending on the quality of the thirds intervals between the notes of a triad, one may call the chord major, minor, augmented or diminished.

To refer to the chord group that can be performed from a scale, the expression harmonic field is used. Within it are contained mainly major and minor chords which constitute a functional relationship to each other. The construction of a harmonic field occurs in order to obtain a chord for each note of the selected scale, using the third overlap and keeping the chord notes within the same scale. For example, for the major C scale, for each note there will be a triad. The first, on the note C, will be with the notes C, E and G; for the second (D), the notes D, F and A, and so on. Chords formed from the major scale can be represented as in Figure 1. In a song elaborated in this way and without any modulation, the assembled chords will be within this harmonic field and each chord will have a different function. [16].

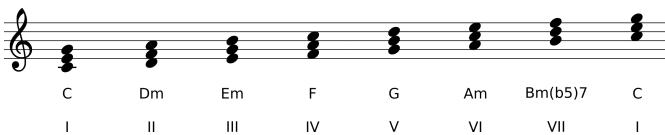


Figure 1: The seven musical chords available from the C major scale, and their respective degrees, represented using musical notation.

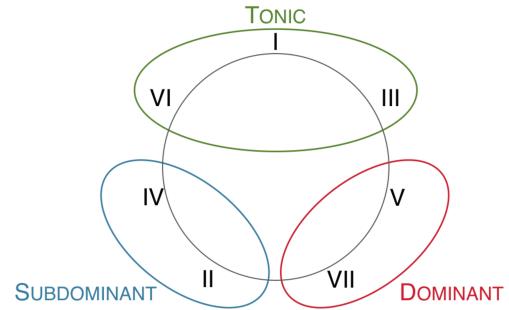
Function is understood, in terms of harmony, as the characteristic of a chord and its value in relation to the others, in the context of the work, which is determined by the relational sum of all chords with a tonal center, by which the concept of tonality. The tone is then defined by the set of categories or functions traditionally called tonic, dominant and subdominant, respectively. Each can have its functions defined as [2], [17]

- **Tonic:** grants a feeling of rest, having start or end functionality. Its only representative is the harmonic field degree I;
- **Dominant:** gives the sense of approach movement, as it generates tension and need for resolution, usually provided by the tonic. It is the degree V of the harmonic field;
- **Subdominant:** also related to movement, but of distance, a middle ground between tonic and dominant, being

represented by degree IV.

For the other chords (or degrees) present in the harmonic field, their functions are inherited from the main functions, called secondary functions, and these chords are then called relative. According to [17]: "(relative) chords have the harmonic meaning of that tonic, subdominant or dominant, of which they are third neighbors". Generally speaking, the one illustrated in Figure 2 applies [18].

Figure 2: Triads arranged in the circle of thirds, labeled by harmonic functions [18].



Respecting the functions of each degree, the standard chord movement within the music is established: the dominant function is commonly preceded by the tonic, giving the resolution of dissonance; the subdominant function has greater freedom, however its sense of movement usually leads to the dominant one; the tonic function in turn is usually applied at the beginning or end of a song or musical phrase, providing the idea of completion [2].

Functional harmony, therefore, aims to study the functioning of these functions, as well as establish rules and concepts for their correct use within music. By understanding the functions it is possible to harmonize songs in a coherent way that make sense to the listeners. Also being rules that, when respected, indicate the quality of the work.

III. PROPOSED METHOD FOR AUTOMATIC MUSICAL HARMONIZATION

The proposed method for this paper follows the steps described: from a database of music encoded in symbolic data containing both melody and harmony, an ANN will be trained and then used to generate new harmonies conditioned by an input melody, in order to generate a chord per measure. This, in turn, will be processed first, then serve as input to a trained neural network based on a database also processed. The result will be a harmony conditioned by the melody, which will be evaluated qualitatively and quantitatively.

A. Collection and Processing of Data

For application of ANN, data will be required to perform the training. In this paper, the database *CSV Leadsheet Database* [14] was selected, which contains 2252 songs from rock, pop, country, jazz, folk, R&B, children's songs, etc., already

processed in CSV format, whose files contain organized information about tempo, measure, tonic and mode of the currently played chord, melody note, duration, and other information for each song.

In order to facilitate the later training phase of ANN, we must first process the data to make it standardized. Then, to serve as input to the system, this processed information needs to be encrypted before the network can continue execution. Therefore, the processing of the database takes place in two stages: standardization of training data and preparation and coding thereof, which will be explained below.

Standardization happens in terms of:

- **Tonality:** the melodies and harmonies are transposed to a common tone so that there is data consistency and comparability. In this case, all songs in the database were transposed to C major tone; This guarantees a set of expected notes (C major scale) and defines a common harmonic field.
- **Notes duration:** in the database, the duration of each note is standardized to the values shown in Table I. Given the fact that each song contains its own time signature, there must be normalization, as for each measure of songs with different tempo, the sum of the note durations of each measure would have different values. To balance the sum of durations, the duration of each note is multiplied by the inverse of the current time signature.

Table I: Default duration values of database musical notes.

Rhythmic figure	Duration
Hole note (.)	16.0
Half note (d)	8.0
Quarter note (j)	4.0
Eighth note (♪)	2.0
...	...

Being d the length of a musical note, what is desired is that $\sum d = 16$ regardless of the time signature of the song. To meet this requirement, being F_c the time signature, must do:

$$d \times F_c^{-1} = d \times \frac{\text{Rhythmic figure type}}{\text{Number of times of the rhythmic figure}}.$$

This way, the measures will always have equal total duration. Finally, for each measure of the input melody, there will be a vector of time-normalized and transposed notes.

- **Chord types:** the problem with chords is the vast amount of possibilities for their formation. Therefore, a simplification was made, keeping only major, major seventh, and minor chords for all chromatic possibilities, resulting in a total of 36 chord classes.

Keeping in mind the need for testing to better determine the performance of the auto harmonizer, five different ways of selecting the data were prepared to perform the ANN training step. These sets take into account different aspects of the relationship between melody and harmony of songs.

However, they all have in common the consideration of only the first chord of each measure, as it is the goal: to generate one chord per measure. These are the selection forms (S_i) described below:

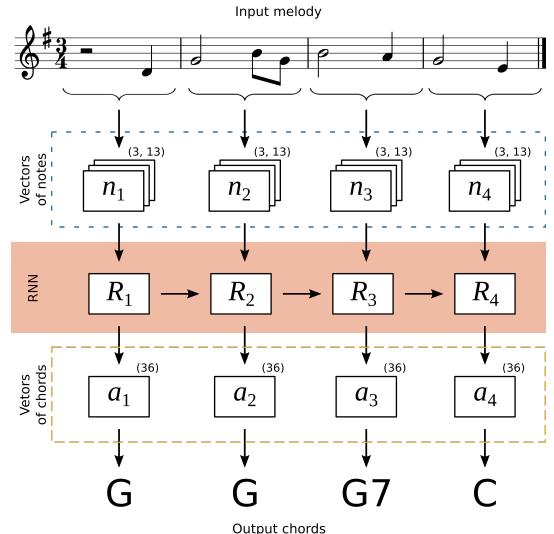
- **Selection 1 (S_1):** only the first note of each measure, to see if this relationship is enough to determine harmony;
- **Selection 2 (S_2):** the first note and the middle note of each measure, as they usually determine the strongest moments, and may give a good indication of the current chord;
- **Selection 3 (S_3):** the set of all notes of the first half of the measure considering the rests;
- **Selection 4 (S_4):** all notes of the bar, also considering the rests;
- **Selection 5 (S_5):** the notes of the bar discarding the rests, giving importance only to the sounds.

The sets are then coded for ANN input. Both notes and chords are vector encoded using the one-hot method [19]. Thus, a 13-position vector represents the notes (12 possible musical notes plus a position to identify a rest), and another 36-position vector indicates the chord (12 major chord classes, plus 12 major with seventh and 12 minor).

B. ANN Functioning

Having input the coded note vectors, these are taken to ANN, which will be responsible for giving chords also coded in the output. The process of data preparation, ANN input and output can be simplified as illustrated in Figure 3.

Figure 3: Overview of the proposed system with melody processing in note vectors serving as input to the RNN, whose output is a vector of chords that are later decoded.



Each note is transformed into a coded vector, so each measure will have multiple note vectors. For the case illustrated, the second measure holds the largest number of notes (three), so for input to be balanced, each measure produces three note vectors. In the other measures, where there is not the same

amount of notes, an empty vector is added. Then the four illustrated bars are processed into four sets of n note vectors, each set having three note vectors of size 13. The ANN then receives at each R iteration these sets of vectors, and back-propagates (indicated by the horizontal arrows). Also, at each iteration, a chord vector a of size 36 is produced, referring to the input measure, which can finally be decoded to a known name, characterizing the output chords. In this work, three recurrent ANN models were used (RNN): simple RNN (RNN_1), one layer LSTM (RNN_2) and also a two layer LSTM (RNN_3).

C. Tests performed

The overall output of the system is harmonies for the melodies served at the entrance. For the test set selected in the database processing period, the generated harmonies are stored separately to perform survey results and to measure the system performance rate quantitatively. This rate survey and measurement takes place by applying the confusion matrix and its respective rates and measures, such as sensitivity (TPR), specificity (TNR), accuracy and F_β -score. In addition, the overall quality of the classification can be verified by averaging the calculated measures for each of the classes, treating them equally, a method called the macro-average (M) [20].

Qualitative tests were performed to infer conclusions about the quality of the system output, and are important because they reveal issues that cannot be easily explained by quantitative measures. To this end, the way proposed to raise qualitative results is described below.

1) *Online Survey*: An interesting factor to note is how the general public tastes behave through the matching results generated by the proposed automatic matching system. Therefore, a survey was made with some questions for two groups: 1) people who have musical experience, either with composition, theoretical or practical study, or even musicians; and 2) people without any technical musical experience beyond their natural prior concepts, since people just by listening to music have an idea of what is appropriate or not intrinsically. Both were led to answer which harmonization they prefer from two possibilities: the original and the automatic for five songs, without revealing to the participants which harmonization is which.

IV. RESULTS

The results of applying the proposed methodology for automatic harmonization of melodies, these being the results of model training and the application of quantitative and qualitative tests, are described in detail in subsequent sections.

In all, 2249 songs were processed, totaling 342723 notes and rests and 75087 chords, the same number of measures as it is considered one chord per measure. To perform the training and testing stages of the proposed ANN models, the song set was divided into two, where 80% was for training, and the remaining 20% for testing, that is, using Holdout 80%.

We have the models with empirically defined hyperparameters according to Table II, for applied data selection and values of neurons, batches, dropout and epochs.

Table II: Empirically selected final settings for each RNN model tested.

Model	Selection	Neurons	Batches	Dropout	Epochs
RNN_1	S_3	256	256	0.5	500
RNN_2	S_5	128	256	0.5	50
RNN_3	S_4	256	128	0.5	15

Following is a deeper analysis of the response of each model with comparative criteria to better determine the final choice for the automatic matching system. An example of the result of a melody harmonized by each model, as well as the original harmony, is shown in Figure 4 for the song excerpt *December, 1963 (Oh What a Night)* from *Frankie Valli & The Four Seasons* between bars 11 and 18. Note that the RNN_1 model line is the most out of line with the original, and that RNN_3 presents closer result with the choice of more similar chords. Between the models, there is total agreement in measures 1, 4 and 7, although these are different from the original, and may infer an alternative of harmony. In bars 5 and 8, RNN_2 and RNN_3 selected the same chord as the original, but using the 7th, showing correct behavior in the understanding that these should be used at higher tension points that will need further resolution. Finally, the objective of achieving a system that presents diversity of the original harmonies is concluded, since for the same melody, three different harmonizations were created. In addition, the diversity can be proved by the possibility of using 36 chord classes.

Figure 4: Eight measures excerpt from the song *December, 1963 (Oh What a Night)* from *Frankie Valli & The Four Seasons* with original harmonies generated by each proposed model.

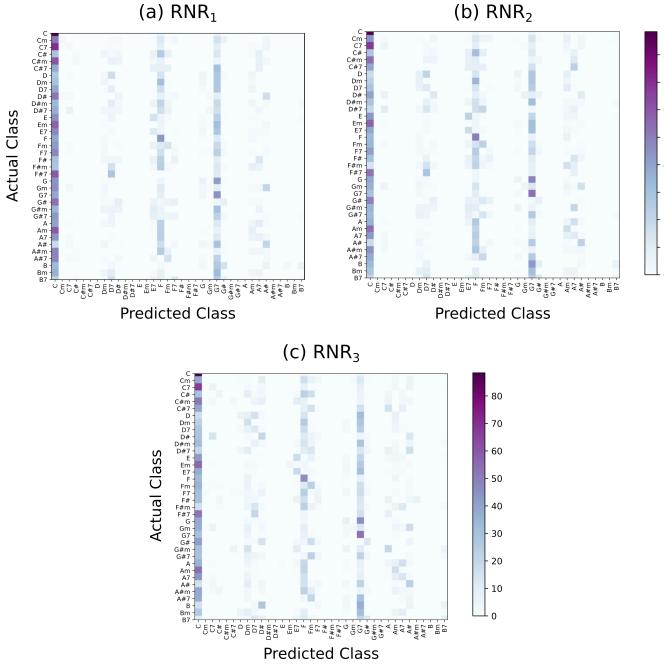
Original: F F F Am RNN ₁ : C C Am C RNN ₂ : C F Am C RNN ₃ : C F F C	
--	--

Original: G F F G RNN ₁ : G C Am C RNN ₂ : G7 F Am G7 RNN ₃ : G7 F Am G7	
--	--

The previous analysis allows us to say that there were differences between the performance and understanding of each model. Further analysis can be done by looking at measures such as those defined in Section III-C.

The confusion matrix of the three systems is illustrated in Figure 5 for (a) RNN_1 , (b) RNN_2 and (c) RNN_3 . Analyzing the confusion matrices, we notice the extensive use of main chords of the tonic functions (C, F and G) and especially the dominant chord (G7). Generalizing between models, some points that draw attention are the correct prediction of the E7 and G7 chord, in addition to the extensive exchange of G for G7. It is also interesting to note the presence of slight vertical

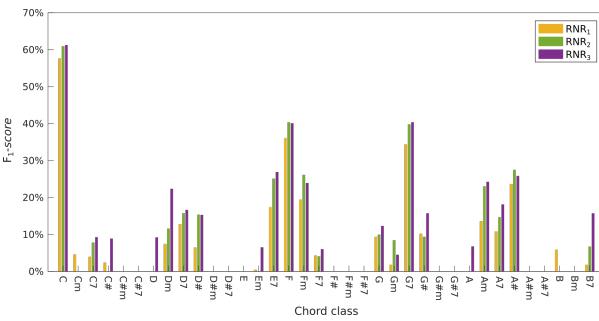
Figure 5: Confusion matrices for the three proposed models: (a) RNN₁, (b) RNN₂ and (c) RNN₃, with percentage of hits referenced by color.



lines for the Dm and A \sharp chords, suggesting the generalization and simplification made by the models during the composition. Finally, among all models, RNN₃ visually performs better than the others, being the one that had the most hits, with an accuracy of 41%.

To continue, the graph of de F_1 -score is presented and analyzed (Figure 6), equally weighting accuracy and sensitivity for each chord class. For several chords, it is clear that there were no hits. Some peaks are noted, for the chords of C, F and G7 mainly, as can also be seen in the confusion matrices of Figure 5. Among all, there is greater participation of the model RNN₃, with higher peaks and sometimes appearing alone, as in chords D and A.

Figure 6: F_1 -score of individual chord classes for each of the proposed models.



The final results of TPR_M , TNR_M , F_1 -score_M and accuracy

are shown in Table III. Although the model RNN₂ has greater accuracy than RNN₁, this has greater accuracy than all, i.e. it is the model able to select the largest number of relevant items. Nevertheless, the model RNN₃ has higher sensitivity value and F_1 -score, being better balanced. Moreover, it is also the model with the highest accuracy value, surpassing the others.

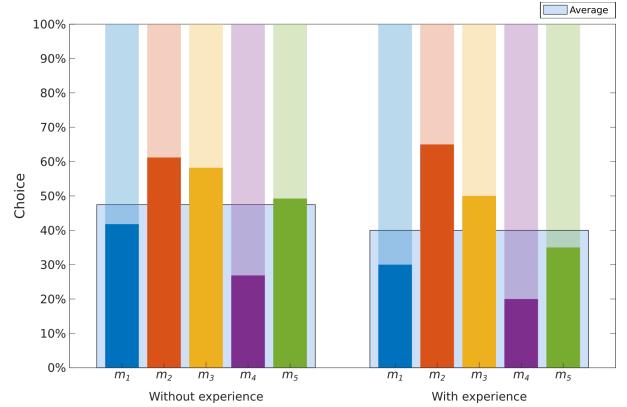
Table III: Macro-average of performance rates (M) and accuracy values for the proposed models.

Model	TPR_M	TNR_M	F_1 -score _M	Accuracy
RNN ₁	18.79%	7.89%	7.91%	37.20%
RNN ₂	13.63%	9.46%	9.64%	40.12%
RNN ₃	18.51%	10.81%	11.40%	41.00%

1) *Online Survey Answers:* In all, six questions were asked: the first about the respondent's musical experience and the other five about works comparison, where the participant listens to short pieces of music of a maximum of 33 seconds for comparison, with two options always containing the original and automatic matching, but this was transparent to the user. The songs used in the form were: Happy Birthday (14 seconds), *O Grande Amor* (32 seconds), Rock of Ages (20 seconds), *Samba Pa Ti* (33 seconds) and *Zie Ginds Komt de Stoomboot* (20 seconds), denoted as m_1 to m_5 , respectively.

In all, 87 people participated in the survey, of which 20 reported having a diverse musical experience, and the remaining 67 did not. The Figure 7 shows the percentage of times automatic harmonizations were chosen individually and the average for both groups surveyed.

Figure 7: Online survey results with percentage of choice for each song and average choice made for each group: experienced and non-experienced, where lower bars are preference for auto harmony and upper bars, original harmony.



Analysis of the graph reveals different preferences between the groups studied. However, there is similarity in the choice of songs: it is possible to see the same decreasing order of preferences for the songs (m_2 , m_3 , m_5 , m_1 and m_4). Notably, participants preferred automatic matching to the original for m_2 , with 61.20% to the group said without experience (1), and 65% to the second group (2). The least favorite song, or

with original harmony as a favorite was m_4 , with 26.87% and 20% for groups 1 and 2, respectively.

Mean and standard deviation values are shown in Table IV for both groups surveyed. Analyzing group 1, it can be stated that the respondents were divided between automatic and original harmonizations, without being able to say exactly their preference, unlike group 2, where it is possible to observe a better definition by the original harmonies. This fact, however, allows us to assume that the system was able to develop realistic harmonies, leading to events where participants had a greater preference for automatic harmonies. In addition, most people preferred the original harmonies averaging 43.73%.

Table IV: Average (μ) and standard deviation (σ) for preference over music harmonized automatically for the two groups analyzed 1) without and 2) with musical experience.

Group	μ	σ
1	47.46%	13.82%
2	40.00%	17.68%
Total	43.73%	15.75%

V. CONCLUSIONS

A system for automatic harmonization based on musical melodies has been proposed, tested and presented. To propose the system, a review of basic musical rules and theory of functional harmony was made. The state of the art was raised, revealing the importance of solving the problem of automatic musical harmonization.

The proposed methodology included the application of a database of songs containing harmony and melody, and ANNs as a predictive and training system for supervised learning. Different ways were presented to evaluate the results quantitatively and qualitatively.

The final system was a recurrent type ANN that had a double layer of LSTM with 256 neurons, 256 batches, 0.5 dropout and 15 training epochs, a model that due to its complexity was best suited to perform the matching task. The accuracy of the model was 41%, a reasonable value considering the results found in the literature and compared to the random result which, taking into account the 36 available chord classes and using a uniform distribution, is likely to select a chord equal to 2.78%.

The results generally showed that the system is actually able to simplify harmonies, since harmony rules were not disregarded during the process. This fact also explains the general preference of users for original harmonizations due to their higher compositional level.

REFERENCES

- [1] I. Guest, *Harmonia - Método Prático*, 1st ed. Lumiar, 2010.
- [2] A. Chediak, *Harmonia e Improvisação*, 7th ed., J. Máximo, Ed. Rio de Janeiro: Lumiar Editora, 1986.
- [3] C. Nougué, “O que é o Sistema Temperado,” 2011. [Online]. Available: <http://www.aboamusica.com.br/2011/11/o-que-e-o-sistema-temperado.html>
- [4] D. Makris, I. Kaydis, and S. Sioutas, “Automatic Melodic Harmonization,” in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, aug 2013, pp. 146–165.
- [5] H. V. Koops, J. P. Magalhães, and W. B. de Haas, “A functional approach to automatic melody harmonisation,” p. 12, 2013.
- [6] A. R. R. D. Freitas, F. G. Guimarães, and A. S. Ruela, “Computação evolutiva multiobjetivo para harmonização de melodias,” pp. 2000–2011, 2011.
- [7] G. Wiggins, G. Papadopoulos, S. Phon-annuisuk, and A. Tuson, “Evolutionary Methods for Musical Composition,” in *International Journal of Computing Anticipatory Systems*, 1998.
- [8] K. Ebcioglu, “An Expert Four-part Harmonizing Chorales,” vol. 12, no. 3, pp. 43–51, 1988.
- [9] S. Phon-Annusuk and G. A. Wiggins, “The Four-Part Harmonisation Problem.pdf,” pp. 28–34, 1999.
- [10] H. Lim and K. Lee, “Chord Generation from Symbolic Melody Using BLSTM Networks,” in *International Society for Music Information Retrieval Conference*, 2017.
- [11] A. Huang and R. Wu, “Deep Learning for Music,” pp. 1–9, 2016.
- [12] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, “MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment,” pp. 34–41, sep 2017.
- [13] H.-M. Liu and Y.-H. Yang, “Lead Sheet Generation and Arrangement by Conditional Generative Adversarial Network,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, dec 2018, pp. 722–727.
- [14] H. Lim, S. Rhyu, and K. Lee, “CSV Leadsheet Database,” 2017. [Online]. Available: http://marg.snu.ac.kr/chord{_\}generation/
- [15] A. Schoenberg, *Theory of Harmony*. University of California Press, 1978.
- [16] W. Kaizer, *Curso de Harmonia*, 1st ed., Vila Velha, 2017.
- [17] H.-J. Koellreuter, *Harmonia Funcional: Introdução À Teoria Das Funções Harmônicas*, 3rd ed. Campos Elíseos, São Paulo: Ricordi Brasileira S. A., 1978.
- [18] K. Shaffer, B. Hughes, and B. Moseley, “Open Music Theory: Harmonic Functions,” 2014. [Online]. Available: <http://openmusictheory.com/harmonicFunctions.html>
- [19] D. M. Harris and S. L. Harris, *Digital Design and Computer Architecture*, 2nd ed. Waltham, MA: Elsevier, 2013.
- [20] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009.