

**PRODUÇÃO MUSICAL POR COMANDOS DE VOZ EM UM APLICATIVO DE  
PERCEPÇÃO SONORA**

***PRODUCCIÓN MUSICAL POR COMANDOS DE VOZ EN UNA APLICACIÓN DE  
PERCEPCIÓN DE SONIDO***

***MUSIC PRODUCTION BY VOICE COMMANDS IN A SOUND PERCEPTION  
APPLICATION***



Leonardo Porto PASSOS  
Universidade Estadual de Campinas (UNICAMP)  
e-mail: leoportopassos@gmail.com



José FORNARI  
Universidade Estadual de Campinas (UNICAMP)  
e-mail: fornari@unicamp.br



**Como referenciar este artigo**

PASSOS, L. P. FORNARI, J. Produção musical por comandos de voz em um aplicativo de percepção sonora. **Revista Hipótese**, Bauru, v. 8, esp. 1, e022020, 2022. e-ISSN: 2446-7154. DOI: <https://doi.org/10.47519/eiaerh.v8.2022.ID414>

**Submetido em:** 10/03/2022

**Revisões requeridas em:** 05/05/2022

**Aprovado em:** 01/07/2022

**Publicado em:** 01/12/2022

*Produção musical por comandos de voz em um aplicativo de percepção sonora*

**RESUMO:** Neste artigo, apresentamos o protótipo de um aplicativo *web* acessível a deficientes visuais para educação em produção musical e treinamento em percepção sonora, com o qual o usuário pode, por meio de entradas por comandos de voz, mixar a música em execução, ao adicionar e remover instrumentos musicais e efeitos de áudio e alterar o *panning*. A mixagem é de fundamental importância no processo de gravação musical, mas muitos ouvintes e músicos desconhecem as ferramentas e técnicas utilizadas nessa etapa, não conseguindo identificar tais procedimentos, que conferem singularidade e características estéticas especiais a uma gravação, o que motivou o desenvolvimento do *app* proposto.

**PALAVRAS-CHAVE:** Produção musical. Percepção sonora. Desenvolvimento de aplicativo.

**RESUMEN:** *En este artículo presentamos el prototipo de una aplicación web accesible para personas con discapacidad visual para la educación en producción musical y la formación en percepción del sonido, con la cual el usuario puede, a través de comandos de voz, mezclar la música en ejecución, agregando y quitando instrumentos musicales, y efectos de audio y cambiar la panorámica. La mezcla es de fundamental importancia en el proceso de grabación musical, pero muchos oyentes y músicos desconocen las herramientas y técnicas que se emplean en esta etapa, no pudiendo identificar dichos procedimientos, los cuales le dan singularidad y características estéticas especiales a una grabación, lo que motivó el desarrollo de la aplicación propuesta.*

**PALABRAS CLAVE:** *Producción musical. Percepción del sonido. Desarrollo de aplicaciones.*

**ABSTRACT:** *In this article, we present the prototype of a web application accessible to the visually impaired for music production education and training in sound perception. Through voice commands, the user can mix the music being played by adding and removing musical instruments and audio effects and changing panning. Mixing is of fundamental importance in the musical recording process, but many listeners and musicians are unaware of the tools and techniques used in this stage, not being able to identify such procedures, which give uniqueness and special aesthetic characteristics to a recording, which motivated the development of the proposed app.*

**KEYWORDS:** *Music production. Sound perception. Application development.*

## Introdução

É muito comum que um ouvinte, principalmente um não músico, possua uma predileção especial por determinada composição musical sem que saiba ao certo o porquê dessa afetividade, considerando haver algo naquela música que evoca suas emoções, mas foge à sua compreensão. E quando a mesma música é apreciada em uma *performance* ao vivo, sem a utilização dos mesmos recursos de estúdio, muitas vezes aquela relação afetiva de antes se rompe, e a sensação já não é a mesma, ainda que a *performance* tenha sido bastante fidedigna à gravação musical tão estimada por aquele apreciador. O que pode ocorrer nesses casos é que o ouvinte possui algum apreço pela estética imbuída pelo produtor musical àquela gravação (sem contar a participação do compositor e do intérprete, já que o nosso foco aqui é a produção musical, mais precisamente a mixagem), conferindo a ela novos contornos (para além daqueles oferecidos por compositores e intérpretes), que muitas vezes são difíceis de serem reproduzidos durante a *performance*, por diversas razões, que vão desde as diferenças de tratamento acústico do ambiente de gravação do estúdio em relação ao local da *performance* até os equipamentos disponíveis para a coleta, registro, processamento e geração sonora.

Nas palavras de David Huron (2015, p. 1), “A música consegue evocar uma ampla gama de estados de sentimento, do trivial ao sublime.”,<sup>1</sup> isso ocorre por conta de quatro tipos de geradores emocionais: 1) associação: “[...] certos sons ou padrões sonoros podem ser associados a experiências emocionais passadas.”;<sup>2</sup> 2) empático: “[...] o ouvinte reconhece características acústicas associadas a emoções particulares.”;<sup>3</sup> 3) cognitivo: “Pensamentos conscientes podem levar o ouvinte a uma experiência particular.”;<sup>4</sup> e 4) sinalização: “[...] um sinal que muda o comportamento do observador.”<sup>5</sup> (HURON, 2012, p. 479).

Além dos papéis do compositor e do intérprete em evocar sentimentos ao ouvinte, há também a participação do produtor musical no processo de mixagem, conforme Richard James Burgess (2013, p. 73): “A mixagem estende todas as técnicas musicais que a precedem, fortalecendo a percepção da música ao reforçar a estrutura, a orquestração e o afeto emocional para a audiência pretendida.”<sup>6</sup> A mixagem pode intensificar a evocação de emoções, ou contribuir para que ela ocorra, por meio do uso adequado de equalizador, filtros, distorção, *chorus*, dinâmicas,

<sup>1</sup> “*Music is capable of evoking a wide range of feeling states from the pedestrian to the sublime.*”

<sup>2</sup> “[...] *certain sounds or sound patterns may become associated with past emotional experiences.*”

<sup>3</sup> “[...] *a listener recognizes acoustic features associated with particular emotions.*”

<sup>4</sup> “*Conscious thoughts can lead a listener to a particular experience.*”

<sup>5</sup> “[...] *a signal is to change the behavior of the observer.*”

<sup>6</sup> “*Mixing extends all the musical techniques that precede it, strengthening the perception of the song by reinforcing the structure, orchestration, and emotional affect for the intended audience.*”

compressor, reverberação, eco, *pitch shifting*, etc. (CASE, 2011), e diz respeito ao seguinte processo:

[...] refere-se à mixagem original de uma faixa quando a instrumentação e os vocais são equilibrados entre si e quaisquer efeitos ou tratamentos necessários são adicionados. [...] Uma mixagem deve otimizar a música, o vocal, as *performances*, o arranjo e a engenharia. Deve soar bem em uma ampla gama de sistemas de alta e baixa gama e em qualquer volume<sup>7</sup> (BURGESS, 2013, p. 102, tradução nossa).

A produção musical – que inclui gravação sonora, arranjo, orquestração, efeitos, mixagem, masterização, etc., como veremos em mais detalhes adiante – é fundamental para que seja possível obter qualidade na captação dos instrumentos e vozes e para que tudo possa ser ouvido com clareza e definição, de acordo com conceitos estéticos variados, que podem até mesmo privilegiar certas “imperfeições” e a baixa fidelidade sonora<sup>8</sup> (*low-fidelity* ou *lo-fi*). Há grande preocupação com a clareza das texturas musicais<sup>9</sup> e com a separação das partes, em um processo tornará imperceptíveis os procedimentos de gravação, ou ao menos reduzir os métodos de produção musical (captação e gravação, adição de efeitos, mixagem, masterização) de modo que a gravação seja percebida como uma representação fiel ou “real” da *performance* musical (TURINO, 2008).

Porém, como afirma Burgess (2013, p. 2, tradução nossa), os processos e as técnicas da produção musical, bem como seus resultados, são desconhecidos por muitas pessoas, até mesmo por músicos: “[...] creio que a produção musical é uma arte mal compreendida, mesmo na indústria.”<sup>10</sup> [musical]. E é com o propósito de oferecer a experiência de realizar uma mixagem musical básica que foi desenvolvido o protótipo de um aplicativo *web*<sup>11</sup> de treinamento em percepção musical, e desse modo, útil para a educação em produção musical, já que permite ao usuário compreender certos recursos e técnicas utilizados por produtores musicais em mixagem e concepções criativas em termos estéticos.

<sup>7</sup> “[...] *mix* refers to the original mix of a track when the instrumentation and vocals are balanced with each other and any needed effects or treatments are added. [...] A mix should optimize the song, the vocal, the performances, the arrangement, and the engineering. It should sound good on a wide range of high- and low-end systems, and at any volume.”

<sup>8</sup> Disponível em: <https://ora.ox.ac.uk/objects/uuid:cc84039c-3d30-484e-84b4-8535ba4a54f8>. Acesso em: 10 jan. 2022.

<sup>9</sup> “O termo textura se refere à maneira como os materiais melódicos, rítmicos e harmônicos são tecidos juntos em uma composição” (*The term texture refers to the way the melodic, rhythmic, and harmonic materials are woven together in a composition*) (BENWARD; SAKER, 2009, p. 145, tradução nossa).

<sup>10</sup> “[...] *I felt that the art of music production was poorly understood, even within the industry*”.

<sup>11</sup> Disponível em: <https://edu.gcfglobal.org/pt/informatica-basica/o-que-e-um-aplicativo-web/1/>. Acesso em: 10 jan. 2022.

Para tal, explicaremos a seguir o que é a produção musical, bem como a definição de percepção musical e as razões que nos levaram a optar pelo desenvolvimento de um aplicativo com entradas (*inputs*) por comandos de voz.

## Produção musical

A história da produção musical teve início com o surgimento da gravação, reprodução e mídia, atribuídas a Thomas Alva Edison com a invenção do seu Fonógrafo, em 1877, um aparelho para gravação e reprodução de sons a partir de um cilindro, o que se configurou como um marco conceitual e estético da produção musical por possibilitar, de fato, a “solidificação”, por assim dizer, da intangibilidade do material sonoro que compõe a música, permitindo toda forma de processamento e análise sonora. Com este novo aparelho e o decorrente desenvolvimento de novas tecnologias para gravação e reprodução sonora – o que trouxe novas oportunidades para o registro musical, antes possível somente pela notação musical (com suas limitações, pois a expressividade de uma *performance* não é registrada pela notação), e para a composição musical –, surgiu a necessidade de técnicas capazes de combinar composição, arranjo, orquestração, interpretação, improvisações, timbres e *performance* em um todo sônico imutável (*immutable sonic whole*) (BURGESS, 2014, p. 1), ou um “objeto sonoro”<sup>12</sup> imutável.

Na definição de Burgess:

A produção musical é a extensão tecnológica da composição e orquestração. Ela captura a plenitude de uma composição, sua orquestração e as intenções performativas do(s) compositor(es). Em sua precisão e capacidade inerente de capturar sutilezas culturais, individuais, ambientais, timbrísticas e interpretativas junto com entonação, tempo, intenção e significado (exceto quando se busca o amorfo), é superior à música escrita e às tradições orais. A produção musical não é apenas representativa, mas também uma arte em si<sup>13</sup> (BURGESS, 2013, p. 5, tradução nossa).

De acordo com o *Grammy Award Eligible Credit Definitions* (RECORDING ACADEMY, 2019), o produtor musical é o responsável por decisões criativas, técnicas e estéticas que atendam aos objetivos do artista e do proprietário dos direitos autorais da gravação de som na criação de conteúdo musical, muitas vezes sendo considerado, quando já não é o

<sup>12</sup> Termo criado por Pierre Schaeffer para referir-se a um trecho de áudio com uma unidade de informação sonora cuja referência imagética é latente ou inexistente (MELO; PALOMBINI, 2006).

<sup>13</sup> “*Music production is the technological extension of composition and orchestration. It captures the fullness of a composition, its orchestration, and the performative intentions of the composer(s). In its precision and inherent ability to capture cultural, individual, environmental, timbral, and interpretive subtleties along with those of intonation, timing, intention, and meaning (except where amorphousness is specified), it is superior to written music and oral traditions. Music production is not only representational but also an art in itself.*”

caso, outro membro do grupo musical, com a mesma ou até maior importância que os músicos. O produtor pode executar, dirigir *performances*, escolher tomadas ou versões finais e supervisionar a seleção de músicas, músicos, cantores, arranjadores, estúdios, etc. É também o responsável por realizar ou supervisionar a mixagem, a masterização<sup>14</sup> e o controle de qualidade geral de uma gravação musical.

Pode-se definir a mixagem como a utilização, de forma criativa e por vezes intuitiva, de técnicas e ferramentas para mesclar, moldar e equalizar o som de um ou mais canais de áudio, com conteúdo de distintas fontes sonoras, para alcançar um objetivo estético específico (ARAÚJO, 2015). A utilização da mixagem de forma criativa, bem como seus aspectos técnicos, pode ser evidenciada de forma mais clara e detalhada:

Mixar música está relacionado com processar apresentações musicais gravadas. O objetivo desse processamento pode ser fazer com que a gravação soe natural e realista, como se você estivesse na sala quando os músicos se apresentaram. E também pode ser usado para alterar drasticamente o caráter sônico da gravação, criando uma paisagem sonora muito diferente que talvez nem seja possível de alcançar na vida real. Para fazer isso, o engenheiro de mixagem possui uma grande variedade de ferramentas, analógicas e digitais. Essas ferramentas são chamadas de processadores de sinais ou efeitos. [...] Os engenheiros de mixagem podem, portanto, usar o processamento de sinal por razões outras que não meramente técnicas. O processamento de sinais pode ser usado de maneiras estéticas e criativas para fazer as coisas parecerem maiores, mais apaixonadas e mais emocionais. Mesmo que o sinal original possa ser fortemente enviesado ou distorcido no processo, fazendo com que soe pouco natural ou com uma qualidade de áudio inferior, muitas vezes é considerado desejável. [...] Ao mixar música, os engenheiros de mixagem às vezes usam o processamento de sinal para suscitar um impacto emocional específico no ouvinte. Por exemplo, uma faixa vocal pode ser mixada com muita reverberação e delay com a intenção de induzir uma emoção onírica ou melancólica<sup>15</sup> (OLSSON, 2015, p. 2, tradução nossa).

6

<sup>14</sup> “A masterização é a etapa final de otimização do material gravado enquanto ele é transferido para o(s) formato(s) que será(ão) utilizado(s) no processo de fabricação” (*Mastering is the final stage of optimization of the recorded material while transferring it to the format(s) that will be used in the manufacturing process*) (BURGESS, 2014, p. 48, tradução nossa).

<sup>15</sup> “*Mixing music is about processing recorded musical performances. The goal of this processing can be to make the recording sound natural and realistic, just as if you were there in the room when the musicians performed. It can also be used to drastically change the sonic character of the recording, creating a vastly different soundscape maybe not even possible to achieve in real life. To do this, the mixing engineer has a wide variety of tools, both analog and digital. These tools are called signal processors, or effects. [...] Mixing engineers may therefore use signal processing for reasons other than mere technical. Signal processing can be used in aesthetic and creative ways as to make things sound bigger, more passionate and more emotional. Even though the original signal may be heavily skewed or distorted in the process, making it sound unnatural or of a lesser audio quality, it is often times found desirable. [...] When mixing music, mixing engineers sometimes use signal processing in a way as to achieve a specific emotional impact in the listener. For example, a vocal track may be mixed with lots of reverb and delay with the intention of inducing a dreamy or melancholic emotion.*”

De acordo com Burgess (2013), a produção musical é antecedida pela fase de pré-produção e seguida pela etapa de pós-produção:

- a) Pré-produção: fase preparatória de tomada de decisões para a seleção, organização e refinamento do material musical;
- b) Produção: preparar (escolher e posicionar) microfones, instrumentos, fones de ouvido, efeitos (*delay*, reverberação, etc.), equalizadores e compressores (pré-mixagem) e realizar as sessões de gravações, tudo com base em escolhas estéticas iniciadas na etapa anterior;
- c) Pós-produção: etapa da realização da mixagem, que consiste em equilibrar e otimizar os componentes da produção para o máximo impacto musical e clareza perceptiva das partes, utilizar recursos como equalização, compressão, *panning*,<sup>16</sup> compressão, limitação, expansão, *gating*, reverberação, *delays* e outros efeitos para otimizar os sons, aumentar seu impacto e garantir que ocupem seu próprio espaço no espectro de áudio.

E por fim, a masterização, sendo a preparação de uma mídia ou arquivo digital único com a junção de todas as demais gravações que compõem uma peça musical (música) ou um conjunto delas (álbum).

É comum que algumas dessas etapas sejam realizadas por profissionais distintos, especializados em funções específicas. Mas “Com a onipresença das estações de trabalho de áudio digital (DAW [*digital audio workstation*]), a maioria dos produtores, desde a virada do século, conseguem gravar e manipular áudio na DAW de sua escolha. Isso confunde ainda mais a distinção entre engenharia e produção de áudio.”<sup>17</sup> (BURGESS, 2013, p. 29, tradução nossa).

É na fase de pós-produção que se concentram as etapas mais relevantes para o presente estudo, em especial a mixagem:

Na fase de pós-produção, o engenheiro de mixagem combina as gravações por meio da mixagem e edição para obter uma mixagem final. Predominantemente, quanto mais habilidoso for o engenheiro de mixagem, melhor será a mixagem final em termos de qualidade de produção. A mixagem de áudio envolve a aplicação de técnicas de processamento de sinal a cada trilha de áudio gravada, por meio das quais o engenheiro manipula as características dinâmicas (equilíbrio e compressão de faixa dinâmica), espaciais (panorâmica estéreo ou *surround* e reverberação) e espectrais (equalização) do material de origem. Uma vez que a mixagem final foi criada, ela é enviada para um estúdio de masterização, em que um processamento adicional é aplicado para que a gravação musical possa ser distribuída para

<sup>16</sup> “O movimento panorâmico, o posicionamento para a esquerda ou direita dos sons entre os alto-falantes” (*Panning, the left/right placement of sounds between the speakers*) (GIBSON, 2005, p. 22, tradução nossa).

<sup>17</sup> “*With the ubiquity of digital audio workstations (DAWs), most producers, since the turn of the century, have been capable of recording and manipulating audio in their workstation of choice. This further blurs the distinction between audio engineering and production.*”

*Produção musical por comandos de voz em um aplicativo de percepção sonora*

audição em um ambiente doméstico ou de clube<sup>18</sup> (RONAN; REISS; GUNES, 2018, p. 1, tradução nossa).

Independentemente da função específica desempenhada pelo profissional de produção musical, a percepção musical aguçada é preponderante para a realização desse tipo de trabalho.

## Percepção musical

A mixagem é uma etapa de fundamental importância para o resultado da gravação de uma peça musical, conforme defende David Gibson (2005, p. 17, tradução nossa): “A mixagem pode ser apenas uma pequena parte de tudo o que é necessário para criar uma ótima gravação geral; no entanto, é um dos aspectos mais poderosos, porque a mixagem pode ser utilizada para esconder pontos fracos em outras áreas.”<sup>19</sup> No entanto, o mesmo autor salienta que “[...] a maioria das pessoas não diferencia as partes individuais que compõem uma peça musical gravada. Elas escutam um ‘som’ geral e raramente separam a mixagem da música.”<sup>20</sup> (GIBSON, 2005, p. 1, tradução nossa).

Para que seja possível realizar uma mixagem satisfatória, é necessário, antes, concentrar-se na percepção musical, uma capacidade humana inata de “[...] perceber auditivamente, refletir e agir criativamente sobre a música.” (BERNARDES, 2001, p. 75), desenvolvida ao longo de nossa evolução enquanto espécie a partir da percepção sonora, um mecanismo de defesa e proteção para que nos mantivéssemos sempre atentos aos acontecimentos no entorno (por isso somos capazes de fechar os olhos, mas não os ouvidos), e que também pode se tornar mais apurada por meio de treinamento.

Para melhor compreensão do que se trata a percepção musical, recorreremos a uma definição mais detalhada:

A percepção musical é a percepção sonora no contexto musical, ou seja, a capacidade de perceber ondas sonoras como parte de uma linguagem musical. A percepção musical envolve principalmente a percepção sonora, que é a capacidade de identificar atributos físicos do som, como volume, timbre e

<sup>18</sup> “*In the post-production stage, the mix engineer combines the recordings through mixing and editing to achieve a final mix. Predominately, the more skilled the mix engineer is, the better the final mix sounds in terms of production quality. The mixing of audio involves applying signal processing techniques to each recorded audio track whereby the engineer manipulates the dynamics (balance and dynamic range compression), spatial (stereo or surround panning and reverberation), and spectral (equalisation) characteristics of the source material. Once the final mix has been created, it is sent to a mastering studio where additional processing is applied before it can be distributed for listening in a home or a club environment.*”

<sup>19</sup> “*The mix may only be one small part of everything that goes into creating a great overall recording; however, it is one of the most powerful aspects because the mix can be utilized to hide weaknesses in other areas.*”

<sup>20</sup> “[...] *most people don’t differentiate between the individual parts that make up a recorded piece of music. They hear an overall “sound” and rarely separate the mix from the music.*”

Leonardo Porto PASSOS e José FORNARI

afinação. Além da percepção sonora, a percepção musical envolve também elementos musicais como melodia (percepção melódica), ritmo (percepção rítmica) e harmonia (percepção harmônica) (MATUNOBU, 2010, p. 22).

A percepção sonora ocorre, por exemplo, quando estamos em meio a uma paisagem sonora<sup>21</sup> complexa, constituída de diversos sons de origens variadas, somos imersos em uma grande quantidade de informações sonoras cuja fonte e natureza muitas vezes desconhecemos. Ainda assim, somos capazes de perceber claramente determinadas nuances desses sons – mesmo que não consigamos compreendê-los em profundidade –, como intensidade, altura, timbre, reverberação, etc. Estes são os aspectos perceptuais sonoros. Na música, um dos elementos fundamentais da percepção sonora diz respeito ao processamento mental, pela audição, de aspectos elementares do som, que descrevem características psicoacústicas<sup>22</sup> do material escutado” (FORNARI, 2010, p. 12).

Quando estamos imersos em uma paisagem sonora conturbada e percebemos informações sonoras por todo o redor, oriundas das mais diversas fontes,

Nossos ouvidos estão recebendo, traduzindo e enviando toda essa informação sonora para o cérebro pelo nervo auditivo, na forma de sinais elétricos. Apesar dessa informação perceptual estar emaranhada nos dois canais de recepção, que são os ouvidos, somos capazes, até certo ponto, de focalizar nossa atenção voluntariamente em uma única conversa, bem como mover nossa atenção de uma fonte sonora para outra, de acordo com nosso interesse, e desconsiderar o restante. Se, nesse ambiente sonoro tumultuado, alguém chama pelo nosso nome, especialmente se constatarmos que se trata de uma voz conhecida, a nossa atenção é imediatamente e involuntariamente deslocada para esta pessoa (FORNARI, 2010, p. 21).

Segundo a teoria da Gestalt (LERDAHL; JACKENDOFF, 1996; TENNEY; POLANSKY, 1980) existem quatro princípios básicos da identificação de objetos sonoros em música (FORNARI, 2010, p. 26-27):

---

<sup>21</sup> Conceito popularizado por R. Murray Schafer (2001, p. 24): “Uma paisagem sonora consiste em eventos *ouvidos* e não em objetos *vistos*”, e pode ser dividida em: sons fundamentais, as notas que identificam a escala ou tonalidade de uma música ou os sons criados pela geografia e pelo clima (água, vento, pássaros, insetos, animais); sinais, sons destacados e ouvidos conscientemente como recursos de avisos acústicos (sinos, apitos, buzinas, sirenes); e marcas sonoras, sons únicos de uma comunidade, que possuem certas qualidades que os tornam especialmente significativos ou notados pelas pessoas daquele lugar (derretimento de geleiras, vulcões em erupção, campos de enxofre fervente) (SCHAFFER, 2001).

<sup>22</sup> “Características psicoacústicas ocorrem em intervalos de tempo suficientemente pequenos, anteriores à formação de um modelo de memorização de informação sonora (assim, não existe distinção entre aspectos psicoacústicos sonoros e musicais). Tais aspectos estão associados a um intervalo de tempo conhecido na psicoacústica como intervalo da persistência auditiva, considerado em torno de 0,1s de duração. Eventos sonoros distintos, separados por intervalo de tempo menor que o da persistência auditiva são percebidos como um único evento sonoro” (FORNARI, 2010, p. 10).

*Produção musical por comandos de voz em um aplicativo de percepção sonora*

1. Estruturação: compreender um conjunto de diferentes eventos como uma estrutura única. Ex.: instrumentos, melodia, harmonia, ritmo, etc.;
2. Segregação: notar um evento que se sobressai em relação aos demais. Ex.: a melodia de um instrumento solo;
3. Pregnância: identificar primeiro e de forma clara as estruturas mais simples e regulares. Ex.: padrões rítmicos simples (ao contrário do que ocorre na polirritmia);
4. Constância: perceber a continuidade nas variações entre eventos consecutivos e entendê-los como pertencentes a um mesmo contexto. Ex.: um carro que passa com música tocando.

Com essa capacidade aprimorada de percepção sonora, R. Murray Schafer (2001, p. 25) afirma que: “O que o analista da paisagem sonora precisa fazer, em primeiro lugar, é descobrir os seus aspectos significativos, aqueles sons importantes devido a sua individualidade, quantidade ou preponderância.”. E essa descoberta deve ser realizada também pelo produtor musical, ao realizar a mixagem das partes que compõem uma peça musical gravada.

Schafer defende que uma das formas de se realizar o treinamento em percepção sonora é o que ele chama metaforicamente de “limpeza-de-ouvidos” (*ear cleaning*):

Começa-se ouvindo sons. O mundo é cheio de sons que podem ser ouvidos em toda a parte. As espécies mais óbvias de sons são também as menos ouvidas, essa é a razão da operação limpeza-de-ouvidos concentrar-se nelas. Alguns alunos limpam tanto seus ouvidos para ouvir os sons que os rodeiam que já podem partir para um estágio posterior e passar a analisá-los. Quando o processo de análise foi acurado, é possível reconstruir sinteticamente, ou ao menos imitar, um som que se ouve. Esse é o ponto em que a limpeza-de-ouvidos dá lugar ao treinamento auditivo (SCHAFFER, 1991, p. 103-104).

| 10

Diante do exposto, nota-se a importância da percepção sonora no viés musical, uma vez que o músico está também imerso em uma paisagem sonora que, no caso, refere-se à *performance* musical, sendo ainda mais significativa em termos de comunicação entre agentes, no caso da *performance* em grupo. Assim, foi desenvolvido um aplicativo web no qual o usuário poderá tanto realizar treinamento em percepção sonora quanto aprender algumas ferramentas utilizadas por produtores musicais, conforme apresentado na próxima seção.

## Modelo computacional de percepção musical

A melhor forma de aprender a ouvir é ouvindo,<sup>23</sup> como bem defende Schafer em seu livro *O ouvido pensante*, de 1991. Diante disso, uma aplicação *web* cujos *inputs* e *feedbacks* se dão por meio de sons apresenta um potencial elevado para o treinamento da percepção sonora, da mesma forma que os *audiogames*,<sup>24</sup> que apresentam algumas vantagens em educação musical pela ênfase nos recursos sonoros (músicas, efeitos sonoros e vozes) e diminuição ou até mesmo ausência de recursos visuais, conforme apontam Rovithis, Mniestris e Floros:

Nos *audiogames* [AGs] em particular, os jogadores precisam se concentrar em estímulos auditivos a fim de compreender e realizar as tarefas de jogo. A redução ou exclusão de informações visuais pode potencializar a aquisição de habilidades, como memória e concentração. Além disso, os AGs podem apresentar a todos, mesmo aos não músicos, conceitos e princípios musicais, servindo como plataformas nas quais os jogadores experimentam e realizam suas ideias. Assim, o design de AG pode desempenhar um papel inovador na pesquisa e na educação, especialmente em currículos relacionados à música e aos estudos de som<sup>25</sup> (ROVITHIS; MNIESTRIS; FLOROS, 2014, p. 1, tradução nossa).

Diante dessas possibilidades, apresentamos um protótipo simples de um aplicativo *web* de treinamento em percepção sonora e produção musical, com o qual o usuário poderá, por meio de entradas por comandos de voz (tornando o *app* acessível a deficientes visuais), mixar a música em execução, ao:

- Adicionar e remover instrumentos, nomeados como bumbo, caixa, tom, pratos, baixo, sintetizador, arpejo, melodia e efeitos;
- Ligar e desligar efeitos de áudio, nomeados como *chorus*, compressor, *delay*, distorção, *flanger*, filtro passa-alto, filtro passa-baixo, reverberação e tremolo;
- Alterar o *panning* para o centro (mono), para a esquerda ou para a direita.

<sup>23</sup> Ou ainda, “escutar escutando”, já que escutar é ouvir com atenção, conscientemente, conforme Houaiss (2009, verbete “escutar”): 1) estar consciente do que está ouvindo; 2) ficar atento para ouvir, dar atenção a; 3) esforçar-se para ouvir com clareza.

<sup>24</sup> *Audiogames* são jogos eletrônicos cujo conteúdo (narrativa, mecânicas, comunicação) é veiculado, principalmente ou exclusivamente, por meio do som (ROVITHIS; MNIESTRIS; FLOROS, 2014).

<sup>25</sup> “*The application of educational theories on computer games, as well as the interactivity that computer games, unlike other forms of entertainment, such as books, music and movies, can provide, have made them an increasingly used medium for education. Particularly in AGs, players need to focus on aural stimuli, in order to understand and accomplish the game-play tasks. The reduction or exclusion of visual information can enhance the acquisition of skills, such as memory and concentration. Furthermore, AGs can introduce everyone, even non-musicians, to musical concepts and principles, by serving as platforms, on which players experiment and realize their ideas. Thus, AG-design can play a groundbreaking role in research and education especially on curricula related to music and sound studies.*”

O protótipo foi desenvolvido na *engine* (*software* otimizado para a criação de *games*) Unity.<sup>26</sup> Sua escolha se deu por conta da possibilidade de integração com a *middleware*<sup>27</sup> Fmod,<sup>28</sup> que permitiu ativar ou desativar os instrumentos musicais e os efeitos de som conforme as entradas do usuário, bem como alterar o *panning* da música.

Com o intuito de manter o usuário do protótipo do app focado nos sons, sem possíveis dispersões por conta de muitos recursos visuais, optou-se pelo uso de entradas por comandos de voz, o que foi possibilitado pelo *plugin* WebGL Speech,<sup>29</sup> escrito na linguagem de programação C# (a mesma utilizada no Unity), que permite o reconhecimento de fala pelos navegadores de internet (*browser*) e que converte a fala do usuário em texto (*speech-to-text*), mais especificamente em uma variável do tipo string (que armazena palavras), que pode ser comparada ou manipulada para que seja possível transformar os comandos de fala do usuário em ações no aplicativo. Assim, o usuário pode, por exemplo, pronunciar o nome de um instrumento musical ou de um efeito sonoro, conforme a *Figura 1*, para desligá-lo, caso esteja ligado, ou ligá-lo, caso esteja desligado, o que foi possível com a criação de uma variável binária do tipo bool (que armazena apenas dois valores possíveis (*true* ou *false*, ou seja, verdadeiro ou falso) para saber se o instrumento em questão está ligado ou desligado, e assim realizar a ação comandada pelo usuário. Quando um instrumento ou efeito de áudio está ligado, seu nome aparece em verde na tela, e quando está desligado, o nome fica vermelho.

<sup>26</sup> Website oficial do Unity: <https://unity.com/pt>.

<sup>27</sup> Um *middleware* é um *software* de computador que fornece serviços para *softwares* aplicativos além daqueles disponíveis pelo sistema operacional.

<sup>28</sup> Website oficial do Fmod: <https://www.fmod.com/>.

<sup>29</sup> Disponível em: <https://assetstore.unity.com/packages/tools/audio/webgl-speech-105831>.

Leonardo Porto PASSOS e José FORNARI

**Figura 1** – Interface de usuário do protótipo, com os nomes dos instrumentos à esquerda, efeitos de áudio à direita e o indicador da reprodução da música (ícone de alto-falante) e a posição do *panning* ao centro.



Fonte: Elaborado pelos autores

A mesma premissa é utilizada para o usuário iniciar ou parar a reprodução da trilha musical, o que é indicado por um ícone de alto-falante no centro da tela, que fica vermelho quando a música não está em execução (*Figura 1*) e se torna verde quando a música está em execução (*Figura 2*). Para reproduzir a música, o usuário deve dizer “tocar música”, e para interrompê-la, deve pronunciar “parar música”.

13

**Figura 2** – O ícone de alto-falante verde ao centro indica que a música está em execução, bem como a cor dos nomes dos instrumentos e efeitos de áudio indica se eles estão ativos (verde) ou inativos (vermelho).



Fonte: Elaborado pelos autores.

Para alterar o *panning*, o usuário pode dizer “centro”, “esquerda” ou “direita”, para que a música seja executada, respectivamente, nos dois alto-falantes, em mono (*Figura 1*), somente no

*Produção musical por comandos de voz em um aplicativo de percepção sonora*

alto-falante da esquerda (*Figura 2*) ou somente no da direita, e uma letra (C, E ou D) aparecerá na parte inferior do centro da tela, abaixo da palavra “Panning”.

Por *default*, o aplicativo foi programado para reconhecer palavras pronunciadas no idioma português, mais especificamente o português brasileiro (PT-BR). Porém, por alguma razão desconhecida, o código utilizado não está funcionando como o esperado, e quando o usuário acessa o aplicativo por um navegador configurado em outro idioma que não o PT-BR, o protótipo não funciona, já que o aplicativo reconhece apenas palavras pronunciadas no idioma PT-BR. Desta forma, o usuário precisa alterar manualmente para PT-BR a configuração de idioma do navegador a fim de que o modelo funcione corretamente.

Além disso, alguns problemas ocorreram nos testes realizados com o reconhecimento de fala, que interpretava de maneira errônea a pronúncia das palavras em inglês ao transcrevê-las para o português. Diante disso, para que certas variáveis do tipo string fossem comparadas e o resultado pretendido fosse possível, foi necessário fazer adaptações ou o aportuguesamento de palavras do idioma inglês, como “*chorus*”, aportuguesado para “*córus*” e também adaptado para “*khorus*”, que são as formas que o reconhecimento de fala do plugin WebGL Speech costuma transcrever a pronúncia de “*chorus*”. O reconhecimento de fala dificilmente identificava a pronúncia da palavra “*flanger*”, e ao invés disso, compreendia a palavra “*ranger*”, e assim foram incluídas essas duas opções para ligar ou desligar este efeito quando o usuário pronúncia o seu nome.

Quando o usuário pronúncia alguma das palavras-chave, sendo os comandos para ativar ou desativar instrumentos ou efeitos, os respectivos parâmetros criados no Fmod são alterados, o que faz com que os efeitos ou instrumentos sejam ativados ou desativados. Ou seja, os parâmetros foram criados dentro do Fmod, mas são manipulados pelo Unity conforme as entradas por comandos de voz captadas pelo algoritmo de reconhecimento de fala do *plugin* WebGL Speech, que converte as palavras pronunciadas em texto (*speech-to-text*), e assim essas palavras são comparadas a variáveis do tipo string, e se a comparação for efetivada, ocorre uma determinada ação previamente programada. Foi esta sinergia entre Unity, Fmod e WebGL Speech que possibilitou o desenvolvimento do protótipo do aplicativo *web*.

O protótipo deste aplicativo está disponível para testes no [itch.io](https://itch.io)<sup>30</sup> (um site para hospedagem e distribuição, paga ou gratuita, de *games* independentes), e pode ser acessado pelo link <https://leopassos.itch.io/musicmixer>.

<sup>30</sup> “*itch.io is an open marketplace for independent digital creators with a focus on independent video games. It’s a platform that enables anyone to sell the content they’ve created. As a seller you’re in charge of how it’s done: you set the price, you run sales, and you design your pages. It’s never necessary to get votes, likes, or follows to*”  
Rev. *Hipótese*, Bauru, v. 8, esp. 1, e022020, 2022.

## Considerações finais

Em algumas sessões de apresentação do Music Mixer, bem como em alguns *playtests*,<sup>31</sup> as pessoas manifestaram entusiasmo e divertimento diante do aplicativo. Algumas melhorias precisam ser realizadas para que a proposta de oferecer educação em produção musical e treinamento em percepção sonora seja mais efetiva e se aproxime ainda mais da prática real de mixagem e de produção musical, ainda que dentro de certos limites, já que nem todas as técnicas e ferramentas de mixagem estão disponíveis no aplicativo. A utilização das entradas por reconhecimento de fala traz alguns entraves, como o *delay* entre a entrada e a ação e o *feedback* pelo aplicativo; os problemas por conta do idioma programado no reconhecimento de fala e o configurado no navegador; a certa constância na imprecisão do sistema de reconhecimento de fala, que muitas vezes não consegue captar corretamente a pronúncia do usuário, principalmente quando ele não está utilizando fones de ouvido, e o som emitido pelas caixas de som por vezes acabam atrapalhando reconhecimento de fala do aplicativo; e as limitações das entradas por reconhecimento de fala, sendo menos precisas e dinâmicas que outros tipos de entradas, como as realizadas por *mouse* ou tela sensível ao toque, que permitem maior variedade de comandos e ações pelo usuário.

Como possibilidades futuras, pretende-se permitir ao usuário: incluir *samples* de áudio | 15 capturados em tempo real; alterar o andamento da música; realizar entradas por toque de tela, em uma versão *mobile*, para aumentar as possibilidades de uso, dentre as quais se incluem a possibilidade de adicionar efeitos de áudio a instrumentos específicos, e não à música como um todo, como o caso atual do protótipo; e permitir o controle de volume e *panning* de cada instrumento. Com essas implementações, o usuário poderá realizar algo muito próximo, senão idêntico, ao que é a mixagem real das partes individuais que compõem uma peça musical gravada. Porém, toma-se como desafio realizar tais implementações sem abandonar a possibilidade de uso dos comandos de voz, para que assim o aplicativo permaneça acessível a deficientes visuais, além das possibilidades de uso da síntese de fala, com o sistema *text-to-speech*, também incluso no *plugin* WebGL Speech.

Como desdobramentos futuros de trabalhos derivados deste, com vias ao desenvolvimento de um aplicativo mais sofisticado e com possibilidades mais elaboradas de uso, pretendemos recorrer à utilização da metodologia da pesquisa-ação (TRIPP, 2005), que consiste

---

*get your content approved, and you can make changes to how you distribute your work as frequently as you like.*" Disponível em: <https://itch.io/>.

<sup>31</sup> Alguns dos *playtests* foram gravados e estão disponíveis em: <https://youtu.be/HrxLHzuhg3w>.

em: 1) desenvolvimento, 2) testes, 3) aprimoramentos, 4) coleta de resultados e 5) reinício do processo a partir da etapa 1. Para as etapas 2 e 4, cogita-se a coleta de dados por meio de *playtests* seguidos de preenchimento de relatório *on-line* por parte dos usuários, com a devida autorização prévia do Comitê de Ética em Pesquisa (CEP) da Universidade Estadual de Campinas (Unicamp).

## REFERÊNCIAS

ARAÚJO, D. V. G. **Uma breve história da mixagem**: Origem, técnicas, percepção e futuros avanços. Campinas, 2015. Dissertação (Mestrado em Música) – Instituto de Artes, Universidade Estadual de Campinas, São Paulo, 2015. Disponível em: <https://revistas.nics.unicamp.br/revistas/ojs/index.php/nr/article/view/190>. Acesso em: 06 maio 2021.

BENWARD, B.; SAKER, M. **Music in theory and practice**: v. 1. 8. ed. New York: McGraw-Hill, 2009.

BERNARDES, V. A percepção musical sob a ótica da linguagem. **Revista da Abem**, v. 9, n. 6, p. 73-82, set. 2001. Disponível em: [www.abemeducacaomusical.com.br/revistas/revistaabem/index.php/revistaabem/article/view/444](http://www.abemeducacaomusical.com.br/revistas/revistaabem/index.php/revistaabem/article/view/444). Acesso em: 05 nov. 2021.

BURGESS, R. J. **The art of music production**: The theory and practice. 4. ed. New York: Oxford University Press, 2013.

BURGESS, R. J. **The history of music production**. New York: Oxford University Press, 2014.

CASE, A. U. **Mix Smart**: Pro audio tips for your multitrack mix. Oxford: Focal Press, 2011.

FORNARI, J. Percepção, cognição e afeto musical. In: KELLER, D. (org.). **Criação musical e tecnologias**: Teoria e prática interdisciplinar. Goiânia: Anppom, 2010. Disponível em: [www.anppom.com.br/ebooks/index.php/pmb/catalog/book/2](http://www.anppom.com.br/ebooks/index.php/pmb/catalog/book/2). Acesso em: 21 jun. 2021.

GIBSON, D. **The art of mixing**: A visual guide to recording, engineering, and production. 2. ed. Boston: Thomson Course Technology, 2005.

HOUAISS, A [Instituto]. **Houaiss Eletrônico**. Versão 3.0. Rio de Janeiro: Objetiva, 2009.

HURON, D. Affect induction through musical sounds: an ethological perspective. **Phil. Trans. R. Soc. B**, v. 370, n. 1664, mar. 2015. Disponível em: <https://royalsocietypublishing.org/doi/full/10.1098/rstb.2014.0098>. Acesso em: 10 fev. 2021.

HURON, D. Understanding Music-related emotion: Leslons from Ethology. In: PROC. INTERN. CONF. ON MUSIC PERCEPTION AND COGNITION, 12.; TRIENNIAL CONF. OF THE EUROPEAN SOC. FOR THE COGNITIVE SCIENCES OF MUSIC, 8., 2012, Thessaloniki. **Anais** [...]. Thessaloniki, Greece, 2012.

LERDAHL, F.; JACKENDOFF, R. S. **A generative theory of tonal music**. 3. ed. London: MIT Press, 1996.

MATUNOBU, Y. **Desenvolvimento de software educativo para treinamento em percepção musical**. 2010. Monografia (Trabalho de Conclusão de Curso em Ciência da Computação) – Fundação de Ensino Eurípides Soares da Rocha, Centro Universitário Eurípides de Marília, São Paulo, 2010.

MELO, F.; PALOMBINI, C. O objeto sonoro de Pierre Schaeffer: Duas abordagens. *In: XVI ANPPOM*, 16., 2006, Brasília. **Anais [...]**. Brasília, 2006. Disponível em: [https://antigo.anppom.com.br/anais/anaiscongresso\\_anppom\\_2006/CDROM/COM/07\\_Com\\_TeoComp/sessao04/07\\_COM\\_TeoComp\\_0404-173.pdf](https://antigo.anppom.com.br/anais/anaiscongresso_anppom_2006/CDROM/COM/07_Com_TeoComp/sessao04/07_COM_TeoComp_0404-173.pdf). Acesso em: 24 maio 2021.

OLSSON, E. **Aesthetic signal processing in music production**: Is the intended emotional response achieved? Lulea. 2015. Monografia (Trabalho de Conclusão de Curso em Engenharia de Áudio) – Department of Arts, Communication and Education, Lulea University of Technology, 2015. Disponível em: <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1018575&dswid=-3321>. Acesso em: 23 fev. 2021.

RECORDING ACADEMY. **Producers & Engineers Wing, Technical Guidelines**. Producer Grammy Award Eligibility Crediting Definitions, March 01, 2019. Disponível em: [www.grammy.com/sites/com/files/producer\\_definitions\\_final\\_03\\_01\\_2019.pdf](http://www.grammy.com/sites/com/files/producer_definitions_final_03_01_2019.pdf). Acesso em: 17 jun. 2021.

RONAN, D.; REISS, J. D.; GUNES, H. An empirical approach to the relationship between emotion and music production quality. **ArXiv**, mar. 2018.

ROVITHIS, E.; MNIESTRIS, A.; FLOROS, A. Educational audio *game* design: sonification of the curriculum through a role-playing scenario in the audio *game* ‘Kronos’. *In: AM 2014*, 9., 2014, New York. **Anais [...]**. New York, NY, USA, 2014.

SCHAFFER, R. M. **A afinação do mundo**. São Paulo: Ed. Unesp, 2001.

SCHAFFER, R. M. **O ouvido pensante**. São Paulo: Fundação Editora da Unesp, 1991.

TENNEY, J.; POLANSKY, L. Temporal Gestalt perception in music. **Journal of Music Theory**, Autumn, v. 24, n. 2, p. 205-241, 1980. Disponível em: <https://www.jstor.org/stable/843503>. Acesso em: 19 fev. 2021.

TRIPP, D. Pesquisa-ação: Uma introdução metodológica. **Educação e Pesquisa**, São Paulo, v. 31, n. 3, p. 443-466, set./dez. 2005. Disponível em: <http://educa.fcc.org.br/pdf/ep/v31n03/v31n03a09.pdf>. Acesso em: 21 set. 2021.

TURINO, T. **Music as social life**: The politics of participation. Chicago: The University of Chicago Press, 2008.

*Produção musical por comandos de voz em um aplicativo de percepção sonora*

## **SOBRE OS AUTORES**

### **Leonardo Porto PASSOS**

Universidade Estadual de Campinas (UNICAMP), Campinas – SP – Brasil. Mestrando do Programa de Pós-Graduação em Música (PPGM) do Instituto de Artes (IA).

### **José FORNARI**

Universidade Estadual de Campinas (UNICAMP), Campinas – SP – Brasil. Pesquisador de carreira Pq do CPG/DM/IA. Doutorado em Engenharia Elétrica (UNICAMP).

**Processamento e edição: Editora Ibero-Americana de Educação.**

Correção, formatação, normalização e tradução.