Geração Procedural de Composição Musical Utilizando Cadeia de Markov

Gustavo Yamauchi Souza Reis, Felipe Danieli Nishino e Alessandro Ranulfo Lima Nery
Centro Universitário Senac Santo Amaro
Av. Engenheiro Eusébio Stevaux, 823, Santo Amaro, São Paulo - SP, 04696-000
yamauchi_gu@hotmail.com, felipenishino@protonmail.com,
limanery@yahoo.com.br

RESUMO

A música é uma produção humana que está inserida em diversas ocasiosões, festas, eventos, em filmes ou apenas sendo ouvida enquanto a pessoa realiza uma atividade ou apenas relaxa. A música pode ser utilizada para realçar as emoções da ocoasião, por exemplo, a entrada da noiva na cerimônia de casamento. Este trabalho apresenta uma proposta de um algoritmo que, por meio da aplicação de conceitos de geração procedural, será capaz de gerar uma composição em tempo real para o usuário. Cadeias de Markov são usadas para determinar como os diferentes componentes da música serão gerados. As Cadeias de Markov serão abstraídas a partir de dados obtidos em análises de composições musicais já existentes junto dos modelos Circumplexo Tellegen-Watson-Clark e Modelo de Thayer, desenvolvidos para a categorização de emoções. O usuário terá o poder de decidir quais emoções serão o foco na geração de músicas, de acordo com seus objetivos. Quando testadas de acordo com a percepção humana, com o uso das Cadeias de Markov foi possível produzir uma música para transmitir uma emoção específica, deixando claro que alterações específicas podem melhorar as composições.

PALAVRAS CHAVE. Cadeia de Markov, Geração Procedural, Composição Musical.

EST&MP – Estatística e Modelos Probabilísticos; SIM – Simulação

ABSTRACT

Music is consumed by many people, accompanying situations and emotions of such moments. This paper presents a proposal for an algorithm that, by applying of procedural generation concepts, will be able to generate a composition in real time for the user. Markov chains are used to determine how the different components of the music will be generated. The Markov Chains will be abstracted from data obtained from analyses of existing music compositions with the Tellegen-Watson-Clark Circumplex and Thayer model, developed for emotion categorization. The user will have the power to decide which emotions will be the focus of the music generation, according to its objectives. When tested according to human perception, with the use of Markov Chains it was possible to produce a song to transmit a specific emotion, making it clear that specific changes can improve the compositions.

KEYWORDS. Markov Chain. Procedural Generation. Musical Composition.

EST&MP – Estatística e Modelos Probabilísticos; SIM – Simulação

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

1. Introdução

1.1. Contexto

A geração procedural é uma técnica de computação que produz conteúdos em ambientes simulados, essa geração se dá através de um algoritmo sem intervenção humana, é uma forma para criação de conteúdos em grande escala ou em ambientes que se beneficiam de um conteúdo gerado de maneira aleatória, como por exemplo, em mapas de jogos digitais, provendo um baixo custo de mão de obra e armazenamento [Togelius et al., 2016]. O desafio de utilizar esta técnica é produzir algo que se aproxima de uma produção humana. Esta técnica é frequentemente aplicada em geração de conteúdos gráficos ou dados para uma aplicação específica, mas existe também a possibilidade de aplicá-la na música, por exemplo.

Tendo como base processos tradicionais de composição musical, mais especificamente, sem o uso de ferramentas autônomas durante sua produção, observa-se que além do componente estrutural lógico, existem também nuances que se originam do artista da música, seja durante a produção ou mesmo durante a performance [Linskens, 2014].

A geração procedural pode ser influenciada e controlada de diversas maneiras, aumentando as possibilidades do produto final. Tratando de geração de composições musicais, algo que tem influência nos ouvintes são as emoções [Simões, 2012]. A alteração das regras da geração procedural de acordo com emoções pode ter como resultado um ambiente mais próximo do real ou próximo de algo feito por um humano.

O conceito de composições musicais geradas totalmente por computadores já é estudado há um tempo considerável, como pode ser notado pelos registros da composição "*Illiac Suite*", gerada a partir do algoritmo projetado por Lejaren Hiller e Leonard Isaacson [Nierhaus, 2008], que foi produzida na década de 50. Desde então, diversos outros estudos abordaram esta área. Observando as metodologias existentes que podem ser aplicadas em algoritmos para composição musical, encontra-se a utilização de modelos de Cadeia de Markov.

A Cadeia de Markov é um processo estocástico que destaca-se pela sua capacidade preditiva, ela depende apenas do estado atual dentro do espaço de estados do sistema sendo processado para transitar para o próximo estado. Tal funcionalidade é definida formalmente por uma matriz de transição composta por probabilidades que descrevem o comportamento estocástico para cada um dos estados possíveis do sistema [Allen, 1990]. Aplicando este esquema de modelagem no contexto musical, cada transição de estado de uma Cadeia de notas leva a uma nova nota musical [Linskens, 2014]. Dessa forma, cada nota depende apenas do estado atual e das probabilidades de transição para cada outra nota possível em dada escala, e não da nota anterior (estado anterior) para cada tempo. Ao final da cadeia existirá uma sequência de notas baseada nas tabelas de transições em relação ao tempo.

1.2. Problemática

A arte é uma atividade humana sendo composta por diversos tipos de expressão, sendo a música uma das formas de expressão da arte. A música está presente em diversos momentos: rádios em carros, festas de aniversário, trilhas sonoras, em apresentações de orquestra entre outros. A música é uma ferramenta que pode potencializar as emoções, por exemplo, em uma cena de um filme, a trilha sonora pode potencializar a emoção do espectadores, cena de um enterro, cena de vitória em uma partida esportiva [Thoma et al., 2013].

Este trabalho se propõe a investigar: Sabendo que as produções musicais influenciam nas emoções, como gerar uma composição musical através de técnicas computacionais que conseguem transmitir emoções?

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Um algoritmo pode dar continuidade em uma música produzida por um humano, chegando bem próximo da original, como observado no trabalho de [do Nascimento, 2021], que busca produzir um trecho musical partindo de trechos existentes utilizando a Cadeia de Markov para geração das notas subsequentes. A discussão sobre utilizar o algoritmo de gerar músicas para substituir o trabalho dos artistas não cabe neste estudo.

1.3. Objetivo

Esse trabalho tem como objetivo produzir um algoritmo de geração procedural de composições musicais, usando o modelo de Cadeia de Markov e suas tecnologias como regra de produção. O algoritmo final será aplicado para obter uma sequência de notas musicais, sendo possível ao decorrer do tempo fornecer uma entrada de dados correspondentes à emoção desejada para os trechos seguintes da música.

1.4. Justificativa

Este trabalho produz uma geração procedural que seja contínua e durante o processo possa ser adaptada de acordo com as informações (emoções) selecionadas por um usuário. O artigo de [Ramanto e Maulidevi, 2017] desenvolve a composição com emoções, contudo não são escolhidas durante o andamento da geração.

A geração sendo adaptável pode produzir uma composição diferente de acordo com as entradas, podendo ser aplicada em ambientes de diferentes escopos como: jogos, ambientes simulados, cinematografia, apresentações, etc. Outro aspecto da solução proposta neste trabalho é a autonomia, que junto à grande variedade de situações em que pode ser aplicado, viabiliza a criação de músicas para artistas independentes e usuários que não são necessariamente músicos.

2. Revisão Bibliográfica

2.1. Música

Um som é gerado a partir de uma alteração na pressão do ar criando ondas que se propagam através de um meio (sólido, líquido ou gasoso). A diversidade de meios implicam que haja uma alteração nas característica do som, podendo alterar, por exemplo, sua amplitude [Olson, 1967]. O som pode ser classificado conforme a frequência, podendo ser um ruído (som irregular) ou nota musical (som regular). Um instrumento produz sons regulares, por exemplo [Med, 1996].

Uma escala consiste em uma sequência de notas musicais consecutivas, ordenada de maneira descendente ou ascendente. Essa sequência pode se restringir a notas disponíveis em um sistema musical em específico, por exemplo. As escalas têm padrões ou regras definidas, sendo elas a quantidade de notas ou então a forma que a progressão entre elas é montada [Olson, 1967].

As notas separadamente tocadas dentro da escala com uma variação de progressão são chamadas de melodia, podendo ser cantada ou então tocada por um instrumento específico. Acordes são três ou mais notas escutadas simultaneamente durante um tempo específico. Assim como as notas, os acordes também seguem regras de composição. [Lyra, 1999]. Uma sequência de acordes dentro de uma escala é definida como um campo harmônico.

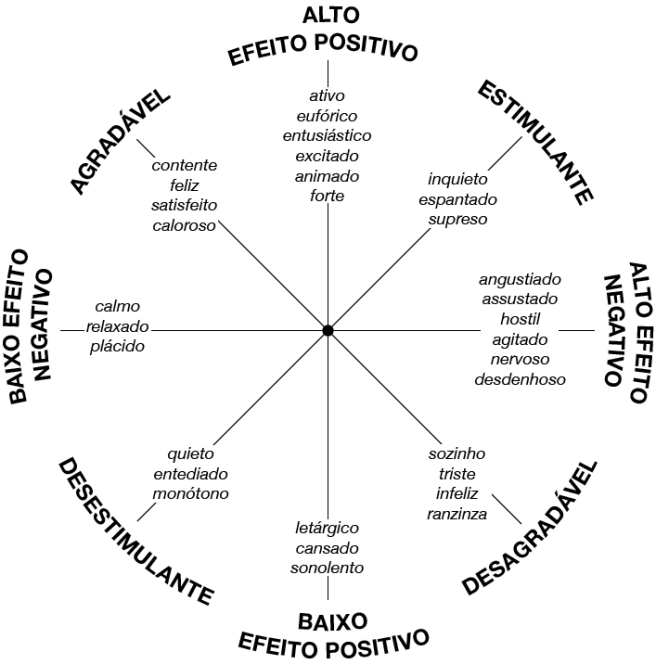
2.1.1. Emoções na Música

As emoções são sentidas de maneira forte e momentânea com uma duração pequena [Cezar e Jucá-Vasconcelos, 2016]. Músicas por serem no geral curtas transmitem emoções, de outra forma, as operas podem gerar sentimentos por terem duração mais elevada.

Para definir uma emoção em uma música existem diversas formas de classificação. Uma delas é utilizando o Circumplexo de Tellegen-Watson-Clark, esse é um dos modelos mais complexos que foi estudado no trabalho de [Ramanto e Maulidevi, 2017]. Apesar de ser bem fundamentado, de acordo com os textos de [Watson e Tellegen, 1985] e o trabalho de Ramanto, o circumplexo se

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Figura 1: Circumplexo de Tellegen-Watson-Clark para classificação de emoções



Fonte: Adaptado de [Watson e Tellegen, 1985]

mostra um modelo de classificação genérico, de forma que aparenta ter uma correlação intuitiva com o contexto de música.

O Modelo de Thayer provê uma melhor classificação de emoções em músicas. Ele se assemelha com o Circumplexo de Tellegen-Watson-Clark, porém de uma forma mais simplificada, montando um plano cartesiano onde o eixo das abscisas representa a valência e o eixo das ordenadas representa a energia [Thayer, 1989] apud [Lu et al., 2006].

3. Desenvolvimento

Neste capítulo serão descritos o planejamento e a implementação do projeto, listando as ferramentas e metodologias utilizadas durante todo o processo.

3.1. Tecnologias

Inicialmente foi realizada uma pesquisa para a escolha de bibliotecas de áudio, o critério de escolha foi utilizar bibliotecas que utilizam as linguagens de programação com as quais os desenvolvedores já estão familiarizados. Há três bibliotecas dentre as mais promissoras, sendo elas a RTAudio, Portaudio e OpenAL Soft, para as linguagens C e C++.

Ao serem analisados os exemplos disponíveis nos repositórios das bibliotecas, juntamente com suas respectivas documentações, a OpenAL Soft apresentou uma grande complexidade para a sua compreensão, pois sua documentação não é clara e as funções disponíveis se mostraram limitadas. Em relação à RTAudio e a Portaudio, foram encontradas funcionalidades muito parecidas, de modo que ambas são boas candidatas a serem utilizadas para o desenvolvimento. Contudo, ao avançar nos estudos dessas bibliotecas foi identificada a necessidade da implementação de sintetizadores de instrumentos, que foge do escopo proposto por elas. Depois de um novo processo de pesquisa, foi

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Figura 2: Modelo de Thayer para classificação de emoções



Fonte: Adaptado de [Yang et al., 2008]

encontrada a biblioteca STK (Synthesis ToolKit), uma extensão da RTAudio implementada em C++ que contém diversos sintetizadores.

O STK cobre grande parte dos requerimentos deste trabalho, porém não foram encontrados recursos para trabalhar com a escrita e leituras de arquivos MIDI. Para suprir tal necessidade, foi selecionada uma ferramenta focada neste caso de uso, a Midifile.

Além das ferramentas de áudio, também foram necessárias outras bibliotecas, já que parte do projeto interage com a internet, tópico que será discutido na implementação. Tais bibliotecas são a cURLpp e a nlohmann/json.

Para a categorização emocional das músicas, foi utilizada a web API da plataforma de streaming Spotify. Foram consideradas a popularidade da plataforma, que conta com mais de 422 milhões de usuários ativos mensais [Spotify, 2022], e o seu sistema de recomendação (sistema responsável por indicar ao usuário músicas, álbuns ou *playlists*). O sucesso desse sistema se origina do robusto algoritmo de análise das músicas disponíveis no serviço, analisando e extraindo diversas características, chamadas de *audio features*. Os dois atributos das *audio features* de interesse são: "Energy" e "Valence". O campo de energia ou *Energy* é a intensidade da música, calculada pelos ruídos, pela velocidade e também pelo volume da música. Já o campo de valência ou *Valence* é a positividade, classificando então o quão feliz a música é.

As tecnologias listadas foram utilizadas para criar um programa dividido em três componentes principais.

1. Agregador: Comunicação com a Web API do Spotify, a fim de obter dados a respeito das

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

características de cada uma das músicas, dada uma amostra predefinida. Destas características são observados os valores de valência e energia. Estes valores foram aplicados no modelo de Thayer representado pela Figura 2 para classificar qual emoção melhor se encaixa com tal música.

2. Analisador: Análise de arquivos MIDI das músicas pertencentes à amostra, a fim de agregar valores como a ocorrência de cada uma das transições entre notas (ex: ocorrências de um dó seguido de um ré) e o tempo entre elas.

3. Gerador: Geração de uma composição musical em tempo real para uma determinada emoção a partir de cadeias geradas com os dados obtidos pelo analisador. Com o uso das Cadeias são construídas instruções no protocolo MIDI considerando escala, tom e tempo; As instruções são rocessadas e reproduzidas com um dos sintetizadores disponíveis no STK.

3.2. Metodologia

Este item irá apresentar e descrever o percurso metodóligo deste trabalho, ou seja, como o trabalho foi desenvolvido. O trabalho foi desenvolvido em cinco partes:

1. Pesquisa, seleção e testagem das bibliotecas

   Foi feita uma pesquisa para a escolha das bibliotecas de áudio a serem utilizadas no desenvolvimento do trabalho. O critério de escolha das bibliotecas foram as bibliotecas cuja linguagem de programação eram conhecidas pelos pesquisadores. Foram selecionadas para consideração as bibliotecas PortAudio, OpenAL, RTAudio e STK. Foram desenvolvidos uma série de testagens de software com as bibliotecas selecionadas para este estudo, para determinar qual delas é a mais adequada para cumprir com os objetivos deste trabalho, sendo que foi determinado que o STK seria a opção mais adequada.

2. Análise da performance o Circumplexo de Tellegen-Watson-Clark

   Com a biblioteca e linguagem de programação definidas, foi agregada a primeira amostra de músicas para análise de acordo com a categorização emocional no Circumplexo de Tellegen-Watson-Clark. Foram obtidos os arquivos MIDI de 45 músicas do gênero clássico. Essa fase de análise foi feita somente com o uso da API do Spotify, com o qual foi gerada uma lista com as classificações para cada uma das músicas. O intuito deste passo foi de verificar se tal modelo de categorização seria suficiente para atingir os resultados esperados.

3. Pivotamento para o Modelo de Thayer

   Ao observar os resultados da análise no Complexo de Tellegen-Watson-Clark, foi decidido que seria interessante o uso de um outro esquema de classificação. Após uma nova pesquisa, foi feito o pivotamento para a análise de emoções com o Modelo de Thayer, utilizando uma nova amostra com 80 músicas.

4. Produção das Cadeias de Markov

   Foi criado um algoritmo responsável por analisar os arquivos MIDI das músicas selecionadas, de onde foram extraídas as informações utilizadas no cálculo das matrizes de transição do algoritmo de geração procedural. Com as informações obtidas foram construídas as Cadeias de Markov das notas.

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

5. Análise das composições geradas por meio de um formulário (Análise 1)

   Finalmente, foi realizado um processo de coleta da percepção de pessoas alheias ao trabalho em relação às composições geradas com o algoritmo desenvolvido por meio de um questionário, não a fim de obter dados estatisticamente válidos, mas para observar como a opinião dos participantes se compara com os resultados esperados.

6. Implementação de novos parâmetros

   Depois de analisar os resultados obtidos no formulário, observou-se que precisaria ser feita uma melhora no algoritmo. Observando a API do Spotify foram encontrados atributos de *"Key"* e *"Mode"*, que seriam implementados na geração.

7. Análise das composições geradas por meio de um formulário (Análise 2)

   Para testar as novas gerações com tom e escala, foi feita uma nova rodada de coleta de dados através de formulário. O formulário tem as mesmas características do primeiro, dessa forma tentando apenas diferenciar as gerações e as respostas.

3.3. Implementação

Com base nos estudos das bibliotecas selecionadas foi montado um projeto conectando as principais funcionalidades para realizar os primeiros testes. Sendo a biblioteca principal o STK, foi necessário configurar a maneira em que o dispositivo de áudio é selecionado para executar os sons. Em uma de suas classes é disponibilizada uma função que lista os dispositivos disponíveis, possibilitando a obtenção do identificador do dispositivo desejado.

Para testar essa implementação foi montada uma matriz de transição com as notas da escala em dó maior, onde a probabilidade de cada nota é a mesma. Como a linguagem traz consigo uma função de randomização, o processo de gerar a cadeia aleatória foi facilitado. Por sequência foi testado o STK com a execução das notas em tempo real no dispositivo desejado.

Durante as pesquisas houve uma dúvida em relação a como poderiam ser definidos os valores de probabilidade para cada emoção, implicando na definição da matriz. Encontrar a API do Spotify contribuiu na resolução dessa questão, pois para obter esses valores do zero, seria necessária uma análise em partituras para cada música, contabilizando a quantidade das transições entre cada uma das notas. Portanto houve uma redução no tempo desta etapa.

Para realizar análise da probabilidade de cada transição dentro de uma música, foi decidido utilizar um banco de arquivos MIDI com 45 músicas clássicas. A combinação da API com os arquivos MIDI foi feita através do modelo da Figura 1, para isso houve uma tentativa de transpor os valores de energia e valência em coordenadas do modelo correlacionando a energia com o eixo do "Estímulo" e valência com o eixo "Agradável".

Na primeira tentativa de gerar o conjunto de dados os valores vieram próximos de duas emoções: "Triste" e "Monótono", pois a API retornou valores parecidos em seus atributos, baixos tanto para "Energia" como a "Valência". O motivo da obtenção de valores baixos é as músicas serem do gênero de músicas clássicas, pois este é o gênero utilizado como próprio exemplo do que seria uma música com baixa "Energia", como escrito na documentação da API. Para atingir o objetivo deste trabalho, é necessário que as amostra coletada seja diversa para que sejam obtidos dados para todas as emoções que poderão ser geradas. Como foram extraídas apenas duas emoções nesta etapa, estes resultados não foram suficientes para prosseguir com o trabalho.

Com isso houve uma mudança de modelo, do Circumplexo de Tellegen-Watson-Clark para o Modelo de Thayer representado na Figura 2. Esse modelo categoriza as emoções por meio de um

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

plano, o que facilita na divisão de novas músicas. Em conjunção dessa mudança, foi feita uma outra busca para encontrar novas músicas, além de uma redução no escopo deste trabalho, descartando a produção de matrizes para as durações das notas e acordes, por conta do tempo utilizado no período de pesquisa sobre o Modelo de Thayer.

A busca desse novo banco de arquivos MIDIs foi separada em estilos musicais, onde foram selecionadas 10 músicas para cada estilo. Os estilos são: *Classical*, *Country*, *Eletronic*, *Hip Hop*, *Jazz*, *Metal*, *Pop* e *Reggae*, os arquivo foram baixados do *Musecore*, um site que disponiliza um grande acervo de partituras. No total foram separados 80 arquivos para serem analisados.

Para aplicar uma música em uma emoção foi feita a divisão considerando o modelo de *Thayer*, o valor de valência e energia relacionado com os eixos do plano. Dessa forma os resultados tiveram uma classificação melhor.

Com todas as músicas classificadas, foi gerado uma matriz de transição para cada música. Para analisar as notas de cada arquivo, foi desenvolvido um algoritmo cujo intuito é verificar um conjunto de notas (nota atual e a nota anterior) e montar uma matriz com esses resultados. Como os arquivos MIDIs poderiam ter mais de um instrumento foi montado uma matriz para cada instrumento, e ao final todas as matrizes foram adicionadas em uma matriz principal. Sendo assim, cada música tem uma matriz de notas. A biblioteca Midifile permitia o acesso ao BPM da música, então foi montada uma lista com os BPMs de cada música de acordo com os valores obtidos no MIDI e o compasso ficou fixo para permitir melhor trabalho com os BPMs.

Como segundo teste foi montado uma interface de entrada via linha de comando para definir qual emoção deveria ser usada pelo gerador, e, por consequência, implementar um sistema de *threads*, onde uma *thread* gera a música e a outra espera uma entrada de usuário.

De acordo com os resultados dessa fase de testagem, mostrou-se necessária uma melhoria no gerador, especialmente para as músicas com alvo de emoção em relaxamento. Como neste momento foram implementados apenas os parâmetros de nota e tempo, foram observados novamente os dados obtidos pelo agregados. Como essa foi a primeira etapa de aperfeiçoamento do gerador, foram escolhidos dois parâmetros que não exigiriam um grande tempo para serem implementados, os atributos de *"Key"* e *"Mode"*. Respectivamente, estes valores representam em qual tom e escala(maior ou menor) a música está. Essas mudanças foram refletidas no analisador com a agregação de todos os possíveis tons em um conjunto e o cálculo da moda das escalas encontradas nas amostras para cada emoção. No gerador esses novos parâmetros foram utilizados para definir quais os valores que as notas podem assumir.

Implementadas tais melhorias, foi produzida mais uma testagem no mesmo padrão da anterior, para obter novos resultados e avaliar os efeitos dos novos parâmetros.

3.4. Testes e Resultados

Durante a execução do trabalho foram feitas 4 fases de testes. As duas primeiras são testes internos para verificar a aplicação dos modelos nas músicas escolhidas.

O primeiro teste foi com o Circumplexo, foi feita uma requisição na API do Spotify com a lista de 45 músicas clássicas e o retorno foi aplicado no modelo. Nessa aplicação do modelo foram utilizados valores de valência e energia de cada música, sendo que tais valores foram transpostos respectivamente para os eixos de “Agradável” e “Estímulo”.

O segundo teste foi feito com o Modelo de Thayer, para esse teste foram selecionadas 80 músicas de 8 categorias diferentes. Seguindo o mesmo princípio do primeiro teste, foi feita uma requisição na API e o retorno, também, da valência e da energia aplicadas nos quadrantes do modelo.

Como apresentado anteriormente na implementação, foi feito um questionário em formato de formulário com o intuito de avaliar as gerações do algoritmo. Cada geração tem uma emoção

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

definida pela interface.

Tabela 1: Tabela de Quantidade de Gerações por Emoção

Emoções	Quantidade de Gerações
Felicidade	2
Tristeza	2
Raiva	3
Relaxamento	3

Fonte: Autores

As gerações foram colocadas em ordem aleatória, para evitar algum nível de viés na apresentação do formulário. Todas as pessoas receberam as perguntas na mesma ordem predefinida. O questionário foi enviado através de mídias sociais para diversas pessoas, então não foi selecionado um grupo específico. Os respondentes tiveram em torno de 5 dias para responder, contudo a maioria das respostas foram entregues nos 3 primeiros dias totalizando 31 respostas.

Nesse momento e para o desenvolvimetno do trabalho foi feita uma opção por uma pesquisa exploratória, e, nesse sentido não foi feito um trabalho para a determinação do tamanho da amostra para a análise dos dados obtidos.

Tabela 2: Tabela de Porcentagem das Respostas por Pergunta do Formulário 1

Música	Felicidade	Tristeza	Raiva	Relaxamento	Opção correta
1	3,3%	**35,5%**	29%	32,2%	Tristeza
2	**32,3%**	25,8%	**32,3%**	9,7%	Felicidade
3	25,8%	16,1%	**48,4%**	9,7%	Raiva
4	**41,9%**	12,9%	**41,9%**	3,3%	Raiva
5	29%	19,4%	**35,5%**	16,1%	Raiva
6	29%	22,6%	**45,2%**	3,2%	Felicidade
7	0%	**67,7%**	12,9%	19,4%	Tristeza
8	0%	19,4%	**41,9%**	19,4%	Relaxamento
9	**41,9%**	16,1%	25,5%	16,5%	Relaxamento
10	**41,9%**	19,4%	35,5%	3,2%	Relaxamento

Fonte: Autores

O quarto teste também foi apresentado em um formato de formulário, dessa forma podendo analisar como as alterações no algoritmo de tom e escala interferiram nos resultados e validando as novas composições.

Para esse teste foram geradas novamente 10 músicas seguindo o mesmo padrão da quantidade de cada emoção, na tentativa de deixar o teste o mais próximo do anterior. Dessa vez tiveram apenas 22 respostas com tempo de recolha das respostas de 6 dias.

3.5. Análise de Resultados

Nas duas primeiras fases de testes envolvendo o Circumplexo e o modelo de Thayer, foram feitas análises internas. O primeiro teste foi feito com o Circumplexo e nele observou-se que as músicas tinham uma tendência para as emoções de tristeza e monótono, com isso foram levantadas questões em relação ao próprio modelo e a quantidade de músicas utilizadas sendo todas do mesmo gênero musical. Os resultados obtidos não foram suficientes.

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Tabela 3: Tabela de Porcentagem das Respostas por Pergunta do Formulário 2

Música	Felicidade	Tristeza	Raiva	Relaxamento	Opção correta
1	22,7%	18,2%	0%	**59,1%**	Felicidade
2	4,5%	**40,9%**	**40,9%**	13,6%	Tristeza
3	**59,1%**	0%	27,3%	13,6%	Raiva
4	4,5%	31,8%	**45,5%**	18,2%	Raiva
5	27,3%	31,8%	**36,4%**	4,5%	Raiva
6	**90,9%**	4,5%	4,5%	0%	Relaxamento
7	36,4%	4,5%	0%	**59,1%**	Relaxamento
8	0%	**63,6%**	9,1%	27,3%	Felicidade
9	0%	**54,5%**	31,8%	13,6%	Tristeza
10	0%	27,3%	4,5%	**68,2%**	Relaxamento

Fonte: Autores

No segundo teste com o modelo de Thayer foi preparado uma base com 80 músicas de gêneros diferentes para uma tentativa de abranger mais valores dentro dos eixos do modelo com os atributos energia e valencia. 80 músicas, sendo 10 músicas para cada gênero: (*Classical*, *Country*, *Eletronic*, *Hip Hop*, *Jazz*, *Metal*, *Pop* e *Reggae*). Os resultados obtidos atingiram as expectativas esperadas, onde as definições das emoções para cada música foi mais variado.

Os resultados obtidos no terceiro teste estão ilustrados na Tabela 2. As músicas 1 e 7 representam músicas de tristeza. Observa-se nelas a baixa presença de felicidade. Sendo assim, as músicas de tristeza tem uma definição clara. No caso da música 1 em especial, as percepções dos respondentes indicavam as emoções de relaxamento e tristeza.

Nas músicas 2 e 6 geradas para a emoção de felicidade, a escolha condizente com o algoritmo ficou em segundo lugar, logo após à raiva. Verificando o posicionamento das emoções no modelo de Thayer, raiva e felicidade encontram-se ambos com um alto valor de energia, se diferenciando pelo valor de valência.

Já as músicas categorizadas como raiva são: 3, 4, e 5, tendo a raiva como a mais selecionada e com os valores de relaxamento inferiores a 20%. Comparando com as músicas de tristeza, o relaxamento também é baixo. No modelo de Thayer, a variação entre tristeza e raiva se dá pelo valor da energia.

Pelos resultados observados, as músicas de relaxamento, posicionadas no formulário nas músicas 8, 9 e 10, apresentaram opiniões conflitantes. Os votos para relaxamento são baixos, e houve uma variação nas emoções mais votadas. No caso da música 8, os votos ficaram entre raiva e tristeza. Já nas músicas 9 e 10 as opções mais escolhidas foram raiva e felicidade. Pelo modelo de Thayer, a escolha recorrente de felicidade se mostra fundamentada, pois difere do relaxamento somente pelo valor de energia. O fato da música 8 ter um alto índice de raiva, indica uma incerteza nos resultados, pois a raiva representa o completo oposto da emoção de relaxamento pelo modelo. Um fator que pode ter contribuído para tais resultados é a falta de refinamento no algoritmo de escolha de tempo, já que as três músicas de relaxamento geradas possuíam um tempo acima de 100 BPM. O intervalo de tempos possíveis para a emoção de relaxamento no modelo gerado é muito grande, variando de 70 a 160 BPM, o que pode justificar a grande variação de emoções percebidas.

Os resultados do quarto teste constam na Tabela 3 com as porcentagens das respostas.

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Nesse teste tivemos 50% das músicas com a opção correta sendo a emoção mais votada. As músicas 1 e 3 tem a opção correta sendo a segunda mais votada, essas duas podem ter causado dúvida por conta da duração da nota ser padrão para todas as emoções.

A música 2 foi provavelmente um caso particular com o instrumento, pois observando a geração foi possível analisar que o instrumento pode ter influenciado no empate entre a opção correta é a segunda opção, Raiva e Tristeza, já que a diferença entre as duas é a quantidade de energia.

As músicas 6 e 8 não tiveram nenhum voto da opção correta, não foi encontrado um motivo específico para elas já que possam estar destoando em relação ao instrumento, BPM, tempo de nota ou a escala.

Comparando os resultados dos formulários, observa-se que a escolha das emoções corretas sobe de 40% para 50%, e a taxa de aceitação de 31,93% para 35,45%, possivelmente indicando um efeito positivo nas melhorias implementadas entre os testes. No trabalho de [Ramanto e Maulidevi, 2017] a taxa de aceitação foi de 62,5%, contudo o método dos testes foram diferentes, portanto não há possibilidade de uma comparação direta entre os resultados.

4. Considerações Finais

Nesse capítulo serão discutidos as conclusões após as análises dos resultados obtidos pelos testes realizados.

Este trabalho se propôs a investigar a viabilidade de produzir uma composição musical usando técnicas de geração procedural com Cadeias de Markov a partir de uma emoção predeterminada. Para esta investigação, foi selecionada uma amostra de músicas de diferentes gêneros, a qual serviu de base para o algoritmo de cálculo das matrizes. Os dados obtidos foram tratados de duas formas, por meio de uma comunicação com um API e por arquivos MIDI. No caso da comunicação com o API, foi feita a comunicação com o API do serviço de *streaming* de músicas, o Spotify, para obter diversas informações das músicas, entre as de interesse para este trabalho a energia e valência. E, no caso dos arquivos MIDI obtidos do site MuseScore, estes foram analisados por um programa autoral para extrair informações dos elementos das composições musicais, sendo elas as transições entre notas e o tempo das músicas.

Com a obtenção de todos esses dados, as músicas selecionadas foram categorizadas de acordo com o modelo de Thayer em quatro emoções: Felicidade, Tristeza, Raiva e Relaxamento. As matrizes de transição de notas juntamente com o parâmetro "tempo" de cada uma das músicas foram utilizados para calcular as matrizes de transição correspondentes a cada uma das emoções, de acordo com a categorização emocional das músicas.

Após análise dos resultados, foi verificado que o algoritmo teve uma taxa de aceitação de 50% entre as coletas de dados, mesmo a amostra seja baixa ainda é possível analisar os dados e definir as próximas otimização para concluir o objetivo do trabalho com a taxa de aceitação mais próxima do 100%. Um fator determinante é que a pesquisa foi realizada sem nenhuma regra de seleção dos participantes, então os participantes são de áreas distintas.

Com a implementação do tom e da escala já foi observado que a base das gerações estão mais precisas, conclui-se que alterar o instrumento e implementar a variação de tempo de nota pode trazer o resultado esperado.

Referências

Allen, A. O. (1990). *Probability, Statistics, and Queueing Theory with Computer Science Applications*. ACADEMIC PRESS, INC.

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856

Cezar, A. T. e Jucá-Vasconcelos, H. P. (2016). Diferenciando sensações, sentimentos e emoções: uma articulação com a abordagem gestáltica. URL http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1807-25262016000100002.

do Nascimento, A. M. A. (2021). Aplicação de cadeias de markov em música. URL https://repositorio.ufrn.br/jspui/bitstream/123456789/34004/1/Aplica%C3%A7%C3%A3oDeCadeias_Nascimento_2021.pdf.

Linskens, E. J. (2014). Music improvisation using markov chains. URL https://dke.maastrichtuniversity.nl/gm.schoenmakers/wp-content/uploads/2015/09/Linskens-Final-Draft.pdf.

Lu, L., Liu, D., e Zhang, H.-J. (2006). Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):5–18.

Lyra, C. (1999). *Harmonia prática da bossa-nova: método para violão*. Irmãos Vitale.

Med, B. (1996). *Teoria da Música*. MusiMed.

Nierhaus, G. (2008). *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer.

Olson, H. F. (1967). *Music, Physics and Engineering*. Courier Corporation.

Ramanto, A. S. e Maulidevi, N. U. (2017). Markov chain based procedural music generator with user chosen mood compatibility. URL https://www.jstage.jst.go.jp/article/adada/21/1/21_19/_pdf/-char/ja.

Simões, A. R. C. (2012). As emoções ao compasso da música : um olhar sobre a influência da música na resposta emocional. URL http://hdl.handle.net/10451/8076.

Spotify (2022). Shareholder deck q1. URL https://s29.q4cdn.com/175625835/files/doc_financials/2022/q1/Shareholder-Letter-Q1-2022_FINAL.pdf.

Thayer, R. E. (1989). The biopsychology of mood and arousal. URL https://books.google.com.br/books?id=ORiwiDNqcbEC&dq=The+Biopsychology+of+Mood+and+Arousal&lr=&hl=pt-BR&source=gbs_navlinks_s.

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., e Nater, U. M. (2013). Emotion regulation through listening to music in everyday situations. URL http://dx.doi.org/10.1080/02699931.2011.595390.

Togelius, J., Shaker, N., e Nelson, M. J. (2016). Introduction. In Shaker, N., Togelius, J., e Nelson, M. J., editors, *Procedural Content Generation in Games: A Textbook and an Overview of Current Research*, p. 1–15. Springer.

Watson, D. e Tellegen, A. (1985). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence.

Yang, Y.-H., Lin, Y.-C., Su, Y.-F., e Chen, H. H. (2008). A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):448–457.

DOI: 10.59254/sbpo-2023-174856
http://live-proceedings.pantheonsite.io/p/174856