

CAMPOREZ, Higor Araújo Fim; FRIZERA NETO, Anselmo; COSTALONGA, Leandro Lesqueves; ROCHA, Helder Roberto de Oliveira. *Interface Computacional para Controle Musical Utilizando os Movimentos dos Olhos*. Revista Vórtex, Curitiba, v.6, n.2, 2018, p.1-17

Interface Computacional para Controle Musical Utilizando os Movimentos dos Olhos¹

Higor Araújo Fim Camporez² | Anselmo Frizera Neto³
Leandro Lesqueves Costalonga⁴ | Helder Roberto de Oliveira
Rocha⁵

Universidade Federal do Espírito Santo | Brasil

Resumo: O último censo demográfico do Brasil relata que cerca de 13,1 milhões de pessoas se declaram com alguma deficiência do tipo motora. O avanço tecnológico tem contribuído na promoção da acessibilidade e inclusão por meio das técnicas de tecnologia assistiva. Este trabalho apresenta uma

¹ *Computational interface for musical interaction using eye movements*. Submetido em: 15/08/2018. Aprovado em: 30/09/2018.

² É mestrando em engenharia elétrica, na área de telecomunicações e tecnologia da informação, pela Universidade Federal do Espírito Santo (UFES). É graduado em Engenharia da Computação pela UFES - CEUNES. É membro do grupo de pesquisa NESCoM (Núcleo Espírito-santense de Computação Musical), auxiliando no desenvolvimento de software e hardware. Participou de um projeto de iniciação científica sobre Smartgrid, financiado pela FAPES. Foi membro da empresa júnior Adapti com cargo de assessor de desenvolvimento. E-mail: higorcamporez@gmail.com

³ Graduado em Engenharia Elétrica pela Universidade Federal do Espírito Santo (2006). Doutor em Eletrônica pela Universidad de Alcalá (2010). Foi pesquisador do Grupo de Bioengenharia do Consejo Superior de Investigaciones Científicas entre 2006 e 2010. Atualmente, é Professor Adjunto do Departamento de Engenharia Elétrica e Docente Permanente do Programa de Pós-Graduação em Engenharia Elétrica, ambos da Universidade Federal do Espírito Santo - Brasil. Atua como membro da Junta Diretiva da Asociación Iberoamericana de Tecnologías de Apoyo a la Discapacidad, contribuindo no apoio à difusão do conhecimento na área de tecnologias de apoio à deficiência. Possui experiência em Eletrônica e Engenharia Biomédica, atuando principalmente nos seguintes temas: robótica de reabilitação, desenvolvimento de sensores e interfaces homem-máquina, análise de marcha (gait analysis), processamento de sinais biomédicos, tecnologias de apoio à deficiência e ajuda à mobilidade. E-mail: frizera@iceee.org

⁴ É professor do Departamento de Computação e Eletrônica da Universidade Federal do Espírito Santo, onde também coordena o grupo de pesquisa denominado NESCoM - Núcleo Espírito-santense de Computação Musical. Doutor em Computação Musical pela University of Plymouth (UK), atua na pesquisa interdisciplinar nas áreas de: multimídia, interação homem-computador, linguagens de programação, inteligência artificial, computação ubíqua, e computação musical, sendo a última a principal linha de pesquisa. Atualmente, além do NESCoM, está ligado aos grupos de pesquisa de Computação Musical (UFRGS), Música Ubíqua (NAP/UFAC) e Computação Aplicada (UFES). E-mail: leandro.costalonga@ufes.br

⁵ Possui Doutorado e Mestrado em Computação Científica e Sistemas de Potência - UFF, Especialização em Administração e Negócios - UCAM, Bacharelado em Administração - UFRRJ e Graduação em Engenharia Elétrica - UFF. Tem experiência na área de Engenharia Elétrica, atuando principalmente nos seguintes temas: Estimação de Estado, Metaheurísticas, Sistema de Potência e Telecomunicações. Foi Tutor Presencial no curso de Tecnologia de Sistemas de Computação oferecido pelo CEDERJ, professor no Instituto Federal do Espírito Santo e do Departamento de Computação e Eletrônica do CEUNES/UFES. Realizou pesquisa no âmbito do programa Pós-Doutorado da CAPES na UFES. Ocupa atualmente o cargo de Professor Classe C - Adjunto II do Departamento de Engenharia Elétrica da UFES. E-mail: helder.rocha@ufes.br

interface computacional para controle musical, porém com foco em promover acessibilidade e inclusão a inúmeras pessoas. Por meio do movimento dos olhos o usuário pode controlar uma interface gráfica que envia mensagens (MIDI) para o software Ableton Live com objetivo de controlar a execução sonora. Visando o uso por pessoas leigas no campo musical, a interface proporciona indicações de escolhas para direcionar o usuário a gerar sons mais conexos. As indicações podem ser configuradas de forma manual ou automatizada com auxílio do algoritmo de agrupamento K-means.

Palavras-chave: Rastreamento de Ocular, Acessibilidade, Controle Musical, Clusterização.

Abstract: The latest Brazilian demographic census reports that about 13.1 million people have some sort of motor disability. Technological advances and Assistive Technology have been developed to promote accessibility and inclusion for this population. This paper proposes an eye-tracking computational interface designed for motor impaired user that allows them to perform live music by triggering MIDI messages that, in turn, selects previous prepared sound samples at Ableton Live. To allow use by non-musicians, the interface indicates possible choices in order to direct the user to generate more related sounds. Furthermore, an algorithmic sample selection based on K-means and clustering is also proposed to show next choices.

Keywords: Eye Tracking, Accessibility, Musical Control, Clustering.

* * *

As pesquisas, realizadas pela Organização Mundial da Saúde, demonstram uma estimativa de que, no mundo, mais de 1 bilhão de pessoas vivem com algum tipo de deficiência (WHO, 2012). De acordo com o último censo demográfico do Brasil (IBGE, 2010), 45,6 milhões de pessoas, cerca de 23,9% da população em 2010, declararam ter pelo menos um tipo de deficiência, seja ela visual, auditiva, motora ou mental/intelectual. O estudo do IBGE classifica deficiência visual, auditiva e motora em três modos: alguma dificuldade, grande dificuldade e não consegue de modo algum (incapacidade). Assim, as pessoas que se declararam com deficiência motora estão divididas em 8,8 milhões com alguma dificuldade, 3,6 milhões com grande dificuldade e 0,7 milhões com incapacidade.

Independentemente do alto número de pessoas com deficiência, garantir a inclusão é algo de extrema importância para proporcionar o bem-estar desses indivíduos. Em janeiro de 2016 entrou em vigor a Lei Brasileira de Inclusão (LBI), que afirmou capacidade e autonomia para esses cidadãos

exercerem seus direitos sociais em igualdade com os demais (BRASIL, 2015). Nesse contexto, a Tecnologia Assistiva (TA) – uma adaptação, ou desenvolvimento específico, de produtos e tecnologias para auxiliar em problemas funcionais de pessoas com deficiência (WHO, 2001) – por meio da acessibilidade, pode proporcionar a inclusão dos indivíduos.

A técnica de rastreamento ocular (*eye tracking*) é utilizada para determinar onde uma pessoa está olhando, o método acompanha os movimentos dos olhos do usuário e observa como as pupilas estão se comportando (POOLE; BALL, 2005), esse método é muito aplicado para interação homem-computador, inclusive no controle de interfaces musicais. Um trabalho com objetivo de proporcionar, a pessoas com deficiência, o uso do computador por meio da técnica de *eye tracking* é descrito por Lupu et al. (2017). O foco do trabalho previamente mencionado está em promover acesso à Internet, podendo-se, com movimento dos olhos, controlar o mouse e ativar funções de um navegador adaptado para obter uma melhor experiência de uso. Outro trabalho que mostra o uso do rastreamento para auxiliar na comunicação de pessoas paráliticas é demonstrado por Alva et al. (2017), onde imagens com atividades diárias, por exemplo comer, são mostradas em uma tela e o indivíduo pode expressar sua necessidade apenas focando sua visão na imagem.

Louro, Ikuta e Nascimento (2005) demonstram algumas adaptações e compensações em instrumentos musicais tradicionais para proporcionar a execução por indivíduos portadores de deficiências, por exemplo, um teclado musical com teclas maiores para que usuário possa usar os punhos quando há dificuldades em usar os dedos. Nesse contexto, o uso da tecnologia assistiva e do computador pode auxiliar nas técnicas adaptativas.

Um trabalho unindo acessibilidade, inclusão, música e técnicas computacionais é descrito por Corrêa et al. (2008). Os autores demonstram um jogo musical chamado GenVirtual para reabilitação de indivíduos com necessidades especiais, no qual é possível seguir uma sequência de cores e sons emitidos a partir de objetos virtuais projetados no mundo real. O projeto faz uso de um computador, uma câmera e algumas cartões. Por meio da realidade aumentada, o computador reconhece os cartões e cria objetos no mundo virtual. O indivíduo pode fazer a obstrução do cartão com a mão, ou até mesmo o pé, dependendo da sua deficiência física. Através da imagem da câmera é identificado o cartão obstruído e executa-se um som relativo ao mesmo. Vale ressaltar que o musicoterapeuta pode fazer todo o planejamento de sequências e sons do sistema, além da posição dos cartões para explorar o movimento motor do paciente.

O conceito de música ubíqua, também conhecido como UbiMus (*Ubiquitous Music*), pode ser definido como o resultado da conjunção de sistemas musicais, sejam eles fixos ou móveis, voltados a dispositivos tecnológicos para geração de sons (KELLER et al., 2009). O trabalho, anteriormente citado, descreve que o objetivo é permitir que qualquer pessoa, até mesmo as leigas, possa utilizar

sistemas musicais de forma criativa, com ênfase em qualidade sonora e na naturalidade das interfaces. Exemplos de UbiMus utilizando dispositivos móveis do cotidiano, como *smarphones* e *tablets*, para gerar sons são apresentados por Keller et al. (2011), Lazzarini et al. (2014a) e Pimenta et al. (2014). Outra afirmação descrita por Fornari (2014) relata que as interfaces gestuais são ferramentas essenciais nas performances de UbiMus. Com base nas afirmativas, utilizar dispositivos tecnológicos para captação dos movimentos dos olhos, juntamente com o mapeamento da informação para gerar sons, se encaixa no contexto de música ubíqua. Assim, a técnica de *eye tracking* pode servir de interface para trabalhos de UbiMus, como por exemplo no controle de instrumentos musicais robóticos, apresentados por Camporez et al. (2016), e também no comando de interfaces computacionais apresentadas por Lazzarini et al. (2014b).

O presente artigo descreve uma interface, denominada Olhar Musical, para controle musical por intermédio do movimento dos olhos, com objetivo de promover acessibilidade e inclusão de pessoas com deficiência motora, além da possibilidade de proporcionar performances de UbiMus. Vale ressaltar que atualmente a interface encontra-se em fase de desenvolvimento e experimentação. Na Seção 1 são apresentados alguns trabalhos que envolvem música e movimento dos olhos. A Seção 2 apresenta a descrição do desenvolvimento da interface computacional do trabalho, além de técnicas para classificação de amostras de áudio. Por fim, na Seção 3 são descritos as conclusões e os futuros trabalhos.

1. TRABALHOS RELACIONADOS COM USO DE *EYE TRACKING*

O sistema EyeMusic (HORNOF; SATO, 2004) utiliza um *eye tracker* comercial, onde as coordenadas, referentes ao ponto de foco do usuário, são enviadas para a linguagem de programação MAX/MSP para geração de sons. O usuário controla o cursor, representado por um ponto branco, que mostra o local em que está focando. Com as coordenadas x e y do cursor o sistema cria um sintetizador granular de acordo com o ponto.

O Oculog é um sistema de baixo custo utilizando a técnica de *eye tracking* (KIM; SCHIEMER; NARUSHIMA, 2007). Os dados do rastreamento ocular do usuário são incorporados à linguagem de programação sônica Pure Data para interação das sequências. O campo de visão do sistema é dividido em quatro quadrantes, onde cada quadrante é um gerador de tom em tempo real e cada gerador é controlado por uma sequência cíclica. As coordenadas horizontais e verticais geradas pela movimentação dos olhos são mapeadas em tom (*pitch*) e intensidade (*velocity*), respectivamente. Onde para o eixo horizontal, olhando para a extremidade mais à esquerda produz o *pitch* com valor 0 e mais à direita com o valor 127. Para o eixo vertical, olhando para baixo produz a *velocity* com o valor 0 e para

cima com o valor 127.

Também utilizando *eye tracking*, tem-se o EyeHarp (VAMVAKOUSIS; RAMIREZ, 2012), um projeto com código aberto de interface musical controlada pelo movimento dos olhos. O projeto é dividido em três camadas básicas: (a) *the pEYE menu*, (b) *the Step Sequencer* e (c) *the Arpeggiator*. As interfaces (b) e (c) são utilizadas para construção rítmica e harmônica em segundo plano, já a (a) é usada na construção melódica. A interface *pEYE menu* apresenta grandes botões posicionados de modo a formar um círculo, no qual representam notas em uma escala diatônica. O tamanho dos botões tem a finalidade de evitar possíveis interferências devido a imprecisão do sistema de *eye tracking*. Algumas características foram implementadas para aprimorar a usabilidade, por exemplo, um cursor identificando o ponto de foco em tempo real e os pequenos pontos brancos que ajudam o usuário a concentrar sua visão em um local exato, além de também ajudar identificar a existência de diferentes tarefas.

Apesar do EyeHarp não ser projetado especificamente para pessoas com deficiências, os autores demonstram alguns testes para reabilitação de pessoas portadoras de paralisia cerebral e aponta indícios de que o trabalho pode, não apenas aumentar a qualidade de vida desses indivíduos, como também ajudar no processo de reabilitação.

Em relação aos trabalhos supracitados, a interface em questão visa avançar em estudos no controle, também por meio do movimento dos olhos, de softwares para produção musical já existente no mercado. Esses softwares possuem várias ferramentas homologadas e testadas, além de pessoas capacitadas para seu uso, o que, de certa forma, pode proporcionar a exclusão da necessidade de aprendizado sobre alguns aspectos de configuração da execução sonora utilizado na interface.

2. A INTERFACE COMPUTACIONAL OLHAR MUSICAL

Este projeto é implementado utilizando a técnica de *eye tracking*, pois a mesma tem se mostrado eficaz na inclusão de pessoas com deficiências e também na construção de interfaces para o campo musical. A Figura 1 ilustra a descrição da interface em linhas gerais, onde as setas indicam o fluxo de dados. O projeto se divide em três peças chaves: o dispositivo (*Eye Tracker*) comercial para rastreamento ocular, Tobii EyeX⁶; a interface gráfica, desenvolvida com a plataforma para desenvolvimento de jogos Unity⁷ e o software para produção musical Ableton Live⁸.

O dispositivo de rastreamento de íris envia as coordenadas do ponto em que o usuário está

⁶ TOBII AB. Tobii EyeX. Disponível em: <<https://tobiigaming.com/product/tobii-eyex/>>. Acesso em: 12 jun. 2018.

⁷ UNITY. Unity - Game Engine. Disponível em: <www.unity3d.com>. Acesso em: 12 jun. 2018.

⁸ ABLETON. Ableton Live 10 - Ableton. Disponível em: <<https://www.ableton.com/en/live/>>. Acesso em: 12 jun. 2018.

focando, na tela do computador, para interface computacional, que verifica se o ponto condiz com algum elemento de ação da cena, e, se for o caso, envia uma mensagem de controle para o Ableton Live. A mensagem segue o protocolo MIDI (*Musical Instrument Digital Interface*) (MOOG, 1986).

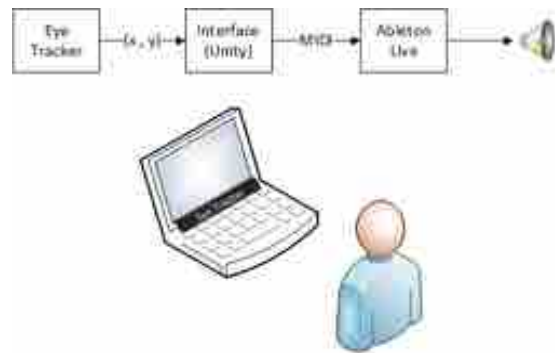


Fig. 1 – Descrição geral da interface.

2.1. A INTERFACE GRÁFICA

A Figura 2 mostra a principal interface gráfica do projeto. No centro tem-se uma matriz, representada por 3 linhas e 3 colunas, onde cada célula contém um determinado evento musical previamente configurado. Os botões *play* e *pause* são utilizados para iniciar ou parar um determinado evento musical, respectivamente. A combinação da escolha de uma célula da matriz com um dos botões representa uma ação sobre o evento musical da célula. Porém, no lado esquerdo da Figura 2 esses botões têm como objetivo executar o evento fora da performance, para que o usuário possa ter conhecimento prévio do mesmo, executando em um canal de saída de áudio específico para tal fim. No lado direito, os botões têm ação direta na performance. Além desses, tem-se o botão *pause all*, que tem função de interromper todos os áudios, e o botão de configuração no canto inferior esquerdo (Seção 2.3). O usuário pode selecionar com o movimento dos olhos qualquer elemento da cena, exceto o de configuração, para isso basta focar em um elemento e piscar os olhos.

Królak e Strumillo (2012) descrevem uma interface para humano-computador baseado na visão, onde detecta o piscar dos olhos e, a partir de então, realiza determinadas ações. Os autores relatam que o tempo médio para piscar de forma involuntário está em torno de 100 milissegundos. Além disso, o piscar dos olhos de forma voluntária é mostrado com um tempo superior a 250 milissegundos, assim o tempo de mínimo e máximo considerado para representar uma ação na interface é de 250 e 2000 milissegundos, respectivamente. Para o trabalho em questão, utilizou-se as mesmas durações, mínima e máxima, para definir uma ação de piscar os olhos voluntariamente. Vale ressaltar que o ponto branco, no centro de cada elemento, tem finalidade de auxiliar o usuário a focar sua visão no momento da escolha.

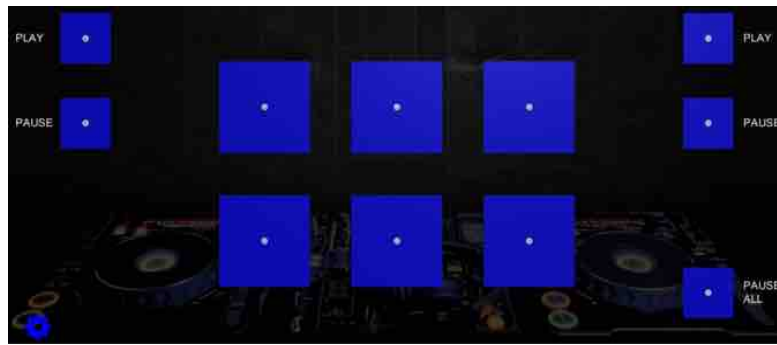


Fig. 2 – Interface gráfica principal.

2.2. CONEXÃO COM O ABLETON LIVE

O software Ableton Live vem sendo amplamente utilizado para estudos musicais em todo o mundo (HANTRAKUL; KACZMAREK, 2014; MARTIN et al., 2012; ROBERTSON; PLUMBLEY, 2007), sua vasta gama de ferramentas tem chamado atenção e se mostrou muito útil no contexto desse trabalho. Assim, toda saída sonora é executada a partir do Live, a Figura Fig. 3 mostra um exemplo de conexão da matriz com o software. Através das setas pode-se notar que cada célula está ligada com um elemento *Audio Track* para a performance. Cada elemento, de 1 a 6, possui uma amostra (*sample*) configurada previamente para ser executada. Cada célula da matriz, ao ser selecionada em conjunto com o botão *play* ou *pause*, envia uma mensagem MIDI ao Live que está mapeada em um *Audio track* de modo a permitir seu controle, podendo parar ou executar o mesmo.

Vale ressaltar que diferentes elementos *Audio Track* executam as amostras de modo a sobrepor os áudios, já as amostras dentro de um mesmo *Audio Track* são executadas de forma excludente, ou seja, interrompe-se a execução da amostra atual para iniciar a próxima, proporcionando a não sobreposição. O *Audio Track* 7 (Figura 3) é utilizado para ouvir o áudio fora da performance principal, cada célula também é ligada a uma amostra dentro do mesmo e todo controle é feito por mensagens MIDI, utilizando os botões *play* e *pause* mostrados no lado esquerdo da Figura 2.

A conexão não se limita apenas ao que foi apresentado na Figura 3, apesar de ter sido tomada como pioneira nos primórdios do desenvolvimento, essa ligação fornece a possibilidade da execução dos áudios de forma mútua (com sobreposição de sons), o que pode ser prejudicial em determinados momentos de uma performance. Com base nessa afirmação, a conexão com Live pode ser configurada de modo a evitar a execução mútua de todas, ou certas, amostras de áudio. A Figura 4 mostra uma forma de ligação onde nenhuma amostra de áudio é executada em simultâneo, devido a essência comportamental do *Audio Track*. Pode-se perceber que a Figura 3 e Figura 4 demonstram os extremos entre os tipos de execução, mas as conexões não se limitam apenas a esses exemplos, assim, uma forma híbrida pode ser configurada para uma melhor adequação a performance. A configuração do

mapeamento das mensagens MIDI (ligação das células e o Live) e escolha dos áudios ficam a cargo do responsável (pessoa capacitada para tal planejamento), que pode adotar uma proposta que mais lhe for conveniente.

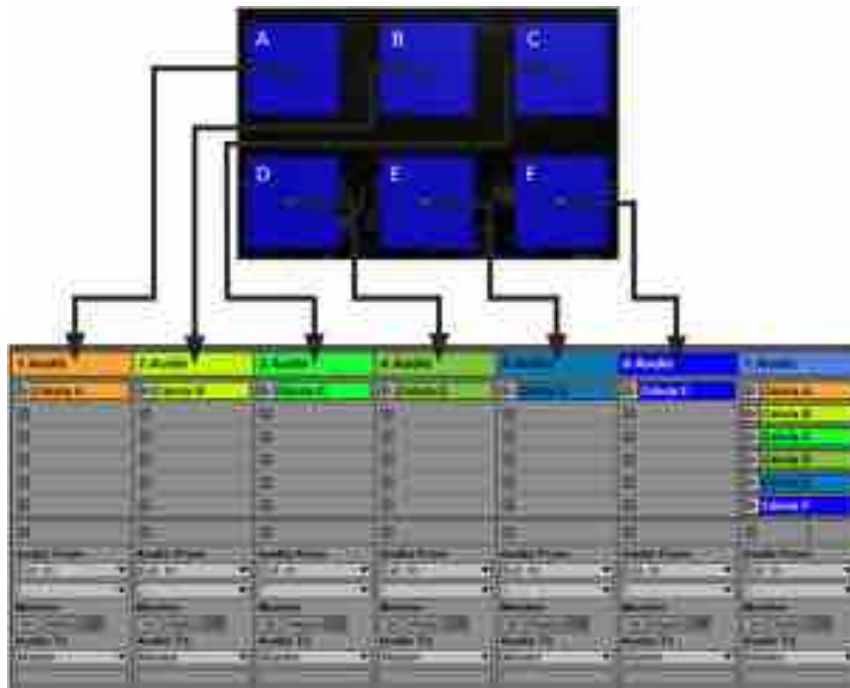


Fig. 3 – Conexão entre a interface e o Live com possibilidade de execução mútua.

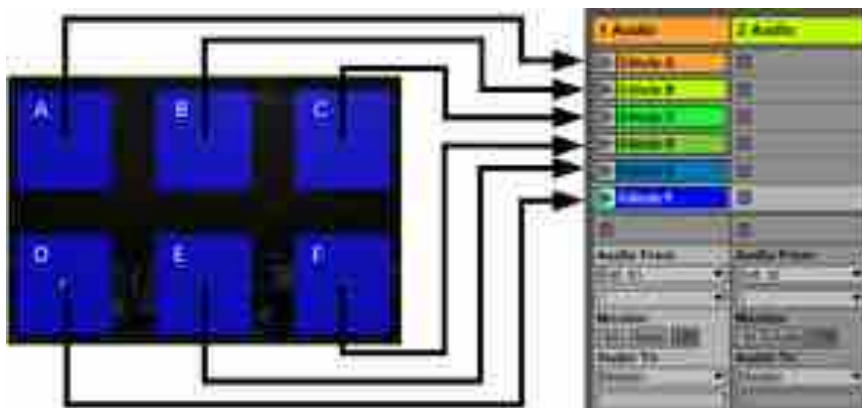


Fig. 4 – Conexão entre a interface e o Live para execução exclusiva.

2.3. CONFIGURAÇÃO DOS CAMINHOS

Para a inclusão de pessoas leigas no campo musical, a interface mostra sugestões para a escolha da próxima célula, ou seja, quando o botão *play* é acionado, os potenciais caminhos são mostrados. Um esquema de seis cores (número máximo de possibilidades de escolha) foi determinado. Quanto mais escuro for o azul da célula, mais potencialmente adequada é a escolha. Além das cores, é possível

desabilitar uma célula, a mesma aparece translúcida. A Figura 5 exemplifica um possível cenário que possui desde uma célula desabilitada (primeira célula) até a escolha com maior potencial (última célula).

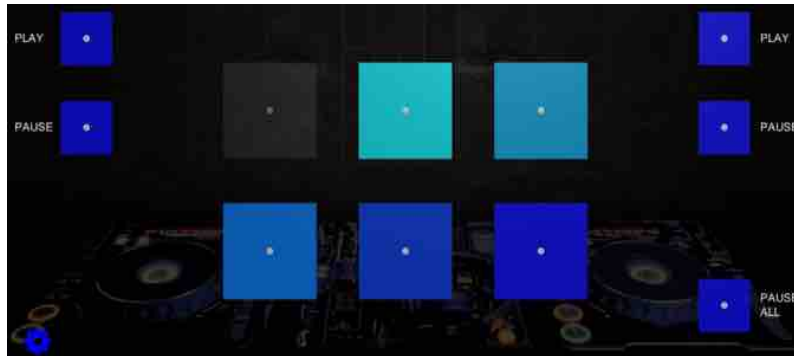


Fig. 5 – Exemplo de uma cena da interface com os possíveis caminhos.

Para melhor exemplificar a Figura 5, podemos utilizar um grafo (Figura 6) que mostra os possíveis caminhos e seus respectivos pesos, ou seja, quanto maior o peso mais potencial tem a escolha. Esta configuração foi obtida a partir da escolha da célula A, nota-se que não há caminhos de A para A o que é representado na Figura 5 como uma célula inativa.

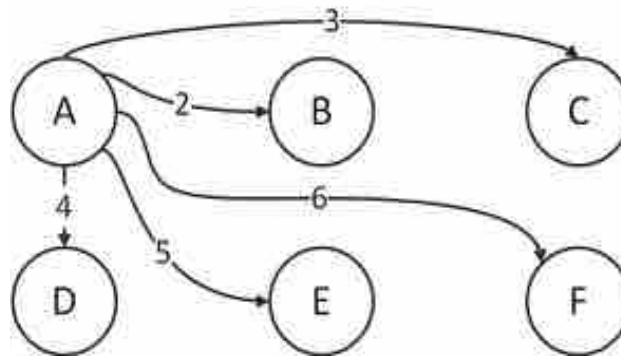


Fig. 6 – Exemplo do grafo dos caminhos. Cada letra representa um elemento que controla uma amostra de áudio e o valor numérico de transição descreve a potencialidade de migração de uma amostra para outra.

Para a configuração dos possíveis caminhos foi desenvolvida uma outra interface (Figura 7), onde existe uma outra matriz de configuração para cada célula e o conteúdo das posições da pequena matriz é o peso do caminho. Assim, após a escolha de uma célula, o programa executa a matriz de configuração para mostrar as próximas escolhas possíveis. O grafo da Figura 6 é representado na primeira matriz da Figura 7. Essa configuração pode ser feita de forma manual por uma pessoa capacitada para a tal função, assim como a escolha das amostras de áudio que executam no Ableton Live.

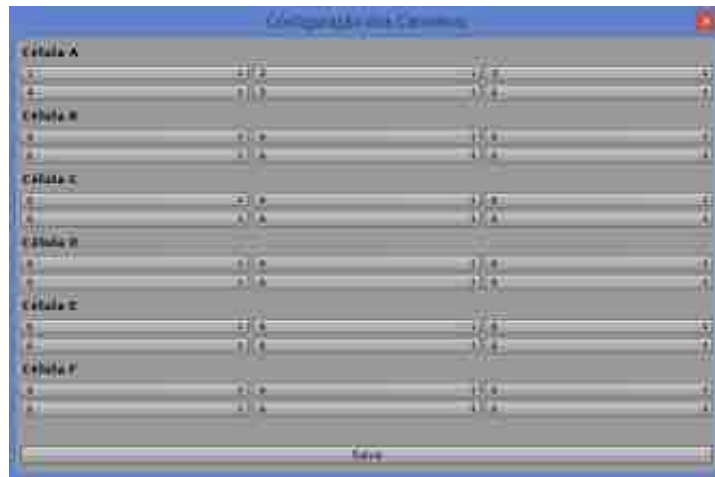


Fig. 7 – Configuração dos caminhos. Interface para definição dos pesos de transição entre as células que representam as amostras de áudio.

Para melhor exemplificar os caminhos como um todo, a Figura 8 ilustra a relação entre todas as células, mostrando todos os caminhos e seus pesos de ponderação. Já a Figura 9 mostra como o grafo foi configurado.

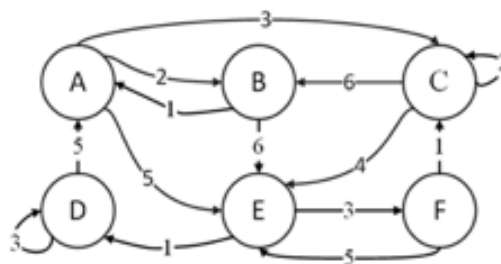


Fig. 8 – Exemplo dos caminhos. Cada valor de transição representa a potencialidade da transição de uma amostra para outra e cada letra representa uma célula que controla uma amostra de áudio.

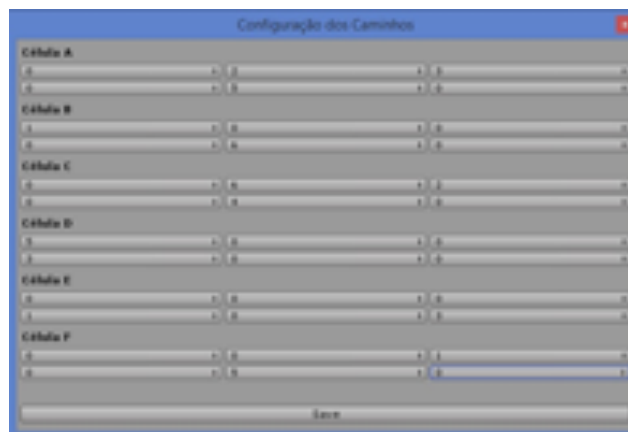


Fig. 9 – Configuração dos pesos para os caminhos. Interface para organização dos pesos de transição entre as amostras de áudio.

2.4. AUTOMATIZAÇÃO DA CONFIGURAÇÃO DOS CAMINHOS

Music Information Retrieval (MIR) é uma área de pesquisa interdisciplinar que engloba ciência da computação e recuperação de informação, musicologia e teoria musical, engenharia de áudio e processamento digital de sinais, além de outras áreas. Seu objetivo é desenvolver modos de gerenciar coleções de material musical para preservação, acesso, pesquisa e outros (FUTRELLE; DOWNIE, 2003). Assim, o auxílio de métodos computacionais para classificação das amostras de áudios e avaliação dos pesos, utilizados nos caminhos entre as células, é de extrema importância para este trabalho. Algumas propostas, ainda em desenvolvimento, são descritas a seguir. Os estudos, totalmente empíricos e experimentais, buscam similaridade, no domínio da frequência, entre amostras de áudio. Para tal fim, estuda-se a aplicabilidade da técnica de agrupamento K-means (JAIN, 2010) e da Transformada Rápida de Fourier (FFT - *Fast Fourier Transforms*) (COOLEY; LEWIS; WELCH, 1969). Por intermédio da técnica FFT pode-se converter um sinal discreto no domínio temporal para o da frequência, e, assim, obter o seu espectro de potência, onde pode-se analisar a amplitude para certas frequências. A fórmula para cálculo da FFT segue o seguinte formato:

$$X(j) = \sum_{n=0}^{N-1} A(n)W_N^{jn}$$

Onde $W_N = e^{\frac{2\pi i}{N}}$, o termo $A(n)$ representa o sinal no domínio do tempo, $X(j)$ o sinal resultante no domínio da frequência com $j = \{0, 1, \dots, N-1\}$ e N que corresponde o número de amostra do sinal A .

2.4.1. CONFIGURAÇÃO DE PESOS POR MEIO DA DISTÂNCIA EUCLIDIANA

A distância euclidiana é o caminho mais curto entre dois pontos (ANTON; RORRES, 2013), para exemplificar, o cálculo da distância entre os pontos $P = (p_1, p_2, \dots, p_n)$ e $Q = (q_1, q_2, \dots, q_n)$, em um espaço com dimensão n , temos a seguinte fórmula:

$$d_{P,Q} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

Assim, após aplicar a FFT para as amostras de áudio obtém-se vetores no domínio da frequência, similares aos pontos Q e P supracitados. Com as informações de cada amostra no domínio da

frequência, e considerando cada valor discreto de frequência como uma dimensão, pode-se encontrar a distância entre cada uma e usá-la para ponderar os pesos da matriz de caminhos, ou seja, quanto menor a distância entre duas amostras, maior é o peso para a transição entre elas. Nesse caso, a menor distância euclidiana significa que são próximas no contexto da frequência, demonstrando, assim, um possível potencial de combinação entre elas. Essa possibilidade de combinação é transmitida ao usuário por intermédio dos pesos dos caminhos entre as células.

2.4.2. CONFIGURAÇÃO POR MEIO DE AGRUPAMENTO

A criação de um cenário para interface requer seis amostras de áudio (uma para cada célula). Assim, quando a escolha se dá por entre uma extensa lista de opções, pode-se aplicar uma técnica de agrupamento para facilitar a seleção. A técnica chamada K-means, em conjunto com a FFT, está sendo estudada para esse trabalho. Essa técnica tem objetivo de diminuir o erro quadrático nos grupos, onde o erro pode ser calculado através da seguinte fórmula (JAIN, 2010):

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

Onde $X = \{x_1, x_2, \dots, x_m\}$ é um conjunto com m pontos de dimensão d para serem agrupados em K grupos, no qual $C = \{c_1, c_2, \dots, c_K\}$ representa os grupos, além de μ_k ser o valor médio do grupo c_k . Os três principais passos do algoritmo são: 1) selecionar uma partição com K grupos; 2) gerar nova partição, atribuindo cada ponto ao grupo com centro (μ_k) mais próximo ao mesmo; 3) calcular novo centro do grupo e 4) repete-se 2) e 3) até que os grupos se estabilizem.

Após a formação dos grupos, pelo algoritmo, pode-se escolher amostras de um determinado agrupamento, uma vez que as mesmas possuem certa similaridade na frequência. Com a escolha das amostras, os pesos podem ser determinados pela distância euclidiana, entre cada uma delas, ou de forma aleatória devido à similaridade entre elas. Vale salientar que o usuário tem livre arbítrio para utilização do método de classificação, com ou sem as distâncias euclidianas, para todas ou apenas algumas amostras. Além, é claro, de ter a possibilidade de não optar por nenhum método e configurar tudo de forma manual.

Como objetivo de estudo experimental, em um contexto fora da aplicação na interface, foi aplicado o algoritmo de agrupamento K-means, de modo a formar seis grupos, em segmentos (cortes sequenciais de 0,5 segundos cada) de algumas músicas. Para cada música, após a sua segmentação, foi

aplicado a FFT em cada segmento e, com os cortes no domínio da frequência, aplicou-se o algoritmo para formação dos seis grupos. Os cortes de cada agrupamento foram novamente concatenados para constituir um único áudio por grupo, respeitando sua sequência na música original. Com os resultados é possível observar uma certa linearidade em alguns pontos, ou seja, a música se reagrupou como a original. Os resultados, experimentais, são mostrados em: soundcloud.com/fftmeans/sets.

2.4.3. TESTES EXPERIMENTAIS PARA CONFIGURAÇÃO AUTOMÁTICA

Observando as amostras de áudio formada pelo exemplo experimental supracitado, a descontinuidade mais perceptível se dá quando a música contém a voz do artista, para tentar contornar tal situação, optou-se por executar teste em músicas instrumentais. Nestes testes, também experimentais, todos os áudios utilizados foram gerados de forma aleatória por intermédio da ferramenta WolframTones⁹, que possibilita a criação de pequenas amostras musicais de acordo com parâmetros configuráveis como: estilo, escala, instrumentos, tempo e etc. Vale ressaltar que o único parâmetro fixo utilizado foi o de estilo, para aos testes decidiu-se utilizar estilo o Jazz. Assim, foram geradas 5 amostras.

A segmentação das amostras segue basicamente o que foi descrito na subseção 2.4.2, porém o tempo de corte foi determinado como sendo 5 segundos, para que o usuário tenha tempo hábil para escolher a próxima célula antes que a anterior termine sua execução. Assim, as amostras geradas pelos WolframTones foram segmentadas sequencialmente em cortes de 5 segundos, os cortes foram convertidos para o domínio da frequência e por fim foram classificados de modo a formar seis grupos.

Após o corte das amostras e formação dos *clusters*, escolheu-se dois cortes pertencentes a um mesmo agrupamento e dois de agrupamentos diferentes para verificar a existência da similaridade e não similaridade. A Figura 10 mostra os sinais dos cortes de mesmo grupo, onde os gráficos a esquerda representam um corte e a direita o outro, e pode-se notar uma certa semelhança nos gráficos de frequência. Já a Figura 11 demonstra os sinais dos cortes pertencentes a grupos diferentes e pode-se observar certa disparidade entre os espectros de potência.

⁹ WOLFRAM. WolframTones. Disponível em: <<http://tones.wolfram.com>>. Acesso em: 18 ago. 2018.

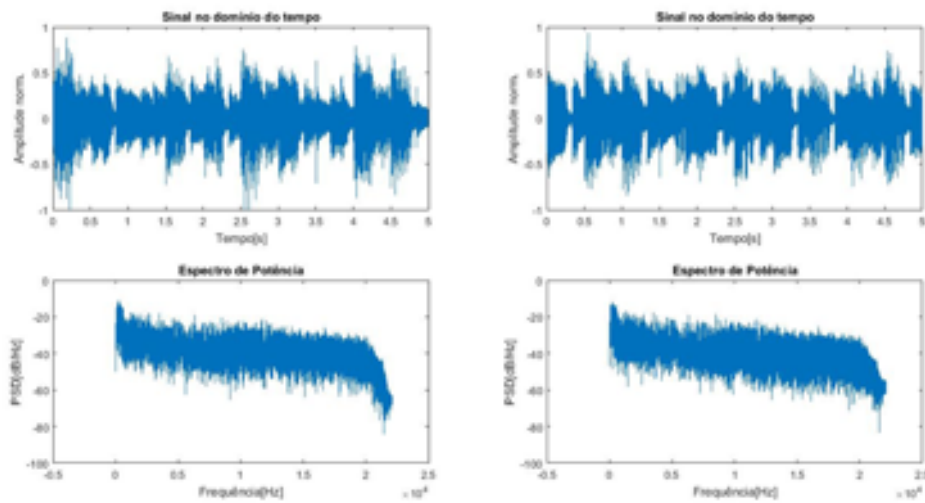


Fig. 10 – Análise de amostras pertencentes ao mesmo cluster.

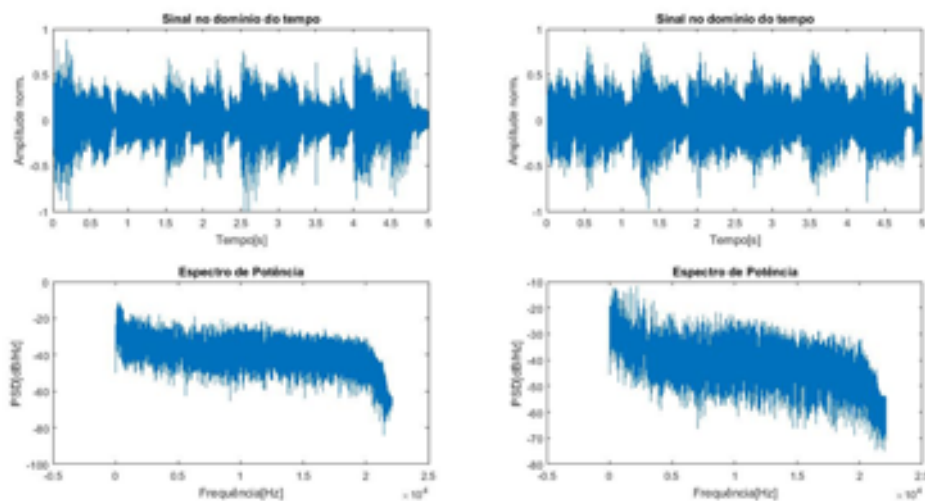


Fig. 11 – Análise de amostras contida em clusters diferentes.

O resultado para os testes com as amostras sobre o estilo musical Jazz pode ser encontrado em: (OLHAR MUSICAL, 2018a, 2018b). Vale ressaltar que foram utilizados cortes pertencentes a um mesmo *cluster*, porém no primeiro teste utilizou-se os pesos dos caminhos definidos de forma aleatória e, no segundo, os pesos definidos de forma proporcional à distância euclidiana entre os cortes, vale salientar que a conexão com o Live foi configurada da mesma maneira apresentada na Figura 4.

Além desses, outros exemplos são apresentados em: (OLHAR MUSICAL, 2018c, 2018d). Porém, nesses casos, as amostras foram escolhidas diretamente, ou seja, não houve processo de agrupamento e a ligação com o Live utilizada foi a demonstrada na Figura 3, porém sem sugestões de caminhos (forma livre).

3. CONCLUSÕES E TRABALHOS FUTUROS

O presente artigo apresentou uma interface para controle musical, pelo movimento dos olhos, com objetivo de promover a inclusão de pessoas no campo da música, e, ao mesmo tempo, uma proposta para incluir usuários com baixo conhecimento musical. Além da possibilidade de performances de UbiMus. Demonstrações da interface podem ser visualizadas em: (OLHAR MUSICAL, 2018a, 2018b, 2018c, 2018d).

O dispositivo de rastreamento ocular (Tobii EyeX) se mostrou bastante usual no contexto de controlar interface computacional para proporcionar o fazer musical, o fato de poder tomar decisões apenas com os movimentos dos olhos o torna extremamente acessível a pessoas com deficiência motora. A migração da execução do som para Ableton Live propiciou uma ampla gama de possibilidades sonoras, pois o mesmo permite aplicar diversas técnicas já desenvolvidas e homologadas, além de ser bastante utilizado por pessoas envolvidas com música, ou seja, existem vários indivíduos capacitados para o seu uso.

Os gráficos apresentados na Seção 3 mostram certo potencial na utilização do método de agrupamento K-means por similaridade em frequência. Para o agrupamento com as amostras de Jazz, geradas automaticamente, foi notado que a maioria dos cortes que continham um *cluster* eram provenientes de uma mesma amostra, que confirma a similaridade. Porém, vale ressaltar, que esse estudo tem objetivo de experimentação sonora e auxílio na configuração automática de caminhos da interface.

Com o apoio do Grupo de EXperimentação Sonora (GEXS) e do Núcleo Espírito-Santense de Computação Musical (NESCoM), formado por pesquisadores e alunos da Universidade Federal do Espírito Santo (UFES), os trabalhos futuros implicam no estudo de incorporar mais opções de controle na interface, montar vários cenários base com diferentes estilos musicais para disponibilizar ao usuário e o estudo de técnicas automatizadas para criar esses cenários. Além disso, aplicação de testes em pessoas com baixa mobilidade, visando buscar melhorias na plataforma e aprovação das mesmas, para que com a aceitação possa ser disponibilizado o código fonte, de forma gratuita, a toda comunidade. Outra perspectiva de futuro é trabalhar em conjunto com musicoterapeutas para utilizar a ferramenta em suas terapias.

REFERÊNCIAS

ALVA, M. et al. An image based eye controlled assistive system for paralytic patients. In: INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS, COMPUTING AND IT APPLICATIONS, 2., 2017, Mumbai. **Anais...** IEEE, 2017.

- ANTON, H.; RORRES, C. **Elementary Linear Algebra: Applications Version**. 11. ed. USA: Wiley Global Education, 2013.
- BRASIL. Lei n. 13.146, de 6 de jul. de 2015. **Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência)**. Brasília, DF, jul. 2017. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/113146.htm>.
- CAMPOREZ, H. et al. RoboMus: Sistema de performances musicais robóticas. In: WORKSHOP ON UBIQUITOUS MUSIC, 7., 2016, Foz do Iguaçu. **Anais...** Foz do Iguaçu: 2016.
- COOLEY, J. W.; LEWIS, P. A. W.; WELCH, P. D. The Fast Fourier Transform and its Applications. **IEEE Transactions on Education**, v. 12, n. 1, p. 27–34, 1969.
- CORRÊA, A. G. D. et al. GenVirtual: um Jogo Musical para Reabilitação de Indivíduos com Necessidades Especiais. **Revista Brasileira de Informática na Educação**, v. 16, n. 1, p. 9–17, 2008.
- FORNARI, J. Aplicações artísticas de ubimus. **NICS Reports**, n. 8, p. 16–26, 2014.
- FUTRELLE, J.; DOWNIE, J. S. Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000–2002. **Journal of New Music Research**, v. 32, n. 2, p. 121–131, 2003.
- HANTRAKUL, L.; KACZMAREK, K. Implementations of the Leap Motion device in sound synthesis and interactive live performance. In: INTERNATIONAL WORKSHOP ON MOVEMENT AND COMPUTING, 14., 2014, Paris. **Anais...** NY, USA: ACM, 2014.
- HORNOF, A. J.; SATO, L. EyeMusic: Making Music with the Eyes. In: INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 4., 2004, Hamamatsu. **Anais...** Singapore: National University of Singapore, 2004.
- IBGE. **Censo Demográfico 2010 - Características Gerais da População, Religião, e Pessoas com Deficiência**. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf>.
- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010.
- KELLER, D. et al. Música Ubíqua: Conceito e Motivação. In: CONGRESSO DA ANPPOM, 19., 2009, Curitiba. **Anais...** Curitiba: PPGM/UFPR, 2009.
- KELLER, D. et al. Convergent Trends Toward Ubiquitous Music. **Journal of New Music Research**, v. 40, n. 3, p. 265–276, 2011.
- KIM, J.; SCHIEMER, G.; NARUSHIMA, T. Oculog: playing with eye movements. In: INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 7., 2007, New York, USA. **Anais...** NY, USA: ACM, 2007.
- KRÓLAK, A.; STRUMILŁO, P. Eye-blink detection system for human-computer interaction. **Universal Access in the Information Society**, v. 11, n. 4, p. 409–419, 2012.
- LAZZARINI, V. et al. Development Tools for Ubiquitous Music on the World Wide Web. In: KELLER, D.; LAZZARINI, V.; PIMENTA, M. (Eds.). **Ubiquitous Music**. Cham: Springer, 2014a. p. 111–128.
- LAZZARINI, V. et al. Prototyping of Ubiquitous Music Ecosystems. In: WORKSHOP ON UBIQUITOUS MUSIC, 5., 2014, Vitória. **Anais...** Vitória: 2014b.
- LOURO, V. S.; IKUTA, C. Y.; NASCIMENTO, M. Música e deficiência: levantamento de adaptações para o fazer musical de pessoas com deficiência. **Arquivos Brasileiros de Paralisia Cerebral**, v. 1, n. 2, p. 11–17, 2005.
- LUPU, R. G. et al. Eye tracking user interface for Internet access used in assistive technology. In: E-

HEALTH AND BIOENGINEERING CONFERENCE, 6., 2017, Sinaia. **Anais...IEEE**, 2017.

MARTIN, A. et al. Creative experiments using a system for learning high-level performance structure in ableton live. In: SOUND AND MUSIC COMPUTING CONFERENCE, 9., 2012, Copenhagen. **Anais...2012**.

MOOG, R. A. MIDI: Musical Instrument Digital Interface. **J. Audio Eng. Soc**, v. 34, n. 5, p. 394–404, 1986.

OLHAR MUSICAL. Olhar Musical - Experimentação Sonora com Jazz e Distância Euclidiana. 2018a. Disponível em: <<https://youtu.be/PAGRCXpyuwg>>. Acesso em: <28 set. 2018>.

OLHAR MUSICAL. Olhar Musical - Experimentação Sonora com Jazz e Distâncias Aleatórias. 2018b. Disponível em: <https://youtu.be/ZKgkqn168_k>. Acesso em: <28 set. 2018>.

OLHAR MUSICAL. Música Eletrônica Feita com os Olhos. 2018c. Disponível em: <<https://youtu.be/YDj8pr3Cnx8>>. Acesso em: <28 set. 2018>.

OLHAR MUSICAL. Olhar Musical - Música Feita com os Olhos #1. 2018d. Disponível em: <<https://youtu.be/PzSIIt603LWo>>. Acesso em: <28 set. 2018>.

PIMENTA, M. S. et al. Methods in Creativity-Centred Design for Ubiquitous Musical Activities. In: KELLER, D.; LAZZARINI, V.; PIMENTA, M. (Eds.). **Ubiquitous Music**. Cham: Springer, 2014. p. 25–48.

POOLE, A.; BALL, L. Eye tracking in human-computer interaction and usability research: Current status and future prospects. In: ENCYCLOPEDIA OF HUMAN COMPUTER INTERACTION. USA: Idea Group Reference, 2006. p. 211–219.

ROBERTSON, A.; PLUMBLEY, M. D. B-Keeper: A Beat-Tracker for Live Performance. In: INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 7., 2017, New York. **Anais...** New York: ACM, 2007.

VAMVAKOUSIS, Z.; RAMIREZ, R. Temporal control in the EyeHarp gaze-controlled musical interface. In: INTERNATIONAL CONFERENCE ON NEW INTERFACES FOR MUSICAL EXPRESSION, 12., 2012, Ann Arbor. **Anais...** Ann Arbor: University of Michigan, 2012.

WHO, (WORLD HEALTH ORGANIZATION). **International Classification of Functioning, Disability and Health**, 2001. Disponível em:

<http://apps.who.int/iris/bitstream/handle/10665/43737/9789241547321_eng.pdf;jsessionid=36A69A0C518E07E7411F5C69E0BFE282?sequence=1>. Acesso em: 28 set. 2018.

WHO, (WORLD HEALTH ORGANIZATION). **Relatório mundial sobre a deficiência**, 2012.

Disponível em:

<http://apps.who.int/iris/bitstream/handle/10665/70670/WHO_NMH_VIP_11.01_por.pdf?sequence=9>. Acesso em: 1 jun. 2018.