UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

João Teixeira Araújo

Proposta de Arquitetura para Interação do Público em Performances Digitais

São João del-Rei

2021

Ficha catalográfica elaborada pela Divisão de Biblioteca (DIBIB)
e Núcleo de Tecnologia da Informação (NTINF) da UFSJ,
com os dados fornecidos pelo(a) autor(a)

A663p

Araújo, João.
Proposta de Arquitetura para Interação do Público em Performances Digitais / João Araújo ; orientador Flávio Luiz Schiavoni; coorientador Edimilson Batista dos Santos. -- São João del-Rei, 2021.
94 p.

Dissertação (Mestrado - Ciência da Computação) -- Universidade Federal de São João del-Rei, 2021.

1. Performances Digitais. 2. Instalações Digitais. 3. Interação do Público. 4. Recuperação da Informação Musical. 5. Computação Musical. I. Luiz Schiavoni, Flávio, orient. II. Batista dos Santos, Edimilson, co-orient. III. Título.

João Teixeira Araújo

Proposta de Arquitetura para Interação do Público em Performances Digitais

Dissertação apresentada ao programa de Pós-Graduação em Ciência da Computação da UFSJ para obtenção do título de mestre em Ciência da Computação.

Flávio Luiz Schiavoni
Orientador

Edimilson Batista dos Santos
Co-Orientador

Tiago Rodrigues Tavares
Convidado 1

Diego Roberto Colombo Dias
Convidado 2

São João del-Rei
2021

Assinado por concordância com ata de defesa realizada por videoconferência

Agradecimentos

Nada é por acaso. Todas as coisas acontecem devido a um propósito e cabe a nós entender e saber gerenciar as oportunidades que estão à nossa volta. Primeiramente gostaria de agradecer aos meus pais por todo o apoio que me deram, durante toda minha vida. Sem eles eu com certeza não teria chegado até aqui.

Comecei minha vida acadêmica em 2012, ao tentar graduação na Universidade Federal de Viçosa (UFV), campus Viçosa. Infelizmente a lista de espera fechou logo à minha frente. Eu era o próximo da lista e acabei não sendo chamado. Devido à este fato, tive que tentar graduação em um campus alternativo, na cidade de Florestal (UFV). Por incrível que pareça, lá pude conhecer um grande amigo meu, Leandro Pessoa, o qual foi primordial para meu futuro ingresso na Universidade de São João del-Rei (UFSJ).

Após a graduação, resolvi tentar pós-graduação na UFV em 2016 e acabei não sendo chamado mais uma vez. Então resolvi fazer algumas matérias por lá, porém como aluno não-vinculado. Por não estar vinculado à faculdade, após 6 meses resolvi estudar na Universidade Federal de Ouro Preto (UFOP), porém desta vez como aluno vinculado, com bolsa de estudos. Tentei iniciar meus estudos na área de Inteligência Artificial mas acabou não dando certo. Consequentemente acabei entrando na área de sistemas distribuídos. Após desenvolver toda a revisão bibliográfica do que até então seria utilizado na minha dissertação, após 1 ano percebi que eu estava no caminho errado e que eu realmente deveria fazer algo que me agradasse.

Até então a única certeza que eu tinha era de que eu gostava bastante de música, além da computação. Procurando mais a fundo sobre sub-áreas da computação, acabei entrando em contato com o Leandro, o qual me apresentou uma área que até então eu nunca tinha visto, a Computação Musical. Além disso, ele falou que havia um pesquisador na UFSJ que trabalhava nesta área, e que inclusive, possuía um modo de trabalho semelhante ao meu. No mesmo dia fui saber mais sobre o programa de pós-graduação da UFSJ. Coincidentemente, faltava apenas um dia para que as inscrições fossem encerradas. Mesmo com a ideia de que eu possivelmente perderia a bolsa da UFOP, me apressei, fui atrás de todos os documentos e acabei conseguindo realizar a inscrição. Pra variar, não entrei de primeira, as vagas haviam sido preenchidas. Porém, um dos alunos acabou não preenchendo a vaga pois havia passado em outra faculdade e acabei ingressando no mestrado em computação da UFSJ em 2018.

Por ter perdido a bolsa na UFOP, fiz uma prova para entrar na fila de prioridades, caso surgisse alguma bolsa no Dcomp (Departamento de Ciência da Computação da UFSJ). Depois de 3 meses sem bolsa recebi a notícia de que agora eu iria passar a receber

para estudar. Esta bolsa ajudou bastante na minha estadia em São João del-Rei e acabou durando todo o tempo do desenvolvimento deste trabalho.

Chegando em São João del-Rei, pude conhecer pessoalmente Flávio Luiz Schiavoni, meu atual orientador. Flávio me ensinou bastante coisa, inclusive coisas que vão além do ambiente acadêmico. Através dele pude ver o lado mais humano das pessoas, o qual é menos formal e mais natural e real. Nunca imaginei que eu sentaria em uma mesa de bar ou até mesmo faria música junto com meu orientador de mestrado...

Logo no início da minha pós-graduação nos juntamos e começamos a submeter alguns trabalhos. Com o passar do tempo fui aprendendo um pouco mais sobre o funcionamento de eventos, revistas e livros acadêmicos, além de como os trabalhos eram submetidos, corrigidos, editados, entre outros... Pude aprimorar minha capacidade de saber buscar, pesquisar, aprender e escrever sobre um determinado assunto, mesmo que eu não o conheça inicialmente. Também cheguei a aprender um pouco sobre edição e consequentemente acabei editando os anais do Simpósio Brasileiro de Computação Musical, organizando-os corretamente em um repositório[1], e o caderno de publicações do evento "Café com Ciência"[2] realizado na cidade de São João del-Rei. Ajudei também na organização do XVII Simpósio Brasileiro de Computação Musical, o que me fez enxergar sobre o quão difícil é para se elaborar, desenvolver e fazer acontecer um evento científico.

Em relação aos trabalhos publicados, todos eles trouxeram experiências positivas. Através deles pude conhecer cidades que até então, eu nem pensaria que visitaria a trabalho, como por exemplo São Paulo (USP), Conselheiro Lafaiete, Belo Horizonte e Mariana. Pude ir à Florianópolis (SC), por exemplo, com meu amigo Fábio Passos, o qual tive uma grande parceria durante todo o mestrado. Em cada cidade que passei, pude conhecer pessoas importantes da área em que atuo. É muito bom ver que existem pesquisadores que estão desenvolvendo trabalhos relacionados ao meu, além de poder receber feedback's positivos e de poder formar possíveis contatos futuros para o desenvolvimento de novos trabalhos.

As apresentações de trabalhos / artigos também me fizeram evoluir bastante. Até então o ato de apresentar qualquer trabalho em público era um grande empecilho para mim. Porém, com o tempo fui me acostumando e hoje em dia enxergo as apresentações em público como um trabalho como outro qualquer, o qual eu consigo realizar de maneira tranquila e fluida.

Por fim, agradeço imensamente por ter participado do "Chaos das 5". Até então eu não fazia ideia do que era performance e arte digital, não fazia ideia de que seria possível unir pessoas tão diferentes em um único projeto. A equipe era composta por futuros

[1] http://compmus.ime.usp.br/sbcm/

[2] https://cafecomciencia.wixsite.com/cafecomciencia?fbclid=IwAR1t7_SqJpaJNySHRusMy6ZRk6OMjcVwXVeWtQxAnvdwXUixybFyLumdLHg

cientistas, artistas, arquitetos, músicos, estilistas e até mesmo filósofos. Ao todo visitamos algumas cidades, passamos por diversos problemas relacionados à criação de eventos, e principalmente, tivemos muita vivência. Conheci muita gente especial que me ensinou bastante coisa: Gabriel Rocha me mostrou muito sobre humildade; Avner Maxiliano com todo o seu companheirismo; Rômulo Vieira com a sua calma e compreensão; Octávio Deluchi com toda sua habilidade musical fora do comum; André Gomes e Fred Rezende que sempre que precisei, me ajudavam durante nossos dias no laboratório; Igino com sua ótima habilidade de escrita e simpatia; Fábio Passos que acabou se tornando uma das minhas maiores amizades em São João del-Rei; Carlos Eduardo com toda sua tranquilidade e simpatia; João Lucas, que além de ser um ótimo skatista, é simplesmente uma das pessoas com o coração mais bondoso que eu já conheci na minha vida; Aneliza Prado, Patrick Veniali, Bruna Guimarães e Aretha que me ensinaram uma forma mais natural de me portar, mesmo estando em ambientes acadêmicos; Waldir Monge com toda sua proatividade e vontade em realizar seus projetos pessoais; Professor Rogério Constante que me ensinou um pouco mais sobre música; Professor Adilson Siqueira que me ensinou um pouco sobre performance teatral; Professor Flávio Schiavoni, o qual possibilitou o desenvolvimento de tudo isso; e aos demais participantes que de alguma forma contribuíram para o acontecimento desta experiência única.

Muito obrigado a todos que contribuíram de alguma forma para o desenvolvimento deste trabalho, evoluí bastante como pesquisador e como pessoa. Esta jornada estará guardada pra sempre em minhas memórias!

"Porque se chamava homem
Também se chamavam sonhos
E sonhos não envelhecem..."
(Clube da Esquina)

Resumo

A participação do público em espetáculos artísticos se faz presente na arte contemporânea como uma tentativa de quebra de papéis onde o público, o criador e o performer podem se misturar na criação e execução de um trabalho, visando proporcionar especificamente ao público um papel ativo e não de mero espectador. Este tipo de visão aliado ao constante avanço tecnológico acabou influenciando o surgimento de trabalhos artísticos os quais envolvem a participação do público através da utilização de aparatos tecnológicos, dando origem às chamadas performances e instalações digitais. Por envolver a utilização de diferentes algoritmos através de dispositivos móveis, computadores e sensores, as formas de interação do público nestes tipos de obras se tornaram bastante amplas. Desta forma, este trabalho tem como objetivo a elaboração de uma arquitetura para descrever como cada interação é dada em uma performance ou instalação digital, do ponto de vista tecnológico. Além disto, serão apresentadas 3 obras artísticas: a performance digital "O Chaos das 5", baseada no uso de dispositivos móveis para prover a interação do público; a instalação digital "Per(sino)ficação" a qual utiliza técnicas de processamento de imagem para realizar diferentes sínteses sonoras de acordo com a imagem captada do participante; e a instalação digital "Leiamídia" a qual utiliza técnicas de machine learning e de processamento de imagem e som para prover a interação dos participantes com a máquina.

Palavras-chaves: Performances Digitais, Instalações Digitais, Interação do Público, Recuperação da Informação Musical, Computação Musical.

Abstract

Public participation in artistic performances is present in contemporary art as an attempt to break roles where the public, the creator and the performer can merge in the creation and execution of a work, aiming specifically to provide the public with an active and non-active role of a spectator. This type of vision coupled with constant technological advancement ended up influencing the emergence of artistic works which involve public participation through the use of technological devices, giving rise to the so-called performances and digital installations. As it involves the use of different algorithms through mobile devices, computers and sensors, the forms of public interaction in these types of works have become quite wide. In this way, this work aims to develop an architecture to describe how each interaction is given in a performance or digital installation, from a technological point of view. In addition, 3 artistic works will be presented: the digital performance "O Chaos das 5", based on the use of mobile devices to provide audience interaction; the digital installation "Per(sino)ficação" which uses image processing techniques to perform different sound syntheses according to the image captured by the participant; and the digital installation "Leiamídia" which uses machine learning and image and sound processing techniques to provide the participants interaction with the machine.

Key-words: Digital Performances, Digital Installations, Audience Interaction, Music Information Retrieval, Computer Music.

Lista de ilustrações

Figura 1 – Performer encenando ao lado de seu piano (“4:33”, John Cage). 18
Figura 2 – Interação do público durante os “Parangolés” de Hélio Oiticica. 19
Figura 3 – Requisição de interação do público durante a performance “Santos Football Music” de Gilberto Mendes. 19
Figura 4 – Exemplo do Janelamento de Hamming sob um áudio digitalizado . . . 33
Figura 5 – Exemplo da função do filtro durante uma convolução em uma CNN. . . 36
Figura 6 – Um exemplo de arquitetura Pipes and Filters. 40
Figura 7 – Arquitetura de um IMD simplificada. 41
Figura 8 – Mapeamento do trabalho Per(sino)ficação (Seção 4.2) 49
Figura 9 – Arquitetura conceitual de interação do público. 52
Figura 10 – Proposta de arquitetura final de interação do público. 54
Figura 11 – Performance realizada pelos artistas no espetáculo O Chaos das 5. . . . 56
Figura 12 – Projeções visuais d’ “Chaos das 5”. 57
Figura 13 – Pré início do espetáculo “O Chaos das 5”. 58
Figura 14 – IMD’s do “O Chaos das 5” 59
Figura 15 – “O Chaos das 5” sob a arquitetura proposta 60
Figura 16 – Fluxograma da instalação Per(sino)ficação. 66
Figura 17 – Captação dos dados do participante através do PureData. 67
Figura 18 – Per(sino)ficação sob a arquitetura proposta. 68
Figura 19 – Apresentação da instalação Per(sino)ficação (SBCM 2019). 68
Figura 20 – Captação de características físicas do público em Per(sino)ficação (SBCM). 69
Figura 21 – Interface inicial da Instalação Leiamídia. 71
Figura 22 – Exemplo de interação da instalação Leiamídia. 71
Figura 23 – Diferentes efeitos visuais proporcionados pela classificação de sentimentos (Leiamídia). 72
Figura 24 – Fluxograma de funcionamento da aplicação utilizada no “Leiamídia”. . 73
Figura 25 – Exemplo abstrato da extração de características de um áudio digitalizado. 75
Figura 26 – Gráfico dos dados de treino, teste e validação (média x desvio padrão). 76
Figura 27 – Resultado do treinamento da CNN. 77
Figura 28 – Leiamídia sob a arquitetura de interação proposta. 77

Lista de abreviaturas e siglas

AIFF	*Audio Interchange File Format*
BMP	*Device Independent Bitmap*
CNN	*Convolutional Neural Network*
DFT	*Discrete Fourier Transform*
DHCP	*Dynamic Host Configuration Protocol*
DJ	*Disc Jockey*
DMTF	*Dual Tone Multiple Frequency*
DNS	*Domain Name System*
FFT	*Fast Fourier Transform*
GAN	*Generative Adversarial Networks*
HDF5	*Hierarchical Data Format 5*
IMD	Instrumento Musical Digital
IA	Inteligência Artificial
IP	*Internet Protocol*
JPEG	*Joint Photographic Experts Group*
JSP	*Java Server Pages*
LDR	*Light Dependent Resistor*
LED	*Light Emitting Diode*
MFCC	*Mel Frequency Cepstral Coefficients*
MHW	*Musical Haptic Wearables*
MIDI	*Musical Instrument Digital Interface*
MIR	*Musical Information Retrieval*
OSC	*Open Sound Control*

PA	*Public Audition*
PHP	*PHP: Hypertext Preprocessor*
PLOrk	*Princeton Laptop Orchestra*
SBCM	Simpósio Brasileiro de Computação Musical
TCP	*Transmission Control Protocol*
TESS	*Toronto Emotional Speech Set*
UDP	*User Datagram Protocol*
URL	*Uniform Resource Locator*
USC	*Ultra Sound Communication*
WAV	*Waveform Audio File Format*

Sumário

1 Introdução . . . 14

2 Conceitos e trabalhos relacionados . . . 16

2.1 A performance musical . . . 16
2.2 A Performance enquanto arte . . . 17
2.3 Performances Digitais . . . 20
2.4 Instalações de Arte Digital . . . 21
2.5 Formas de Interação do Público em performances digitais . . . 23
2.6 Avaliação da Participação do Público em performances digitais . . . 27
2.7 Recuperação de informação musical . . . 29
2.7.1 Descritores de áudio . . . 30
2.7.2 Mel Frequency Cepstral Coefficients - MFCC . . . 32
2.8 Processamento de áudio via Aprendizado de Máquina . . . 34
2.8.1 Redes Neurais Convolucionais . . . 35
2.8.2 Aprendizado de Máquina x Arte Digital . . . 37
2.9 A arquitetura Pipes and Filters . . . 40

3 Proposta de modelo para a participação do público . . . 41

3.1 Entrada . . . 42
3.2 Saída . . . 45
3.3 Mapeamento . . . 48
3.4 Comunicação entre os componentes . . . 51
3.5 Revisando o modelo . . . 52

4 Performances e Instalações Digitais Desenvolvidas . . . 55

4.1 Espetáculo “O Chaos das 5” . . . 55
4.1.1 Formas de interação de “O Chaos das 5” . . . 57
4.1.2 Avaliação da participação do público . . . 60
4.1.3 Lições aprendidas . . . 62
4.1.4 Análise Geral . . . 64
4.2 Instalação Digital Per(sino)ficação . . . 65
4.2.1 A instalação . . . 66
4.2.2 Aspectos Técnicos . . . 66
4.2.3 Avaliação da instalação . . . 67
4.3 Instalação Digital Leiamídia . . . 70
4.3.1 Funcionamento do sistema . . . 71

4.3.2 Aspectos de implementação 72
4.3.3 Avaliação da Instalação 77

5 Considerações Finais 79

Referências 83

1 Introdução

Enquanto em alguns espetáculos teatrais ou performances artísticas os dispositivos móveis ainda são convidados a serem silenciados durante a apresentação, em performances mediadas por tecnologias digitais a participação do público vem sendo estimulada cada vez mais, partindo da ideia de que tais dispositivos se tornaram ubíquos, presentes em todos os lugares e de forma tão transparente que chegamos a não perceber que os mesmos estão lá. Do ponto de vista musical, esta forma de participação mediada pelo uso de diferentes tecnologias demonstra uma mudança conceitual no ato de se produzir e escutar música, especialmente no que tange os papéis envolvidos nesta relação.

A música clássica de tradição europeia, por exemplo, baseia-se em três papéis: compositor, responsável pela criação; intérprete, responsável pela execução; e público, responsável pela escuta. A proposta da arte contemporânea de convidar o público a fazer parte do espetáculo quebra esta tradição por permitir outros papéis e um outro diálogo sonoro / musical, além de propiciar uma experiência única e especial do público na performance artística [Hödl et al., 2017]. Neste ponto, a mediação tecnológica digital acaba permitindo a participação do público na criação, na produção e/ou execução das chamadas performances digitais ou até mesmo em instalações digitais.

Levando em consideração o contexto abordado, esta dissertação foi desenvolvida baseada em 3 trabalhos: a performance digital "O Chaos das 5", desenvolvida pelos laboratórios ALICE e Ecolab do Gtrans (Grupo de pesquisa Transdisciplinar) da Universidade Federal de São João del-Rei (UFSJ), a qual envolve a participação do público em sua execução utilizando para isto dispositivos móveis para produção / síntese sonora e para o fornecimento de determinados dados; a instalação digital "Persinoficação" a qual envolve a participação do público via processamento de imagem para realização de diferentes sínteses de timbres sineiros, tomando como base a cidade de São João del-Rei, a qual possui uma cultura muito forte neste quesito; e a instalação digital "Leiamídia" a qual une técnicas de aprendizado de máquina e processamento de som para processar a fala do participante ao ler uma notícia sobre determinado assunto, classificando automaticamente sua emoção em tempo real.

Neste trabalho, o foco é dado apenas em interações as quais possuem alguma relação com algum tipo de tecnologia. Desta forma, o objetivo do trabalho é desenvolver uma arquitetura para esclarecer como a interação do público em performances / instalações digitais é dada, a partir do ponto de vista dos aparatos tecnológicos e das técnicas de processamento de dados utilizados. Resumidamente, a arquitetura irá demonstrar um fluxograma para cada tipo de interação a ser proporcionada em uma performance / ins-

talação digital de maneira simples e padronizada, a qual pode ser utilizada para facilitar o entendimento das interações como um todo ou até mesmo para comparar determinadas interações entre diferentes performances ou instalações tecnológicas.

Como prova de conceito da arquitetura, a performance digital "O Chaos das 5" e as instalações tecnológicas "Persinoficação" e "Leiamídia" serão validadas para uma melhor visualização do funcionamento e da eficácia da arquitetura proposta sob obras já consolidadas. Dentre os objetivos específicos deste trabalho, temos:

- Estudar sobre o surgimento das performances e instalações digitais no meio artístico;
- Estudar sobre a área de Music Retrieval Information (MIR), demonstrando brevemente o funcionamento das técnicas utilizadas na instalação digital chamada "Leiamídia";
- Definir uma proposta de arquitetura de participação do público em performances e instalações digitais;
- Apresentar a performance digital "O Chaos das 5" e as instalações "Persinoficação" e "Leiamídia", as quais foram implementadas no decorrer do curso, utilizando como base a arquitetura proposta.

Por fim, esta dissertação está dividida entre 5 capítulos. Seguido desta introdução, temos o Capítulo 2, o qual ficará responsável por apresentar conceitos envolvidos e trabalhos relacionados e que irá descrever historicamente o processo da imersão de técnicas participativas e de tecnologia em obras artísticas. Além disso, o capítulo irá exemplificar algumas destas formas de interação, baseando-se em obras digitais mais atuais além de descrever a área de MIR, abordando algumas características do som e sobre as técnicas utilizadas na obra "Leiamídia", passando pelo processamento do som via descritores de áudio e classificação automática a partir de redes neurais convolucionais. Já o Capítulo 3, ficará responsável por descrever a arquitetura proposta, a qual será utilizada nas 3 obras desenvolvidas para facilitar o entendimento sobre como a participação do público é dada em cada obra. O Capítulo 4 irá descrever as obras digitais criadas, passando pela definição e motivação de cada uma, além da validação da arquitetura proposta, incluindo a análise e avaliação da participação do público em cada uma. Por fim, o Capítulo 5 irá descrever as considerações finais, esclarecendo quais foram as lições aprendidas durante a criação das 3 obras digitais e o que a validação da arquitetura proposta pode proporcionar.

2 Conceitos e trabalhos relacionados

Neste capítulo serão apresentados alguns conceitos relacionados às obras desenvolvidas, além de alguns trabalhos relacionados que envolvem performance artística digital e participação do público em arte digital.

2.1 A performance musical

Desde o final do século XIX, a pesquisa em performance musical tem buscado entender o universo de variações e diversidades existentes na interpretação de uma mesma partitura pelo mesmo músico ou por músicos distintos, utilizando para isto a medição de parâmetros acústicos que podem ser manipulados pelo músico ao tocar seu instrumento e que influenciam a comunicação entre o intérprete e o público [Clarke, 2004]. Na música, é comum encontrar variações na performance envolvendo pequenas variações de durações, intensidades, afinações e articulações para comunicar suas intenções musicais que são desvios devidos à interpretação do performer em relação ao que está escrito na partitura. Além disto, há muitas manipulações do som que podem influenciar a construção de uma interpretação e alterar significativamente a expressividade musical de uma performance como a qualidade do ataque de uma nota, a sustentação da nota, a qualidade de transição entre notas, o agrupamento de notas, a articulação entre notas e o agrupamento de legatos ou de notas desligadas [Loureiro et al., 2009].

Se historicamente estas variações são analisadas de forma subjetiva, devido à dificuldade de se quantificar estes parâmetros, diversas pesquisas vêm tentando encontrar formas de mensurar e explicar com parâmetros quantitativos a individualidade, expressividade e inteligibilidade musical [Teixeira, 2010]. Uma das possibilidades de estudo destes parâmetros se dá pela análise e manipulação de atributos timbrísticos na performance musical. Entendendo que a altura e duração das notas de uma peça são definidas pelo compositor, é possível entender que as características timbrísticas de uma execução e sua expressividade são definidas pelo intérprete e somam o que o intérprete consegue trazer em uma peça para além do que o compositor especificou na partitura [Clarke, 2004]. Algumas características da performance podem ser notadas pelo compositor, como a escolha do instrumento, a dinâmica das notas, a articulação, a variação dinâmica e/ou variações temporais como *rallentando* e *accelerando*.

No entanto, tal notação pode não trazer parâmetros claros para a execução e são passíveis de interpretação pelo instrumentista e por isto podem ser consideradas como parte da performance mesmo quando há notações explícitas para a sua execução. Pesquisas na área artística / musical indicam ainda que músicos profissionais conseguem

manipular estas características de maneira bastante clara, demonstrando intencionalidade em suas performances enquanto músicos amadores não alcançam o mesmo êxito em sua intencionalidade [Williamon, 2004].

2.2 A Performance enquanto arte

O conceito de performance em artes pode ser bastante amplo e pode incluir diferentes visões dependendo da área artística em questão. Enquanto para a música a performance é uma área de pesquisa relacionada com o tocar um instrumento musical, para as artes da cena, ou artes performáticas, como o teatro, a dança e o circo, por exemplo, o termo performance tem um significado diferente e remete à participação do público em apresentações artísticas. Esta área tem sido fortemente influenciada pela tecnologia e pelos aparatos tecnológicos que tem permitido que este tipo de participação seja mediado pela tecnologia.

A relação do público com a arte é bastante antiga e já passou por diferentes abordagens dependendo do período histórico. Já houve uma tentativa de agradar o público, de entreter e depois de chocá-lo [Danto, 2008]. Mais recentemente, artistas vêm buscando envolver o público na arte e graças a isto a arte e a interação do público com a arte vem tomando mais e mais espaços, tirando o foco da arte para o objeto artístico e colocando em foco o processo da arte. Com isto, surge a possibilidade de uma experiência do público enquanto resultado do processo em diálogo com o objeto artístico em si.

Por volta do século XX começaram a surgir alguns movimentos artísticos que tinham por objetivo colocar o público como parte do espetáculo. Tal conceito foi explorado por artistas como Allan Kaprow em seus *happenings* com o grupo Fluxus. Este grupo, em conjunto com outros artistas dos anos 1950 e 1960 estabeleceram uma estrutura sobre um evento (*happening*) que pode ajudar a criar trabalhos artísticos fortemente marcados pela participação do público [Taylor, 2017]. Nestes eventos, o público possuía tarefas a serem executadas e precisavam tomar parte do espetáculo de maneira ativa.

Os *happenings* logo deixaram este formato de tarefas pré-estabelecidas, causando o surgimento da arte performática, onde o público toma parte do espetáculo de maneira livre e não de maneira guiada como anteriormente. Com isto, quebra-se a barreira entre artistas e público e também a relação entre a obra de arte como objeto final de um processo. Um exemplo de obra performativa é 4'33 de John Cage (1912-1992), onde o público foi o responsável pelo resultado sonoro do espetáculo [Rocha, 2005]. Nesta peça, o compositor não possui controle sobre o resultado sonoro e o músico no palco terá uma performance musical de difícil execução mas que não irá requerer domínio sobre o instrumento e que não envolve tocar notas em seu instrumento (Figura 1) [Santos, 2008].

O público, aguardando o músico executar sua peça, irá começar a fazer barulhos

Figura 1: Performer encenando ao lado de seu piano ("4:33", John Cage).

e este é o resultado estético esperado pelo artista. Cage defende:

> "Os próximos passos eram sociais, e ainda estavam sendo dados. Precisamos antes de tudo de uma música em que não apenas os sons são sons, mas em que pessoas são só pessoas, não sujeitas a leis estabelecidas por uma delas, mesmo que seja o "compositor" ou o "maestro". Finalmente nós precisamos de uma música que não mais provoque debates de participação do público, porque nela a divisão entre músicos (performers) e público não mais existe: uma música feita por todos. O que precisamos é de uma música que não requeira nenhum ensaio" [Cage, 1973].

Entre os resultados obtidos com esta nova forma de fazer arte está o indeterminismo, proposto por Nam June Paik, em 1965 [Ballard, 2013]. A indeterminação do resultado é o que guia este processo de criação e com isto, o artista não tem mais o controle total sobre o que irá acontecer e o público passa a influenciar o resultado da obra sendo que este resultado será a experiência estética em si.

No Brasil, os trabalhos artísticos participacionais possuem dois grandes expoentes como Lygia Clark e Helio Oiticica. Nos Parangolés de Helio Oiticica [Salomão, 2015] o público era convidado a participar da obra, vestir as capas coloridas e ser a obra de arte em movimento (Figura 2). As obras da artista mineira Lygia Clark *Caminhando* [Andrade et al., 2003] é caracterizada pelo convite para a participação do público e a experiência de fazer a obra é o resultado da própria obra.

Outra iniciativa pode ser visualizada em "Santos Football Music" (Figura 3), do compositor Gilberto Mendes, que em 1992 criou uma obra que integra o som instrumental em permanente transformação (sem motivo, série, desenvolvimento etc.), visando simular alguns aspectos de uma partida de futebol, além de um som concreto gravado em tape, teatro musical e a participação do público diretamente como executante, que participa da

Figura 2: Interação do público durante os "Parangolés" de Hélio Oiticica.

performance reagindo a diversas placas informativas apresentadas pelo performer [Chagas, 1992].

Figura 3: Requisição de interação do público durante a performance "Santos Football Music" de Gilberto Mendes.

Porém, com o avanço tecnológico novos tipos de performances artísticas envolvendo a participação do público foram surgindo, como as chamadas performances digitais, as quais serão explicadas na Seção 2.3.

2.3 Performances Digitais

A junção da arte com a tecnologia é bastante antiga visto que a arte sempre foi criada com base nos meios tecnológicos existentes em cada época. Porém, a apropriação que ela faz dos aparatos tecnológicos na forma de dispositivos eletrônicos, é algo que começou a ser consolidado por volta de 1966, a partir da performance "9 Evenings: Theatre and Engineering" [Bardiot et al., 2006]. Esta performance englobou artistas e engenheiros os quais se uniram com o objetivo de explorar a tecnologia, baseando-se na utilização de microfones, projetores e outros dispositivos eletrônicos para criação / composição artística. Estas performances, as quais utilizam algum tipo de tecnologia computacional como elemento primordial e não apenas como uma ferramenta utilizada para prover determinada interação, passaram a ser chamadas de performances digitais [Dixon, 2015]. A introdução do computador na arte, especificamente no Brasil, se deu através do surgimento da música eletroacústica, através de Jorge Antunes e Waldemar Cordeiro, o qual também forneceu embasamento para o surgimento de conceitos como: vídeo-arte; *computer art*; *computer music*; arte comunicação; holografia; poesia intersemiótica; e intersecção arte ciência [Melo et al., 2010].

Logo após algumas décadas, os dispositivos tecnológicos passaram a ser utilizados em conjunto com o corpo do performer, como por exemplo em Stelarc [Atzori and Woolford, 1995]. Nesta performance o artista utiliza diversos sensores atrelados em seu corpo com o objetivo de modificá-lo biologicamente, fazendo com que ele receba diferentes estímulos, os quais influenciam diretamente na performance realizada. Em Stelarc até mesmo o próprio dispositivo tecnológico pode ser considerado como um performer, visto que ele é totalmente responsável por gerar os estímulos, os quais podem ser caracterizados como parte de um ato performático.

Mais adiante, os dispositivos móveis passaram a se tornar ferramentas ubíquas, as quais passaram a estar cada vez mais presentes no cotidiano das pessoas. Devido à acessibilidade e aos diversos sensores acoplados a ele, os dispositivos móveis começaram a ser utilizados em performances digitais, nos primeiros anos do século XXI [Taylor, 2017]. Benjamin Taylor estuda a performance digital chamada Dialtones [Levin et al., 2001] como prova de conceito sob os potenciais musicais deste novo tipo de tecnologia, destacando algumas motivações expressas pelos compositores da época: a ubiquidade presente nos dispositivos móveis; a intenção de ilustrar o wireless em um espaço social; e a intenção de usar estes dispositivo como meio de criação / composição de arte.

Desde então, cada vez mais os dispositivos móveis, sensores, projetores, entre outros meios tecnológicos foram sendo utilizados, criando novos tipos de interações tanto em performances digitais quanto instalações tecnológicas, as quais serão explicadas na Subseção 2.4.

Em geral, a inserção destes aparatos tecnológicos no fazer artístico acabam vindo como "uma forma de traspor os limites físicos e possibilitar uma maior liberdade de alcance criativo" [Kraiser et al., 2011]. Desta forma, a tecnologia acabou permitindo ao criador da obra, não apenas utilizar objetos físicos em sua performance, mas também captar e processar qualquer evento externo e imprevisível que possa acontecer no ambiente, fazendo com o que os aparatos tecnológicos utilizados passem a ser uma espécie de "organismo vivo, mutável, dinâmico, que permite não um resultado único, mas uma variabilidade de criações e de propostas criadoras" [Melo et al., 2010] ou como uma espécie de performer, por também realizar atos performáticos ao captar, gerar ou processar algo em uma performance digital. Por fim, outra forma de se criar uma obra artística tecnológica / digital pode ser visualizada nas chamadas instalações digitais, as quais serão explicadas na Seção seguinte.

2.4 Instalações de Arte Digital

Logo no final do século XIX, o termo "instalações" passou a ser cada vez mais utilizado no meio artístico. Analisando as instalações artísticas como uma outra forma de arte, Mamedes et al. [2015] as define como "um gênero independente da escultura ou outras formas de arte, atribuindo importância ao espaço, manifesto através da percepção integral do espaço expositivo como uma única obra". Este trabalho (Mamedes et al. [2015]) será utilizado como base para descrição de alguns conceitos sobre instalações digitais, além de ser utilizado para elaboração de algumas métricas para avaliação destas formas de ambientes.

Por dependerem do espaço a ser empregado, as instalações acabam se tornando uma ótima forma para incluir a imersão e interação do participante com a obra artística desenvolvida. Até mesmo a atividade de outros visitantes no local ou a forma de apresentar determinadas mídias podem influenciar na experiência do participante, sendo que este fator pode ocorrer ao mesmo tempo, em diferentes níveis.

Segundo Reiss [2001] citado em Mamedes et al. [2015], em uma instalação:

> o artista trata todo um espaço interno (grande o suficiente para pessoas entrarem) como uma única situação, ao invés de [tratá-lo] como uma galeria para exibição de obras separadas. O espectador é de alguma forma considerado como integral para realização da obra. Embora o termo "instalação" tenha se tornado amplamente utilizado, ele ainda é relativamente inespecífico.

Por serem desenvolvidas em um espaço fechado e predeterminado, em relação à interação proporcionada ao público, pode se dizer que ela tende a ser mais intensa por ocorrer de maneira multissensorial, sob qualquer sentido humano, como: tato; paladar; olfato; audição; e visão. Estes sentidos humanos estão diretamente relacionados ao tipo de

feedback que uma performance / instalação digital pode prover ao público, o que justifica a presença do *feedback* no componente de saída da arquitetura de interação proposta no capítulo 3.

Diferente de esculturas ou pinturas, nas instalações digitais é possível empregar a participação do público, o que justifica o fato de que estes tipos de instalações são uma forma artística que se difere das demais. Em exposições artísticas, por exemplo, o processo de imersão do indivíduo na obra é dado, na maioria dos casos, de maneira visual. Já em instalações digitais, na maioria dos casos o participante é imerso em um ambiente fechado e controlado, o qual irá obter uma experiência que é dada de acordo com o sentido estético / conceitual e o objetivo artístico da obra, tornando possível a utilização de diferentes sentidos estéticos para diferentes níveis de interação.

Por fim, Mamedes et al. [2015] caracteriza as instalações sob três propósitos: percepção integral do espaço pelo visitante como uma única obra; experiência multissensorial pelo público; e engajamento e imersão do público. Todos estes propósitos são dependentes do nível de engajamento do participante com a obra, o que acaba sendo algo bastante individual, variando bastante de acordo com cada um. Ou seja, a experiência a ser obtida em uma instalação está relacionada à vontade e compromisso do participante durante toda a imersão, além do seu conhecimento prévio sobre o que está sendo apresentado.

Já Sogabe [2011] foca na descrição de cinco elementos físicos os quais são primordiais em instalações digitais: espaço, público, interfaces, gerenciador digital e dispositivos. Em relação ao espaço, antigamente o termo instalações interativas eram bastante ligadas à ideia de uma sala escura e fechada, bastante semelhante às sessões de cinema. Com o tempo, devido ao avanço tecnológico, este tipo de ambiente passou a ser uma opção e não uma condição obrigatória. Desta forma, o espaço para se criar uma instalação interativa passou a depender unicamente da intenção e da experiência artística a ser proporcionada, podendo ser dado sob: ambientes fechados, os quais geralmente possuem menos iluminação, além de limitar o espaço de interação; ambientes abertos, quando a experiência a ser proporcionada pela instalação vai além do simples uso das ferramentas e interações criadas exclusivamente para a instalação; ambientes divididos, quando há intenção de se criar ambientes distintos ou utilização de corredores para preparação dos participantes; ou até mesmo uma estrutura física qualquer com o intuito de "forçar" o participante a seguir uma determinada sequência de tarefas.

Já o público começa a ser estudado como uma entidade mais ativa, modificando a ideia de estudá-lo apenas como receptor. Desta forma, a experiência a ser obtida em uma instalação interativa acaba sendo bastante individual e dependente do engajamento de cada participante com a obra, como abordado por Mamedes et al. [2015], nesta Subseção.

Em relação às interfaces, este é um conceito bastante amplo que pode representar tanto as telas disponibilizadas para prover o contato do público com a aplicação desen-

volvida, quanto a própria aplicação, código ou até mesmo sensor utilizado. Sogabe [2011] estuda as interfaces como "o aparato físico que capta as ações do público na instalação, sendo a parte sensível do sistema tecnológico". As possibilidades de dispositivos a serem utilizados como interface serão abordados na Seção 3.1, a qual explica o componente de entrada da arquitetura de interação proposta.

As interfaces em uma instalação interativa operam sob um gerenciador digital, o qual geralmente é dado através de um microcontrolador e um programa. Desta forma, o gerenciador digital fica responsável por captar as informações coletadas pelos sensores, mapeá-las, decidindo qual ação deverá ser tomada, gerando uma determinada saída. Os possíveis dispositivos a serem utilizados para prover a interação do público serão abordados na Seção 3.1; já os diferentes tipos de *feedback* sob as interações serão abordados na Seção 3.2; e por fim, os conceitos sobre as possíveis formas de mapeamento serão explicados na Seção 3.3.

Além dos elementos físicos, Sogabe [2011] também cita três elementos que ocorrem nas instalações interativas durante o tempo: evento; interação; e processamento de informações. Neste contexto, chamamos de evento toda e qualquer experiência obtida pelo participante durante sua imersão em uma instalação interativa. Estes eventos ocorrem a todo momento e são independentes do que há no ambiente da obra. A maioria das instalações interativas operam sob imagem e som, porém, mesmo com a ausência de qualquer objeto interativo no ambiente, o participante irá passar por um determinado evento o qual irá acabar gerando determinada reflexão e consequentemente, uma nova forma de interação no ambiente.

Segundo Plaza [2003], citado em Sogabe [2011], estas interações podem ser dadas sob três níveis: primeiro grau, onde o público participa da obra sob sua interpretação individual e conhecimento artístico; segundo grau, onde o público não apenas interpreta a obra, mas participa fisicamente podendo escolher o rumo em que alguns eventos irão ocorrer; e terceiro grau, que se refere às obras interativas digitais, as quais são estudadas neste trabalho e necessitam da existência ou do contato com o público para que elas ocorram. Alguns exemplos de diferentes formas de interações em ambientes digitais performáticos serão apresentados na próxima seção (Seção 2.5).

Por fim, o processamento de informações pode ser definido como a forma em que o gerenciador digital opera, podendo processar os dados coletados de diversas maneiras, através do processamento de imagem e de som, por exemplo.

2.5 Formas de Interação do Público em performances digitais

A ascensão das performances digitais possibilitaram diversas inovações no que diz respeito às possibilidades da utilização de ferramentas/instrumentos inovadores e tecnoló-

gicos, além do surgimento de novas formas de interação do performer com o público. Uma das possibilidades que o avanço da tecnologia trouxe para a performance artística é o de envolver o público por meio de dispositivos móveis, como uma interface mediadora em performances e composições musicais, por exemplo [Clinch, 2013]. Hoje em dia, os dispositivos móveis se tornaram ferramentas que fornecem diversas possibilidades de interação, além de ser um dispositivo ubíquo na vida das pessoas, justificando a sua utilização em espetáculos de arte, envolvendo cada vez mais o público nas performances digitais.

Diversas iniciativas têm utilizado os dispositivos móveis como mediadores para a participação do público em performances artísticas e há diversas possibilidades de design para permitir uma experiência de participação do público-alvo baseada em dispositivos móveis [Egozy and Lee, 2018]. No trabalho "*12*" composto pelo grupo Radius temos a caracterização de um papel envolvente, individualizado e influente na performance da música ao vivo, por parte do público [Egozy and Lee, 2018].

Outra iniciativa para promover a interação entre o público via dispositivos móveis pode ser visualizada em SWARMED [Hindle, 2013]. Neste trabalho, vale mencionar sobre os aspectos técnicos que os autores utilizaram para a aplicação criada. Os autores usaram uma rede local de um roteador Wi-Fi popular, encarregado de direcionar todo o tráfego para um *laptop*, além de ter um *captive portal*[1] para redirecionar os usuários para o aplicativo. O *laptop* usado atuava como *gateway* da Internet e também era responsável por atuar como um *Domain Name System* (DNS), reescrevendo todas as solicitações de DNS de volta para o próprio *laptop*, além de atuar como servidor através de um *Dynamic Host Configuration Protocol* (DHCP), atribuindo *Internet Protocols* (IP's) a todos os participantes. Ao todo, sete Instrumentos Musicais Digitais (IMD's) distintos foram desenvolvidos e o sistema operou com uma média de 20 dispositivos móveis, simultaneamente. Em geral, a dificuldade dos participantes para se conectarem à rede fornecida foi o único problema obtido, o qual poderia ser tratado a partir de instruções simples impressas em um panfleto.

Outra performance que utiliza dispositivos móveis como IMD, porém não apenas para a audiência, pode ser visualizada em Echobo [Lee and Freeman, 2013]. Este trabalho visou performances musicais de grande escala e se baseou no uso de dois tipos de instrumentos implementados em um aplicativo de dispositivo móvel: um para o maestro e outro para o público. Em Echobo, o maestro fica responsável por definir a progressão dos acordes e controlar a harmonia dos sons produzidos pelo público. E então o público, a partir da restrição harmônica estabelecida pelo condutor, pode tocar nota por nota dentro do campo harmônico definido. A audiência mencionada foi entre 20 e 120 participantes e o *feedback* recebido foi que, ao usar o Echobo, o público se sentiu mais imerso e mais

[1] Recurso que ajuda a administrar a entrada de usuários em uma rede sem fio de um ambiente público ou privado.

conectado à música e aos outros músicos. Ou seja, a criação de instrumentos com interfaces de fácil utilização parece ser uma ideia muito promissora em relação à satisfação do público ao utilizar o IMD criado.

A partir de um servidor em nuvem, com a utilização do Web Audio[2], Junior et al. [2016] desenvolveu um IMD a partir de uma página web (interface) focada em música colaborativa, que funciona em rede, parecendo ser bem suscetível em relação à performances de grande escala. O trabalho é baseado em uma conexão de três vias, sendo elas: humana; via instrumento musical e servidor em nuvem. Desta forma, o instrumento fica responsável pela comunicação com o servidor, permitindo a interação social entre o público, a partir do envio de mensagens.

Por fim, em relação à utilização de dispositivos móveis como IMD's, temos um estudo geral sobre o comportamento do público em relação à sua familiaridade técnica e musical com o uso de um IMD específico em uma performance musical [Bin et al., 2016]. Por meio do trabalho desenvolvido, é possível afirmar que o nível de interesse do público em uma performance, em geral, não seja afetado de acordo com o conhecimento de cada participante sobre o uso de IMD's. Ou seja, explicar os aspectos técnicos de um IMD para todos os participantes de uma performance pode ser uma ação exaustiva e desnecessária.

A utilização de dispositivos móveis como alto-falantes espalhados para se produzir música tem sido também uma opção para envolver o público na performance. Porém, é possível que um servidor seja utilizado para o processamento do áudio produzido, sendo assim, controlado pela audiência em tempo real. Gimenes et al. [2016] apresentam a performance "Without Borders", a qual o público fica encarregado por alterar alguns parâmetros do mecanismo de áudio em tempo real e utilizar seus dispositivos móveis como alto-falantes para fornecer suporte à todo o som produzido.

Em Sense of Space [Hirabayashi and Eshima, 2015] é demonstrada uma forma de interação do público via iPads ou iPhones. Este trabalho utiliza um método próprio para comunicação, baseado em um *Dual Tone Multiple Frequency* (DMTF) com alta frequência nomeado de *Ultra Sound Communication* (USC), que por fim é utilizado como um identificador na escolha dos sons que o público poderá tocar. A performance acontece com os performers utilizando um sistema de *Public Audition* (PA)[3] principal para tocar música e junto deles o público utiliza o alto-falante dos dispositivos móveis, o que possibilita ao público, sentir o som vindo de diferentes locais, dentro do espaço definido.

Em performances digitais, os dispositivos móveis também podem ser utilizados de maneira em que não sejam IMD's tocados pelo público. A performance TweetDreams

[2] A Web Audio é uma API desenvolvida em JavaScript de alto nível para processar e sintetizar áudio em aplicativos da web.

[3] Sistema de endereço público é um sistema eletrônico que compreende microfones, amplificadores, alto-falantes e equipamentos relacionados.

[Dahl et al., 2011] traz uma abordagem diferente, em que os dispositivos móveis são utilizados para gerar Tweets realizados pela audiência durante a performance. Estes dados são coletados e então agrupados em gráficos de tweets relacionados, onde os tweets associados recebem melodias semelhantes. Esse tipo de abordagem acaba gerando música de acordo com a rede de relacionamentos que um tweet em particular tem com outros tweets, o que escapa da ideia de apenas criar música a partir do sentido léxico de cada um, gerando música também pelo sentimento de cada um. A opinião do público em relação à uma performance digital é um aspecto importante, e os dispositivos móveis também podem ser utilizados para transformar estas opiniões em música.

Zhang et al. [2016] desenvolveu uma aplicação web para permitir a contribuição/colaboração entre a audiência que engloba um sistema de votação que permite ao público optar por determinados atributos musicais predeterminados, através de seu voto. Assim, o sistema fornece *feedback* a todos os participantes na forma de visualizações que são tratadas pelos artistas, que, por sua vez, geram um certo tipo de som para o público. A audiência estudada foi de 13 participantes e o aplicativo foi testado sob 9 aspectos: facilidade de uso do aplicativo; satisfação do usuário; facilidade de entendimento; engajamento; interação proporcionada; utilidade da interface; simplicidade da interface; influência pessoal e a partir de outros participantes na performance. No geral, o público se mostrou muito satisfeito com o uso da aplicação, mas insatisfeito com a influência proporcionada por eles.

Já no trabalho chamado massMobile [Weitzner et al., 2012] era possível ao público votar utilizando seus dispositivos móveis para modificar determinados aspectos da iluminação e consequentemente influenciando a performance dos dançarinos. Um aspecto importante relatado neste trabalho foi que pequenos grupos começaram a se formar, respondendo aos votos uns dos outros, formando assim "líderes" que modificavam seus votos, influenciando seus "seguidores" e realizando uma mudança / interação coletiva.

A participação do público em performances digitais não é limitada ao uso de dispositivos móveis, também é possível utilizar sensores para este fim. Tentando fugir da ideia de se ouvir música de maneira unidirecional, assim como a música eletrônica consumida em clubes, tocada a partir do *disc jockey* (DJ) ou produtor para o público, Hout et al. [2014] desenvolveu um trabalho chamado Experio, sob o conceito de dança interativa, aplicada em performances ao vivo. O trabalho tem três objetivos principais: permitir a participação de pessoas com habilidades musicais limitadas; permitir a exploração, expressividade e improvisação; incentivar a criação entre o moderador e o público. Em geral, o Experio permite que o público dance sob várias áreas designadas, marcadas e iluminadas por raios laser. Assim, este laser acaba sendo constantemente interrompido em diversas áreas, modificando o som produzido pelo moderador musical, que utiliza uma interface para controlar determinados parâmetros.

Outra forma de interação que foge da ideia da utilização de dispositivos móveis como instrumento musical digital é apresentada por Turchet and Barthet [2018]. O autor apresenta um sistema de comunicação bidirecional no qual o intérprete se encarrega de tocar um IMD, chamado de Smart Mandolin, que influencia e é influenciado diretamente pelo público. Já o público, utiliza *Musical Haptic Wearables* (MHW's) em seus braços, produzindo sons a partir de um aplicativo desenvolvido através do Pure Data [4], a qual sintetiza estímulos táteis por meio de técnicas de modulação por largura de pulso.

Até mesmo as ondas cerebrais de um participante podem influenciar no som reproduzido em uma performance. Eaton et al. [2014] desenvolveu um trabalho baseado na eletroencefalografia, o qual capta ondas cerebrais através de um capacete adaptado com eletrodos e durante a apresentação de uma peça voz e piano, as emoções dos *performers* e do público são lidas e gravadas para permitir que elas selecionem as próximas frases musicais, que por sua vez, foram pré-produzidas para a notação voz e piano.

Até então foram abordados exemplos em performances digitais as quais não são necessariamente musicais. As performances musicais podem ocorrer em diversos locais, desde uma roda de viola entre amigos até grandes concertos executados por orquestras. Em performances musicais ao vivo, a interação do público com os artistas geralmente se dá por meio de aplausos, com a finalidade de expressar seu respectivo engajamento com a mesma, ou até mesmo através da solicitação de suas músicas preferidas para que elas sejam tocadas pelos artistas. Baseando-se neste conceito, Lins and Novaes [2017] implementou um aplicativo para facilitar esta forma de interação tornando possível, ao público, solicitar músicas aos artistas em tempo real, além de possibilitar ao artista visualizar quais são as músicas mais votadas em determinado tempo de sua performance. A partir do aplicativo desenvolvido, o autor desenvolveu 3 casos de teste para avaliar a interação do público e em geral, o uso do aplicativo mostrou-se bastante promissor, caso aplicado em bares. O aplicativo também facilitou a interação do público com o artista, do público com o público e serviu também para que o artista soubesse o perfil das pessoas que estavam ouvindo sua performance, em tempo real. Lins and Novaes [2017] ainda cita aplicativos como o Radiola Digital[5] que foram desenvolvidos para serem utilizados em bares, porém o artista acaba não entrando como usuário e o dono do bar acaba tomando a posição de maestro, controlando quais músicas devem ou não ser tocadas.

2.6 Avaliação da Participação do Público em performances digitais

Na Subseção anterior foram apresentadas diversas formas tecnológicas para se prover a interação do público em performances e instalações digitais. Para avaliar a ex-

[4] O Pure Data é uma linguagem de programação visual voltada para criação de música eletrônica, música electroacústica, música interativa e trabalhos multimídia.

[5] http://www.radioladigital.com.br

periência relacionada à interação do público, além do som reproduzido nestes ambientes, neste trabalho serão utilizados as métricas propostas por Mazzanti et al. [2014] e Mamedes et al. [2015].

Mazzanti et al. [2014] cita algumas métricas para avaliar plataformas voltadas para performances participativas:

- **Liberdade de controle** (*Control Design Freedom*), que descreve o quão livre a interação do público pode ser proporcionada na plataforma;
- **Versatilidade do sistema** (*System Versatility*) referente à uma visão geral sob o quão simples é a performance e o conforto dos artistas no palco;
- **Transparência na interação do público** (*Audience Interaction Transparency*) referente a clareza da relação entre a manipulação da audiência e seus efeitos;
- **Distribuição da interação do público** (*Audience Interaction Distribution*) que é o quanto a interação pode ser localizada em relação aos participantes (interface fortemente centralizada vs. cada participante utilizando uma interface);
- **Foco** (*Focus*) em relação a quão facilmente o público pode se concentrar livremente em diferentes aspectos da performance (o palco, sua interação, visuais, música, etc.);
- **Afinidade ativa / passiva do público** (*Active/Passive Audience Affinity*) se referindo à diferença entre os participantes ativos e passivos.

Porém, de acordo com as peformances apresentadas neste capítulo, nota-se que a participação do público passa a envolver características que vão além das apresentadas por Mazzanti et al. [2014], as quais são focadas apenas na plataforma ou aplicação utilizadas nestes ambientes.

Por ter relações com a participação do público, a maneira como o som é "criado" e reproduzido em performances digitais também deve ser levada em consideração. Surge então um conceito importante, o qual será utilizado como uma métrica de avaliação de performances digitais chamado design sonoro. O design sonoro possui uma definição muito ampla e começou a ser utilizado no cinema, com o objetivo de caracterizar a concepção que um artista utiliza ao desenvolver a direção do ambiente sonoro de um filme.

Em relação ao sentido industrial, o design sonoro pode ser definido como a interação entre um produto a ser projetado e desenvolvido, as condições para sua produção e sua utilidade prática [Bürdek, 2005]. Este tipo de definição acaba englobando aspectos tanto criativos quanto técnicos durante o desenvolvimento de um produto, fazendo com o que o design esteja relacionado à "um projeto de execução, o rascunho / desenho que representa o objeto do projeto, a interação do objeto com o usuário e o ambiente e, por

fim, as características de estilo que o diferenciam dos produtos similares" [Mamedes et al., 2015].

Desta forma, Mamedes et al. [2015] acaba utilizando o conceito de design no sentido industrial, definindo o design sonoro como "parte de um processo criativo que procura desenhar o som de acordo com o projeto conceitual da obra em questão, otimizando suas relações referenciais e as relações de unidade estabelecidas, tanto da mídia sonora e sua respectiva articulação expressiva em relação as demais mídias presentes na obra, quanto da mídia sonora e seu sentido próprio dentro do projeto artístico". A partir desta definição, o autor apresenta 4 métricas para avaliação do **design sonoro** em um ambiente performático / artístico:

- Objetivo de proporcionar uma experiência estética considerando o ponto de vista do usuário;
- Adequabilidade de condições técnicas para sua realização;
- Clareza do projeto;
- Validade, ou sua utilidade prática e efetiva.

Estas métricas foram utilizadas para desenvolver o design sonoro das instalações "Abstrações", "Cerejeira" e "Caminho das águas" [Mamedes et al., 2015]. Apesar do trabalho ter sido focado para estudar e mapear os gestos dos participantes, a definição do design sonoro em cada instalação ajudou o autor a compreender melhor a evolução dos materiais sonoros utilizados em cada etapa da experiência fornecida nas instalações, além de observar a relação entre os objetivos conceituais, estéticos, e as soluções que foram implementadas de acordo com a disponibilidade de recursos.

2.7 Recuperação de informação musical

O som pode ser definido como: um som musical, que corresponde à pressões acústicas periódicas e frequências múltiplas, as quais são executadas com um propósito e uma série de regras; um ruído, o qual não possui periodicidade bem definida, além de conter diversas frequências em seu espectro; ou até mesmo uma palavra dita por alguém, que consiste na mistura de sons ordenados que apresentam periodicidade com diversas frequências vizinhas.

Para que o computador interprete determinado som, é necessário que ele seja processado. Este é um subcampo da área de processamento de sinais voltado para a manipulação eletrônica de sinais de áudio, englobando diversas formas de representações eletrônicas e suas possíveis aplicações sob ondas sonoras [Paolo Prandoni, 2008].

Dentre os fundamentos do processamento de áudio, está a intensidade do som no ar, que pode ser alterada através de: modulação no áudio; compressão, que reduz a faixa dinâmica de ganho do material; densidade, responsável pela amplitude de um sinal de áudio; e muitos outros, como equalização, limitador de picos e compressão multi-bandas[Zölzer, 2008, Hodgson, 2010].

A partir do processamento das ondas sonoras é possível "ensinar" o computador, através de técnicas de aprendizado de máquina (Seção 2.8), a distinguir o som reproduzido por um instrumento ou por uma pessoa.

Atualmente, o avanço tecnológico tem possibilitado a criação de trabalhos que envolvem a extração e análise de informação musical a partir do processamento de áudio, utilizando o computador como ferramenta para tal fim. Esta análise musical envolve a área conhecida como MIR, a qual tem por o objetivo alcançar um alto nível de entendimento sobre o comportamento e as características do som. As pesquisas em MIR envolvem análise, segmentação e classificação das características do som a partir de seu processamento.

O processamento de áudio é uma tarefa um pouco complexa, visto que o áudio é armazenado no domínio do tempo, o que acaba trazendo pouca ou nenhuma informação sobre seu conteúdo para além da sua respectiva amplitude e envelope. Além disso, o áudio digitalizado possui alta dimensionalidade, por exemplo, um áudio de apenas um minuto, amostrado a 44100Hz possui aproximadamente 2 milhões e meio de amostras para serem processadas, o que acaba tornando necessário a aplicação de alguma técnica para realçar algum tipo de característica específica, a qual é extremamente importante de ser definida para a solução de um determinado problema.

2.7.1 Descritores de áudio

Após a digitalização do áudio, podemos utilizar os dados coletados para extrair suas características de acordo com o objetivo da aplicação, como por exemplo detectar padrões para classificar o timbre, o ritmo, o instrumento ou até mesmo qual palavra foi dita naquela amostra. A extração das características são dadas através dos descritores de áudio. Estes descritores são ferramentas analíticas que representam características de um sinal sonoro em curvas unidimensionais, reduzindo a complexidade da informação ao focar em aspectos específicos.

Em suma, os descritores apresentam caráter reducionista, além de poderem ser correlacionados com atributos subjetivos da percepção, como "brilho", "opacidade" ou ainda "maciez" do som [Simurra, 2015]. Pires [2011] os caracteriza como: descritores timbrísticos temporais, através do centróide temporal; descritores timbrísticos espectrais, através do desvio de espectro harmônico; e os descritores de parâmetros de sinais, através da

frequência fundamental e da harmonicidade.

Estes descritores podem auxiliar a detectar características como frequência fundamental, energia, ganho, brilho, silêncio, ruído entre outras, além de possibilitar a criação de aplicações que envolvem o reconhecimento automático de gêneros musicais, sugestão musical automática, transcrição automática de partitura, identificação automática de instrumentos ou de músicos, reconhecimento de harmonia, entre outras [Magalhães, 2015, Madison et al., 2011]. Eles podem atuar no domínio do tempo, também chamados de descritores básicos, envolvendo o processamento dos sinais de áudio amostrados, ou o domínio da frequência, utilizando descritores espectrais, como por exemplo o *Mel Frequency Cepstral Coefficients* (MFCC) [Hamel and Eck, 2010], melhor explicado na Subseção 2.7.2.

Estes descritores ficam responsáveis por reduzir o áudio à uma estrutura simplificada, sem perda de informações, as quais podem ser definidas de acordo com o tipo de descritor utilizado.

O resultado de uma análise por descritores de áudio pode retornar o quão brilhoso, suave ou rugoso é uma determinada nota ou trecho musical mas este valor nem sempre consegue ser parametrizado para sua avaliação. Por isto, normalmente a avaliação de uma determinada interpretação musical é feita com a comparação da mesma com sua escrita simbólica, com a performance de outros instrumentistas ou com uma performance anterior do mesmo instrumentista sendo que a escolha de qual performance seria usada de base de comparação implica em aplicar critérios subjetivos sobre a própria performance musical.

Por serem implementados com convoluções, na forma de equações matemáticas como a *Discrete Fourier Transform* (DFT) e a *Fast Fourier Transform* (FFT), os descritores de áudio acabam fazendo parte da etapa de pré-processamento dos dados, seguidos da aplicação de algum algoritmo de aprendizado de máquina[6]. Estes algoritmos podem ser: supervisionados, onde a máquina é treinada a partir de exemplos já classificados; ou não supervisionados, onde a máquina processa a base de dados e os classifica por ela mesma. Em geral, estes algoritmos são capazes de realizar classificações e agrupamentos dos dados de maneira automatizada, podendo em alguns casos, realizar todo o processo em tempo real. A Seção 2.8 trará mais informações sobre a área de aprendizado de máquina

Atualmente, existem diversas pesquisas em relação ao desenvolvimento de ferramentas tecnológicas para a extração de descritores de áudio, tanto na forma de bibliotecas para programação quanto softwares gráficos ou extensões para ferramentas existentes voltadas para extração e visualização das características de um áudio. Dentre as ferramentas existentes, temos: as bibliotecas Aubio [Brossier, 2006] e Librosa [McFee et al., 2015]; frameworks como jAudio [McKay et al., 2005], CLAM Music Annotator [Amatriain et al., 2005], jMIR [McKay and Fujinaga, 2009] e Sonic Visualiser [Cannam et al., 2010]; e a

[6] Ramo da IA baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana.

toolbox para MatLab MIRtoolBox [Lartillot et al., 2008].

2.7.2 Mel Frequency Cepstral Coefficients - MFCC

Um dos descritores de áudio mais utilizados para extração de características do áudio, a nível de frames, baseando-se no domínio da frequência, está o Mel Frequency Cepstral Coefficients [Hamel and Eck, 2010]. Por ser bastante eficaz para detecção de similaridade timbrística, o MFCC tem sido utilizado em diversos trabalhos [Pachet and Aucouturier, 2004, Foote, 1997, Hoffman et al., 2008, Levy and Sandler, 2006, Logan and Salomon, 2001, Mandel, 2010, Mandel and Ellis, 2005, Pye, 2000, Seyerlehner et al., 2008].

O cálculo do MFCC é dado sob 4 passos[Pampalk, 2006]: cálculo do espectro de energia; aplicação da mudança de escala para frequências Mel[7]; segunda mudança na escala para decibéis; e o cálculo da transformada discreta do cosseno. Já Logan et al. [2000], citado por Fachini and Heinen [2016], caracteriza o processamento do MFCC em 5 etapas: janelamento; aplicação da DFT; aplicação de um filtro sob a amplitude resultante da DFT, a partir de janelas triangulares na escala Mel; aplicação de um logaritmo sob o resultado da etapa anterior; por fim, a aplicação da Transformada Discreta de Cosseno, resultando nas amplitudes como Coeficientes Mel-Cepstrais.

Levando em consideração as etapas descritas por Logan et al. [2000], temos a primeira etapa referente à fase de janelamento, que é basicamente a segmentação dos sinais de áudio em quadros. Na literatura existem alguns tipos de técnicas de janelamento para janelas fixas, como a Janela de Bartlett, Janela de Hanning, Janela de Blackman e Janela de Hamming. Fachini and Heinen [2016] afirma que o Janelamento de Hamming é bastante indicado para tarefas de processamento de sinais de áudio, e além disso, devido à sua facilidade de projeção na fase linear e sua garantia de estabilidade, visto que não há realimentação da saída para a entrada, este método foi escolhido para o desenvolvimento deste trabalho.

A primeira tarefa para processar os coeficientes Mel é segmentar o sinal em quadros de mesmo tamanho, através da função de janelamento *w(n)*:

$$w(n) = \begin{Bmatrix} 0,54 - 0,46 * cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0, & out \end{Bmatrix}$$

Onde n é o número de um quadro qualquer do sinal de áudio e N é o tamanho das janelas. De acordo com o valor obtido em *w(n)*, a janela de Hamming pode ser calculada segundo a equação:

$$x_j(n) = x(n)w(n)$$

[7] As frequências Mel são uma escala que visa imitar as características únicas perceptíveis pelo ouvido humano.

Onde cada sinal de entrada *x(n)* é multiplicado ao resultado de sua respectiva janela *w(n)*, fornecendo resultados para elaborar o gráfico da quantidade de amostras por amplitude, referentes ao janelamento de Hamming (Figura 4).

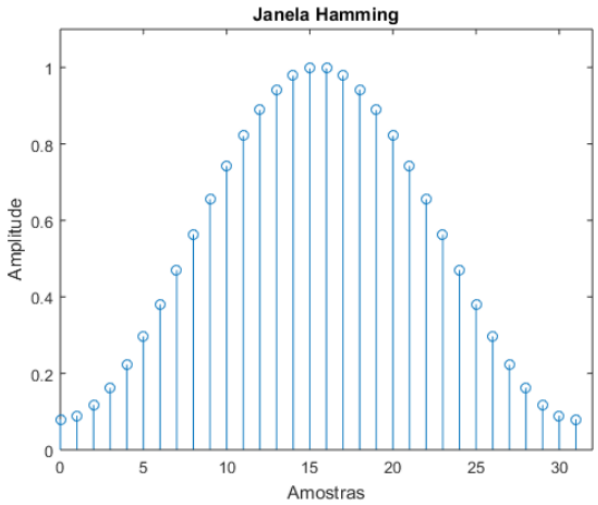

Figura 4: Exemplo do Janelamento de Hamming sob um áudio digitalizado

As características a serem coletadas a partir do MFCC geralmente são extraídas quadro a quadro, o que acaba não solucionando o problema da alta dimensionalidade, visto que para um quadro de tamanho 1024, aplicado em um áudio de 1 minuto amostrado a 44100Hz, teremos mais de 2 mil características que ainda devem ser multiplicadas pela dimensão da característica local. Porém, um aspecto interessante do MFCC é que durante suas etapas de processamento, um espectro de 1024 pontos pode ser reduzido de 15 a 40 pontos que podem ser utilizados para verificar a similaridade ou a distinção entre dois sons [Brent, 2010]. Vale ressaltar que a escolha da quantidade de pontos pode variar, mas geralmente utiliza-se o valor mencionado entre 15 e 40 pontos.

A tarefa de janelamento dos sinais de áudio é seguida então da aplicação da DFT, para transformar um determinado sinal em seu respectivo coeficiente espectral. Desta forma, a equação a se utilizar é:

$$F(n) = \sum_{k=0}^{N-1} f(k)e^{-i2\pi\frac{kn}{N}}$$

Onde N é o tamanho do sinal de entrada, o qual é representado pela função *f(k)*, $e^{-i2\pi}$ é a função base que define para cada ponto de *F(n)* no domínio da frequência, seus respectivos valores complexos, à partir do número complexo "i".

Como resultado, temos então a função *F(n)* que representa os coeficientes espectrais do som, em Hertz, os quais podem ser utilizados para diversas finalidades, como por exemplo para detecção do timbre de um instrumento, a partir da análise de seu espectro.

2.8 Processamento de áudio via Aprendizado de Máquina

Em geral, como mencionado anteriormente, a área de aprendizado de máquina é um subcampo da área de Inteligência Artificial (IA) e lida com diversos algoritmos, tanto estatísticos quanto de mineração de dados, focados na extração de características e informações a partir de uma determinada base de dados. Ou seja, a partir de uma certa quantidade de dados passou a ser possível treinar / ensinar o computador a classificar algo ou até mesmo a gerar um novo dado automaticamente. Desta forma, bases de dados cada vez maiores foram surgindo e sendo disponibilizadas para toda comunidade, como por exemplo bibliotecas de MIDI contendo diversas performances artísticas realizadas ao redor do mundo.

Até então, eram utilizados apenas dados simbólicos sob algoritmos de Regras de Produção ou Árvores de decisão, como o CN2 e o C4.5, citados em Martineli [1999]. Uma das vantagens de se utilizar estas técnicas é que todo o conhecimento processado e gerado é algo compreensível para o ser humano onde os dados utilizados podem ser simbólicos ou contínuos.

Com o avanço nas pesquisas na área de aprendizado de máquina, além do processamento de dados simbólicos, diversos áudios, imagens ou filmes passaram a ser também processados, ampliando ainda mais as possibilidades de criações artísticas. Consequentemente, bases de dados mais complexas começaram a surgir, englobando, por exemplo, imagens históricas, artigos científicos, gravações de performances artísticas, entre outros tipos de dados. Ou seja, além de dados simbólicos como o MIDI, também foram anexados dados contínuos e multidimensionais. Desta forma, os dados passaram a ser armazenados sob diversos formatos como por exemplo MP3, *Audio Interchange File Format* (AIFF) e *Waveform Audio File Format* (WAV) para áudio e RAW, *Joint Photographic Experts Group* (JPEG) e *Device Independent Bitmap* (BMP) para imagens.

Uma das técnicas que tem sido bastante utilizadas para a criação de arte é dada através da utilização de redes neurais. As redes neurais são modelos computacionais com nós interconectados que funcionam de forma baseada nos neurônios do cérebro humano, sendo capazes de realizar o aprendizado de máquina bem como o reconhecimento de padrões. Atualmente, na literatura existem diversos tipos de redes neurais, como por exemplo as *Generative Adversarial Networks* (GAN's) e as *Convolutional Neural Networks* (CNN's) e na área da arte digital, estes tipos de redes neurais têm sido mais utilizados como gerador de conteúdo do que como classificadores automáticos.

Além da complexidade dos dados, inicialmente a entrada das redes neurais ainda era unidimensional, o que impulsionou diversas pesquisas sobre formas eficientes para redução do espaço de representação dos dados sem que haja perda de informação. Estas pesquisas deram origem à diversas técnicas de pré-processamento de dados, como a convo-

lução, extratores de características, filtros e descritores de áudio, fornecendo suporte para a ascensão da área multi-disciplinar chamada de MIR que envolve áreas como aprendizado de máquina e computação musical.

Atualmente, em relação ao áudio como dado a ser processado, existem diversas bases de dados para diversos fins, como por exemplo bases de dados: para tarefas de reconhecimento de voz como o AudioSet[8] e Common Voice[9]; musicais como Million Song Dataset[10] e BallRoom[11]; ou até mesmo mais específicos, voltados para classificação de timbres de instrumentos[12].

2.8.1 Redes Neurais Convolucionais

Por volta de 1988, inspirados em processos biológicos [Matsugu et al., 2003], Yann LeCun et al. [LeCun et al., 1998] criaram as chamadas CNN's. Inicialmente, a proposta era de que estas redes fossem voltadas para o processamento de estruturas de dados unidimensionais ou bidimensionais, tornando as CNN's bastante promissoras para o processamento de áudio e imagem. As imagens são compostas por *pixels* que formam uma estrutura 2D porém o áudio pode ser tratado tanto como uma estrutura unidimensional, se analisarmos as variações das ondas sonoras no tempo, ou bidimensional, se analisarmos seu respectivo espectrograma, sendo tratado de maneira semelhante à uma imagem.

Uma CNN é basicamente uma variação das redes de Perceptrons[13] de Múltiplas Camadas, porém voltada para um tipo de processamento bastante semelhante ao de visão computacional, possibilitando a aplicação de filtros sob o dado a ser processado, mantendo a relação entre seus respectivos *pixels* [Vargas et al., 2016].

Em geral, este tipo de rede neural tem sido bastante utilizado em diversos tipos de aplicações como: reconhecimento de imagens de instrumentos musicais [Kuhn et al., 2019]; reconhecimento de fala, seja ela em aplicativos móveis ou até mesmo em jogos [Torres et al., 2016]; classificação de emoções em sistemas de recomendação de música [Lopes et al., 2018]; reconhecimento de produtos [Juraszek et al., 2014]; detecção da presença de pedestres [Vargas et al., 2016], e quaisquer outras aplicações as quais podem automatizar processos a partir do reconhecimento de imagem.

Em relação ao funcionamento de uma CNN, como o próprio nome já diz, elas operam principalmente sob uma equação matemática chamada de convolução:

[8] https://research.google.com/audioset
[9] https://voice.mozilla.org
[10] http://millionsongdataset.com
[11] http://mtg.upf.edu/ismir2004/contest/tempoContest/node5.html
[12] https://magenta.tensorflow.org/datasets/nsynth
[13] Cada Perceptron é um modelo matemático que recebe determinadas entradas e produz uma única saída binária.

$$s(t) = (x * w)(t) = \sum_a x(a)w(t-a)$$

A convolução é um operador linear que se baseia no cálculo de duas funções: a função *kernel*, também chamada de *filter*, é descrita como w e se equivale a uma *matrix* estática a qual é aplicada diversas vezes em diferentes regiões da imagem; e a soma dos produtos gerados pela função x, em cada região da imagem / dado processado.

Em relação à arquitetura de uma CNN, elas podem ser classificadas como uma junção de diversas camadas, as quais podem ser classificadas como: Camadas convolucionais; camadas de *Pooling*; e as camadas totalmente conectadas [Guimarães, 2018].

As camadas convolucionais são compostas por diversos neurônios[14], os quais são conectados, segundo um determinado peso, a um conjunto de *pixels* fornecidos pela camada anterior. Estes neurônios são responsáveis por aplicar filtros sob os *pixels* captados através de funções *kernel*, e o resultado deste processamento é então propagado para a camada seguinte.

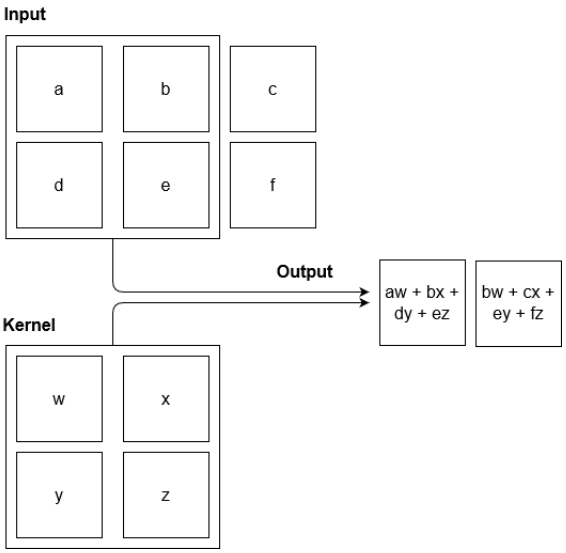

Figura 5: Exemplo da função do filtro durante uma convolução em uma CNN.

A figura 5 mostra o exemplo de um filtro a ser aplicado. Nesta etapa, o *kernel* e a entrada fornecida são convoluídos através do deslizamento da janela que neste caso, irá se deslocar duas vezes, gerando duas saídas. É importante mencionar sobre a variável *stride* que diz respeito a quantos *pixels* serão pulados, visto que nem sempre a janela irá se mover de *pixel* em *pixel*, gerando uma entrada de tamanho 2 para a próxima camada, como demonstrado na figura 5. Além disso, nem sempre é possível às janelas, convoluírem até o final. Para isso utiliza-se o *zero-padding*, que é uma técnica que se baseia na adição de zeros nas bordas, possibilitando a convolução sob todos os *pixels*. Após a convolução,

[14] Um neurônio de uma rede neural é um componente que calcula a soma ponderada de várias entradas, aplica uma função e passa o resultado adiante.

é comum a aplicação de uma função de ativação não-linear, como as utilizadas em redes neurais comuns.

Um aspecto interessante sobre as CNN's é a possibilidade de se utilizar filtros mais sofisticados, se comparados aos clássicos utilizados em processamento de imagens, que geralmente são bidimensionais. Ou seja, é possível aplicar um filtro que possui altura, largura e até mesmo profundidade, otimizando a extração de informações fornecidas pelos dados de treinamento.

Logo após a convolução em conjunto com a função de ativação, temos a camada de *pooling*, responsável por diminuir a dimensão dos dados. Por exemplo, dada uma entrada com janelas bidimensionais, um valor é selecionado para representá-la através de uma função, que geralmente é a função chamada de *maxpooling* a qual seleciona o valor máximo dentro de uma vizinhança retangular.

Atualmente existem outras arquiteturas mais sofisticadas como a LeNet-5 [LeCun et al., 1998] que possui duas camadas de convolução seguidas de *pooling* e mais uma camada de convolução ou a GoogLeNet [Szegedy et al., 2015] que possui cinco camadas de convolução seguidas de *pooling*, ambas citadas em Vargas et al. [2016].

Por fim, após a camada de *pooling*, para problemas de classificação, geralmente utiliza-se mais uma camada totalmente conectada. Esta camada fica responsável por realizar as decisões que a rede irá tomar, dada as saídas recebidas pelas camadas anteriores. Nesta camada o *score* para cada situação é gerado, sendo utilizado para calcular o resultado final de cada classificação.

2.8.2 Aprendizado de Máquina x Arte Digital

Por volta dos anos de 1970 e 1990, Myron Krueger realizou diversos estudos na área de realidade aumentada[15] criando diversas instalações digitais, as quais utilizavam técnicas computacionais para a criação artística, reforçando ainda mais a relação do computador com a arte [Kogan, 2016]. Tais aplicações acabaram inspirando toda uma geração de artistas / cientistas que passaram a enxergar o computador como uma ferramenta bastante promissora para a criação de arte.

Consequentemente, a área de aprendizado de máquina foi tomando cada vez mais espaço, englobando cada vez mais o uso de diferentes tecnologias voltadas para criação de um novo tipo de arte, chamada de arte digital. Além do surgimento de diversas discussões entre pesquisadores e artistas sobre este modo de arte, a arte digital tem estado cada vez mais presente em diversos locais como Museus (Uncanny Mirror, Museu de Arte Moderna e Barbican Centre), Casas de Leilão (Christie's e Sotheby's) ou até mesmo em

[15] Designa-se realidade aumentada a visualizações do mundo real através de uma câmera e com o uso de sensores de movimento como giroscópio e acelerômetro.

determinados eventos científicos.

Uma obra de bastante sucesso, que une a utilização de redes neurais no mundo artístico é a "ToTa Machina", desenvolvida por Katia Wille[16]. Esta instalação capta as emoções do público por meio de técnicas de reconhecimento facial e, através da robótica, conta com projeções visuais que se movimentam de acordo com o público. Através dos estímulos faciais, a relação do ser humano com a máquina acaba ficando ainda mais próxima, reforçando ainda mais o engajamento e a apreciação da obra [Calais, 2020]. Um fato interessante que reforça ainda mais a promissão de instalações digitais, é que durante as apresentações da obra, o público permaneceu bastante tempo na instalação, onde as crianças ficavam curiosas para saber mais sobre a estrutura da instalação e os adultos interagiam criando diferentes afeições para obtenção de diferentes resultados.

Um outro exemplo mais atual de redes neurais voltadas para produção artística é a obra "Edmond de Belamy", criada em 2018 pelo grupo Obvious, formado por três pesquisadores da Universidade de Montreal que utilizaram uma GAN para elaboração de seu trabalho [Kaufman, 2019]. A obra é composta por um conjunto de 11 retratos gerados pela rede neural, retratando membros da família fictícia Belamy. Tal obra teve muito sucesso, sendo leiloada por US$ 432.500, reforçando a ideia de que o computador é capaz de ser uma ferramenta para produções artísticas [Pearson, 2018].

A inclusão de tal obra causou diversas discussões envolvendo tanto artistas quanto pesquisadores na área de aprendizado de máquina que se perguntavam sobre a possibilidade de algoritmos por si só, serem artistas. Mark Riedl, professor-associado de aprendizado de máquina e IA do Instituto de Tecnologia da Georgia classifica os algoritmos como "pincéis muito complicados com muitos parâmetros matemáticos, que possibilitam a criação de um efeito que seria difícil de obtê-lo de outra maneira" [Pearson, 2018].

Por fim, um exemplo da utilização de CNN's para criação de arte pode ser visualizado em Ribeiro and Mendes [2016]. Neste trabalho o autor treina a rede neural a partir de pinturas de estilos diferentes, fazendo com o que o algoritmo utilizado, reconheça automaticamente quais seriam os possíveis valores dos *pixels* de uma imagem referente a cada estilo. Desta forma, é possível à CNN, elaborar antigas obras de arte, sob diferentes estilos de pinturas, criando obras artísticas / digitais nunca vistas até então.

As redes neurais geralmente necessitam de uma quantidade de dados massiva para se obter um resultado e todo o conhecimento adquirido é então utilizado para gerar novos resultados. Ou seja, o produto final é diretamente ligado à um longo processo de seleção de dados, parâmetros matemáticos e seleção dos resultados obtidos. Desta forma, não podemos dizer que a máquina é totalmente responsável pela obra gerada, visto que o próprio pesquisador passou a ser um artista por ter sido o responsável por toda a

[16] http://museuartesacra.org.br/tota-machina

elaboração dos parâmetros e da arquitetura utilizada para a obtenção de tal resultado. Surge então a dúvida sobre quem seria o verdadeiro artista, o algoritmo por ser treinado durante meses para gerar determinada obra ou o pesquisador realizando diversos testes e ajustes para obter determinado resultado?

Sob a questão levantada acima, como toda obra de arte, devemos analisar todo o processo de criação e não apenas o produto final. Desta forma, vale refletir sobre o pensamento de três artistas contemporâneos: Anna Ridler, artista britânica que utilizou uma rede neural relacionando a volatilidade do mercado de criptomoedas com a Febre das Tulipas, que entende as técnicas de aprendizado de máquina como ferramenta, por possibilitar criações específicas, e também como processo, visto que a artista necessita de rotular objetos e modelar os dados a serem utilizados; Gauthier Vernier, artista francês que participa do grupo Obvious, citado acima, que possui um pensamento menos humano no sentido de quem é o responsável pela arte, fornecendo o crédito à máquina e assinando a obra com a própria fórmula do algoritmo, pois segundo o autor, foi ele mesmo quem criou a obra; e Mario Klingemann, artista alemão pioneiro no uso de IA, afirma que o que a máquina faz para criar uma obra de arte é basicamente imitar aspectos humanos, e sob a tecnologia atual, ela acaba falhando ao tentar interpretar aspectos subconscientes e emocionais que muitas vezes são difíceis de quantificar [Gonzaga and Fernandes, 2019].

Há discussões que precedem a questão sobre quem seria o artista em uma obra desenvolvida através de IA, como por exemplo a definição se o resultado obtido é ou não uma obra artística. Existem pessoas com um pensamento mais futurístico como a galerista Luisa Strina que defende a ideia de que para ter qualidade, a arte não deve seguir tendências. Ela ainda afirma que "Usada com coerência conceitual e estética, as novas tecnologias abrem uma porta para pesquisas que ainda não foram feitas na arte" [Calais, 2020]. As redes neurais são técnicas novas, se analisadas na campo artístico, e por não serem totalmente automatizadas, visto que o programador necessita de ajustá-la para determinados resultados, elas acabam abrindo espaço para inserção de conceitos estéticos e conceituais durante a etapa de criação, fazendo com o que o resultado final possa ser diretamente relacionado com a intenção do artista / programador. Consequentemente, as redes neurais acabam possuindo um grande potencial para criação de diversas obras artísticas de qualidade, nos trazendo mais uma questão levantada por diversos artistas. Em relação à criação de arte, futuramente as máquinas irão substituir os artistas?

Sob a questão acima levantada, vale ressaltar três pensamentos: Anna Ridler afirma que as pessoas pagam caros por obras simples e minimalistas, mas se analisarmos obras em que há bastante repetição, talvez a área de aprendizado de máquina otimize estes processos, facilitando a criação e talvez tornando a obra mais barata; Gauthier Vernier afirma que a máquina não irá substituir o homem, dado que o humano cuida de boa parte do processo criativo e é ele quem fornece propósito para o que foi criado, ou seja,

a máquina não é capaz de criar uma obra de arte por si só, é necessário que haja a colaboração humano / máquina; por fim, Mario Klingemann aponta que, para artistas que realizam trabalhos repetitivos, a área de aprendizado de máquina tem grande potencial de substituí-los futuramente [Gonzaga and Fernandes, 2019].

2.9 A arquitetura Pipes and Filters

A arquitetura Pipes and Filters é uma arquitetura que visa modularizar e simplificar um determinado sistema de processamento, particionando-o em uma sequência de processamentos menores, os quais são representados a partir de filtros / filters (Figura 6).

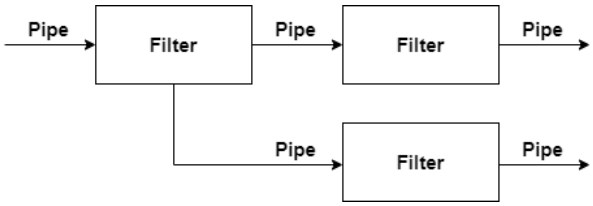

Figura 6: Um exemplo de arquitetura Pipes and Filters.

Estes filtros agem de forma individual e podem se conectar com qualquer outro tipo de filtro, sob diversas maneiras. Neste tipo de arquitetura, o filtro é basicamente uma interface a qual recebe uma entrada, processa uma mensagem e publica / propaga o resultado como saída para outro filtro. Tanto a entrada quanto a saída são chamados de pipes os quais são responsáveis por conectar os filtros.

Este tipo de arquitetura será utilizado como base para a criação da arquitetura de interação do público proposta neste trabalho.

3 Proposta de modelo para a participação do público

Os Instrumentos Musicais Digitais (IMD's) podem ser vistos como algoritmos, os quais ficam responsável por realizar sínteses sonoras que são dadas de acordo com as informações captadas através de uma interface, que estabelece a comunicação do músico com o instrumento. A influência destas informações sob o som reproduzido é de total responsabilidade do programador e geralmente é chamada de mapeamento [Rovan et al., 1997]. Patrício [2010] define a estrutura de um IMD como interface, mapeamento e síntese sonora. Instrumentos musicais convencionais possuem a mesma estrutura de um IMD, onde uma ação é mapeada para uma reação específica. Em um piano, por exemplo, temos as teclas que agem como uma espécie de interface entre o músico e o instrumento. Ao pressionar uma tecla, o músico fornece uma entrada para o sistema que mapeia diretamente qual saída ele deve retornar, ou seja, qual corda ele deve tocar.

Uma das vantagens ao se criar um IMD é a possibilidade de realizar mapeamentos mais complexos, os quais chamamos de "um-para-um", "muitos-para-um", "um-para-muitos" e "muitos-para-muitos" [Hunt et al., 2000]. Estes mapeamentos serão explicados na seção 3.3.

Igualmente à estrutura por trás de um IMD, podemos visualizar os exemplos de interação do público em performances digitais como arquiteturas que se encaixam sob três componentes principais, dentre eles a entrada, o processamento/mapeamento e a saída (Figura 7).

Figura 7: Arquitetura de um IMD simplificada.

Estes componentes podem se relacionar de diversas formas, como por exemplo a saída realimentando a entrada, a entrada passando por diversos mapeamentos, realimentando outros sistemas, ou até mesmo a relação mais convencional, onde a entrada passa por um mapeamento e gera uma saída.

Nota-se que as interações do público em ambientes performáticos digitais podem ser representadas de maneira semelhante à estrutura apresentada. Toda interação parte de uma determinada entrada, a qual engloba a coleta de dados a partir do público, passa

por um mapeamento, o qual dirá como os dados coletados serão utilizados, e por fim, gera uma determinada saída, referente ao *feedback* que a interação irá proporcionar. Levando em consideração os aspectos técnicos que podem envolver protocolos de rede, diferentes formas de mapeamento e de comunicação entre os componentes, além das formas de interação do público em performances digitais e como elas podem ser realizadas, este trabalho utiliza a arquitetura de um IMD como base de sua criação.

Este capítulo ficará responsável por propor e definir a arquitetura de interação do público mencionada anteriormente. A Seção 3.1 irá abordar sobre quais são as possibilidades de dispositivos tecnológicos para se criar uma performance digital interativa. Já a seção 3.2 irá abordar sobre o que estas entradas podem gerar, como por exemplo síntese sonora, síntese de imagens ou algum tipo de *feedback* para o público. As possíveis formas de mapeamento serão abordadas na Seção 3.3. As formas de comunicação entre estes componentes, os quais geralmente envolvem protocolos de rede, por exemplo, serão abordadas na Seção 3.4 e por fim, a Seção 3.5 irá apresentar um resumo do modelo proposto.

3.1 Entrada

O componente inicial a ser analisado, tomando como base a arquitetura de um IMD, é aqui chamado de entrada. Este componente engloba as possibilidades existentes para a coleta de informação do público. É neste componente que a interação física em si, do público com o sistema, é dada, visto que os componentes de mapeamento e saída são responsáveis por apenas definir o que a interação, já realizada, irá proporcionar.

Dentre as possíveis ferramentas para captação de informação em um ambiente performático digital, temos os dispositivos móveis, os computadores e os sensores. Apesar do dispositivo móvel possuir diversos sensores, neste trabalho ele será analisado como um dispositivo a parte. Além disso, os computadores e os dispositivos móveis serão analisados em conjunto, dado a semelhança entre suas formas e propósitos de utilização.

Em relação aos sensores, eles são componentes ou circuitos elétricos que são capazes de captar alguma condição/informação do ambiente. Estes sensores podem ser divididos entre duas categorias [Patsko, 2006]: sensores analógicos, que se baseiam em sinais analógicos limitados entre dois valores de tensão, podendo assumir infinitos valores intermediários e possibilitando atribuir a cada condição medida, um valor de tensão correspondente; e os sensores digitais, que se baseiam em níveis de tensão bem definidos, sendo eles Alto (*High*) ou Baixo (*Low*).

Patsko [2006] exemplifica alguns tipos de sensores analógicos: *Light Dependent Resistor* (LDR) para captar a variação na luminosidade do ambiente; termistor para captar a temperatura do ambiente; sensor de peso para medir o peso de algum objeto ou uma força aplicada sob ele; potenciômetro utilizado para ajustar algum parâmetro ou

atuar como sensor de posição, fornecendo sua atual angulação dado um evento; microfone para captar o som do ambiente; sensor de vibração para captar a vibração de objetos; sensor de distância; acelerômetro para medir a aceleração sofrida por um corpo; sensor de pressão atmosférica para medir a força exercida pelo ar sobre alguma superfície; e por fim, o sensor de campo magnético.

Já em relação aos sensores digitais: chaves digitais para indicar se estão pressionadas ou não; chave magnética que são chaves digitais que são ativadas através de ímãs; e os pares ópticos que englobam um *Light Dependent Resistor* (LED) emissor de infravermelho e um fototransistor, geralmente utilizados para transmissão de dados sem fio levando em consideração seu posicionamento.

É possível criar diversos tipos de interação em performances digitais à partir dos sensores e durante muito tempo eles eram utilizados em pequena escala devido ao seu alto valor de mercado. Para que o público tivesse acesso aos meios de interação criados era necessário toda uma instalação digital equipada com sensores distintos fornecidos pelo próprio criador da obra [Hindle, 2013].

Com o avanço tecnológico os dispositivos móveis, os quais possuem diversos sensores acoplados, passaram a se tornar ubíquos a ponto de estarem bastante presentes no cotidiano das pessoas de forma tão transparente que chegamos a não perceber que os mesmos estão lá. Além da ubiquidade, por terem se tornado mais acessíveis, estes dispositivos acabaram sendo utilizados cada vez mais em performances digitais, visto que eles possuem diversos sensores aliados à alta capacidade de processamento, que possibilita a criação de novas formas de interação nestes ambientes. Inicialmente eles eram utilizados em orquestras e grupos musicais que compartilhavam a mesma aplicação a qual sintetizava o som de cada músico através de alto-falantes, sob *Musical Instrument Digital Interface* (MIDI) ou via mixer [Hindle, 2013].

Em geral, os dispositivos móveis podem ser utilizados através de um aplicativo, onde o público baixa o conteúdo necessário para realizar a interação, ou através de aplicações web, onde a participação é dada de forma imediata. De acordo com os exemplos apresentados na subseção 2.5, podemos concluir que a utilização de aplicativos pode causar uma certa exaustão / insatisfação por parte do público, devido à necessidade de ter que se envolver com aspectos técnicos, como por exemplo a pré-configuração de hardware para acessar um sistema ou apenas encontrá-lo em uma plataforma para baixá-lo, o que para leigos, nem sempre é uma tarefa fácil [Hindle, 2013].

Porém, o uso de aplicações web tem sido uma ótima solução para facilitar a interação público/aplicação em performances digitais. Durante um tempo, houve uma certa dificuldade por parte dos programadores em implementar uma aplicação multi-plataforma que fosse compatível com todos os possíveis tipos de dispositivos móveis do público [Weitzner et al., 2012]. E apesar da necessidade de informar o *Uniform Resource Locator* (URL)

da aplicação ao público, as aplicações web são ainda promissoras pois ao acessar o URL, o público tem acesso direto às funcionalidades esperadas. Além disso, ao utilizar aplicações web, desenvolvidas localmente, há a necessidade de informar ao público sobre como se conectar à rede, porém em determinados casos este tipo de aplicação pode ser mais vantajosa por não utilizar Internet, o que às vezes pode ser um empecilho visto que nem todos os participantes podem não possuir acesso a ela.

Por fim, levando em consideração o ponto de vista musical, nota-se que os dispositivos móveis modificaram conceitualmente o ato de se produzir e escutar música, principalmente em relação aos papéis envolvidos nesta relação. Na música clássica de tradição europeia, existem três papéis: o intérprete, responsável pela execução; o público, responsável pela escuta; e o compositor, responsável pela criação. Ao analisar sob o ponto de vista do público não apenas como responsável pela escuta, mas também por exercer o papel de artista/compositor em performances digitais, temos o surgimento de um novo diálogo sonoro/musical que quebra a tradição europeia, propiciando uma experiência única e especial do público na performance artística [Hödl et al., 2017].

Além do uso de dispositivos móveis, o computador também tem sido utilizado para criações e composições musicais. Um dos primeiros grupos de músicos a se denominarem uma orquestra regida através de computadores, foi criado no fim do ano de 2005, o qual era chamado de Princeton Laptop Orchestra (PLOrk) [Trueman et al., 2006]. Esta orquestra é composta de 15 IMD's, os quais eram chamados de meta-instrumentos, onde cada um é composto por um laptop, uma interface e um alto-falante.

A PLOrk[1] ampliou o cenário musical por trazer uma abordagem bastante radical sob a composição eletrônica e acústica. Os compositores tiveram que deixar de lado a definição padrão dos instrumentos acústicos existentes para repensar sobre quais instrumentos poderiam ser criados, dado a infinidade de possibilidades que computadores / sensores podem prover. Consequentemente, não apenas instrumentos foram criados, mas também novas formas de interação do músico com seu respectivo instrumento.

O ato de se produzir música através de um computador pode ser dado em conjunto, como em orquestras digitais, mas também individualmente, de maneira em que o músico pode realizar sua própria performance musical, sem depender de outro músico. Sob este ponto de vista, começaram a surgir os programadores / compositores que utilizavam o computador como um IMD, o qual produzia música à partir de algoritmos. Este novo tipo de performance contemporânea foi chamada de *live coding* e acabou fomentando a criação de linguagens específicas como a ChucK [Wang et al., 2003], criada em 2002 e utilizada posteriormente pela PLOrk, que é uma linguagem multi-paradigma com os tipos de dados fortemente estruturados, compilada com instruções executadas em tempo real. Collins et al. [2003] e Ward et al. [2004] definem o *live coding* como uma performance

[1] https://plork.princeton.edu

onde o compositor necessita de ter um certo conhecimento sobre algoritmos e mapeamentos à partir de códigos, os quais são desenvolvidos em tempo real e consequentemente, prestigiados pelo público.

Um exemplo de performance baseada em *live coding* pode ser visualizado em Devir [Antar, 2016], a qual se trata de uma performance executada durante a série "Música?", desenvolvida pelo núcleo de pesquisa em sonologia da Universidade Federal de São Paulo (NuSom-ECA). A série tem como objetivo causar certa reflexão ao público sobre o que é realmente música, visto que atualmente existem pessoas que não concordam com a afirmação de que o *live coding* é uma forma de criação e composição musical. Segundo os autores, a Devir busca explorar a relação que o ambiente possui com o público, analisando aspectos visuais e temporais do mesmo.

Em uma performance musical, onde o compositor geralmente utiliza um instrumento acústico, a relação compositor-público se dá de acordo com o fluxo sonoro produzido e os gestos do intérprete, onde os estímulos causam respostas instantâneas / imediatas. Porém, pela ausência de instrumentos acústicos em performances de *live coding*, a interação vai além da físico-mecânica ou corpo-instrumento, também sendo físico-abstrata, dada por corpo-código. Desta forma, o público acaba demorando mais para traduzir ou entender a proposta do som gerado, ampliando as possibilidades durante o processo de composição [Antar, 2016].

Por fim, com base na Subseção 2.5, podemos definir as formas de interação do público em performances digitais como: a reprodução de gestos corporais [Hout et al., 2014, Turchet and Barthet, 2018]; a utilização de dispositivos móveis como IMD's [Hindle, 2013, Lee and Freeman, 2013, Junior et al., 2016], como dispositivos responsáveis por influenciar a iluminação do local, como alto-falante [Hirabayashi and Eshima, 2015] ou como interface para que o público possa fornecer dados para aplicações/algoritmos [Gimenes et al., 2016, Dahl et al., 2011, Zhang et al., 2016, Weitzner et al., 2012, Lins and Novaes, 2017]; tarefas ou jogos que fazem com que o público interaja com o ambiente, ou com os performers, procurando dicas de como realizá-los; a utilização de sensores para captar características físicas, como a altura ou a largura de cada participante [Eaton et al., 2014]; ou até mesmo a emissão de algum som através do participante, que pode ser captado para determinado fim ou não, como em 4'33 de John Cage [Rocha, 2005]. Estas formas de interação podem ser realizadas através de dispositivos móveis, sensores, microfones, teclados, *webcam's*, ou até mesmo objetos físicos que irão induzir o público a realizar determinada tarefa.

3.2 Saída

O componente de saída é responsável por fornecer o *feedback* para o público ou para os artistas na forma de: sons, imagens, luzes ou instruções para determinado participante

realizar. Analisando de forma geral o modelo proposto neste capítulo, observa-se que a interação do público com o sistema deve sempre estar presente no componente de entrada. Além disso, o performer nunca deve ser ator no componente de entrada, visto que neste caso, ele iria interagir com o público, de forma ativa, e o público iria apenas receber um determinado *feedback*, atuando de forma passiva, sem interagir com o sistema. Já o componente de saída pode definir a interação tanto do performer quanto do público, ou seja, se analisarmos os casos em relação aos possíveis atores dos componentes de entrada-saída, temos:

- **público-público**, onde a interação é exclusivamente entre o público e o sistema, sendo dada de forma individual, onde o participante é totalmente responsável pelo resultado gerado, ou de forma coletiva, onde cada participante influencia parcialmente no resultado, fazendo com o que o mapeamento seja mais complexo, como por exemplo uma performance em que o músico reproduz a música de fundo de um ambiente, através de instrumentos físicos e o público através de instrumentos digitais, onde o performer / músico em si não realiza interação com o sistema, fazendo com o que a interação a ser proporcionada seja inteiramente por parte do público;
- **público-performer**, onde a ação do público influencia diretamente na forma em que o performer deve realizar determinado ato, como por exemplo o público, responsável por votar sobre qual música o performer deve tocar, através de dispositivos móveis, fazendo com o que estes dispositivos se tornem uma interface para estabelecer a comunicação entre o público e o performer.

Este trabalho não levará em consideração casos em que ambos os componentes de entrada e saída são fornecidos pelo performer, já que neste caso não há interação do público. Em geral, sob performances digitais existem cinco tipos de *feedback's*, dentre eles:

- **Sonoro**, sendo qualquer tipo de reprodução sonora que tenha sido mapeada através de informações captadas do público. Desta forma, o som produzido pode ser uma música escolhida através de votos [Zhang et al., 2016, Weitzner et al., 2012, Lins and Novaes, 2017], um som tocado através de dispositivos móveis [Hindle, 2013, Lee and Freeman, 2013, Junior et al., 2016, Bin et al., 2016, Gimenes et al., 2016, Hirabayashi and Eshima, 2015], um som sintetizado com características (tempo, duração, timbre, dentre outras) mapeadas à partir de dados fornecidos pelo público [Dahl et al., 2011, Hout et al., 2014, Eaton et al., 2014] ou até mesmo sons que o performer pode emitir. Em relação aos possíveis sons em que o performer pode realizar, existem sons que são reproduzidos para complementar a paisagem sonora do ambiente ou sons que são reproduzidos para estabelecer uma comunicação verbal do performer com o público, geralmente sendo dado na forma de explicações sobre o funcionamento da performance ou do aplicativo a ser utilizado.

- **Visual**, sendo dado através de gestos realizados por performers, alterações na iluminação ou na disposição de objetos em um cenário performático ou projeções visuais as quais podem ser criadas sob o mapeamento de diversas informações, estabelecendo uma relação interativa do público com o que é projetado. Estas projeções podem ser imagens do próprio público sob determinados filtros, imagens geradas através de gestos, características físicas ou expressões faciais do público ou do performer [Mancini et al., 2007], ou até mesmo imagens geradas através do mapeamento de características sonoras de uma performance digital/musical, como por exemplo o tempo e a dinâmica [Dixon, 2005, Katayose and Okudaira, 2004, Friberg and Battel, 2002];

- **Háptico**, sendo relacionado às interações de toque que ocorrem pela percepção ou manipulação de objetos. Este tipo de *feedback* visa aumentar a interação homem-máquina, visto que ele é capaz de aprimorar a percepção do usuário sob o ambiente, tornando a interação mais intuitiva e natural [Caetano, 2008]. Em geral, o *feedback* háptico em performances ou instalações digitais pode ser dado através do uso de vestimentas com atuadores para que o público consiga sentir vibrações relacionadas à música do ambiente [Turchet and Barthet, 2018] ou ao próprio músico [Thorn and Lahey, 2019]; o uso dispositivos robóticos como uma luva, para ensinar o usuário a tocar um instrumento musical [Berdahl and Huber, 2015]; ou até mesmo da junção de sistemas mecânicos / digitais com instrumentos comuns, como em uma flauta, para auxiliar o músico no seu processo de aprendizagem [Zhang et al., 2019];

- **Paliativo e/ou Olfativo**, sendo os tipos de feedbacks menos utilizados em performances digitais. Isso se dá pelo fato de que o olfato e o paladar são sentidos bastante particulares e que podem causar maiores insatisfações ao público, se comparados com a visão e audição, além de toda a dificuldade técnica para se implementar um sistema que provê este tipo de interação. Ou seja, o cheiro de um ambiente ou o gosto de algo tem grandes chances de não agradar a todos, fazendo com o que este tipo de *feedback* não seja uma escolha comum para prover a interação do público em performances digitais.

Estes *feedback's* podem ser fornecidos tanto para o público, quanto para os artistas. Além disso, levando em consideração a arquitetura proposta, os *feedback's* podem ser implementados como a saída final do sistema, onde ele é gerado e a interação é finalizada, ou podem realimentar um novo sistema.

O componente de saída fica então responsável por receber uma informação já mapeada, e em seguida executá-la, decidindo se o artefato produzido será propagado ou não para a camada de entrada. A informação coletada pode ser obtida à partir de mapeamentos de sistemas distintos, os quais podem ser descritos como um só, se forem

analisados sob apenas um componente de saída. A saída pode ser então classificada como Recursiva, quando realimenta a entrada de um sistema ou Não-recursiva quando o *feedback* é fornecido ao usuário, a interação usuário/sistema é finalizada e os artefatos gerados não são propagados para nenhum outro componente.

3.3 Mapeamento

Dentre os componentes mais importantes da arquitetura proposta está o mapeamento. Este componente é responsável por definir o quão expressivo um sistema de interação entre o usuário e a máquina pode ser. Levando em consideração o mapeamento existente na estrutura de um IMD, Hunt et al. [2000] o define como "a ligação ou a correspondência entre o controle de parâmetros, obtidos à partir de uma ação performática ou um gesto, e os parâmetros de síntese sonora". O mapeamento em IMD's define a influência que cada parâmetro coletado irá proporcionar aos parâmetros que serão utilizados para realizar uma determinada saída.

O controle do usuário sob as possíveis técnicas de mapeamento pode ser dado em dois níveis: microscópico, quando o mapeamento é dado em baixo nível, fazendo com o que o usuário tenha acesso a cada parâmetro sonoro em detalhes, como timbre, ataque ou *delay*, o que geralmente causa uma certa insatisfação por parte do usuário, por ter que possuir determinado conhecimento musical para obter o resultado desejado; ou macroscópico, onde o usuário possui pouca liberdade, podendo apenas escolher à modelos pré-determinados, alterando por exemplo, o ritmo de uma composição ou qual partitura deve ser tocada. Para prover o engajamento do usuário, é necessário definir uma forma de mapeamento que procure balancear o nível de controle do usuário com o resultado a ser proporcionado.

Em geral, podemos definir os tipos de mapeamentos de qualquer sistema sob quatro tipos distintos:

- **Um-para-um ou mapeamento direto**, quando uma ação independente gera uma entrada que está diretamente ligada à alteração de apenas um parâmetro de saída. Este é o tipo de mapeamento menos expressivo, visto que para cada ação já existe um resultado esperado.

- **Muitos-para-um ou convergente**, quando duas ou mais ações alteram um parâmetro de saída. Este mapeamento requer uma certa experiência do usuário para que ele obtenha controle efetivo do sistema, visto que ele precisa entender todo o conjunto de ações que levaram à alteração de um parâmetro específico de saída. Este mapeamento acaba sendo mais expressivo, se comparado ao mapeamento um-para-um, visto que o usuário possui diversas formas de produzir determinado resultado;

- **Um-para-muitos ou divergente**, onde uma ação resulta na modificação de dois ou mais parâmetros sonoros. Neste tipo de mapeamento as ações do usuário são bastante limitadas, chegando a um nível de controle macroscópico, visto que o usuário fornece apenas uma informação para obter um resultado. Desta forma, o mapeamento um-para-muitos acaba não permitindo que o usuário altere parâmetros mais complexos, que exigem mais informações para serem alterados, como por exemplo o envelope de um som;

- **Muitos-para-muitos**, sendo o mapeamento mais complexo dentre os citados, onde diversas relações podem ser implementadas, fazendo com o que um conjunto de ações possa influenciar na alteração de um conjunto de parâmetros de saída.

Um exemplo simples de mapeamento pode ser demonstrado na instalação digital Per(sino)ficação [Carvalho et al., 2019].

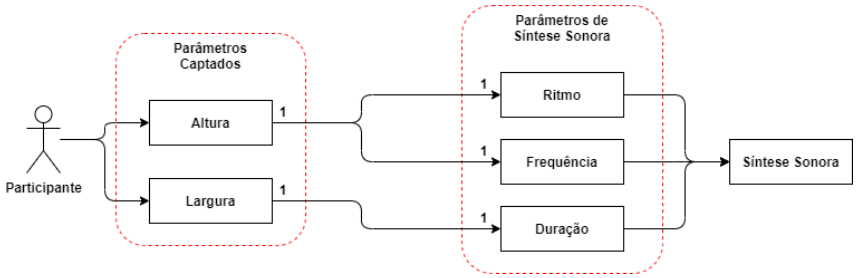

Figura 8: Mapeamento do trabalho Per(sino)ficação (Seção 4.2)

Em Per(sino)ficação, a altura e largura do participante são captadas através de uma *webcam* e são então mapeadas para modificar os parâmetros de um sintetizador de sinos (Figura 8). O valor referente à altura do participante é modificado a partir de duas equações matemáticas, as quais retornarão dois números que irão representar qual a frequência e qual o ritmo do som a ser sintetizado. Já a largura, é mapeada e processada a partir de uma equação matemática para definir qual será a duração do som. Sob a definição de Hunt et al. [2000] sobre mapeamento, a altura e largura do participante podem ser descritas como os parâmetros captados, e o ritmo, frequência e duração como os parâmetros de síntese sonora. A relação entre estes parâmetros define o mapeamento, que neste caso pode ser nomeado como "um-para-muitos", visto que o parâmetro coletado à partir da altura pode alterar mais de um parâmetro sonoro.

Em um sistema de interação também é possível utilizar mais de um mapeamento, como por exemplo o modelo apresentado no ambiente de síntese sonora ESCHER [Wanderley et al., 1998]. Neste ambiente, o mapeamento é dado em duas camadas, onde a primeira camada fica responsável por detectar qual instrumento foi tocado, além de mapeá-lo para um modelo específico, o qual será processado na segunda camada. Existem trabalhos que definem estas duas camadas como camadas *input* e *outuput*, como Ungvary and

Kieslinger [1998], Mulder et al. [1997], citados em Hunt et al. [2000]. Desta forma, na arquitetura proposta neste trabalho, o componente Mapeamento poderá ser classificado como mapeamento unitário ou mapeamento com multi-camadas.

Em outra possibilidade, segundo Hunt et al. [2000] o mapeamento pode ser dado através de mecanismos generativos, os quais geralmente operam sob técnicas de IA e aprendizado de máquina, fazendo com o que o mapeamento seja criado por meio de adaptações internas a partir de um determinado treinamento, ou mapeamento explícito, onde são definidas estratégias que explicitam as relações existentes. Podemos definir então duas formas de mapeamento: dinâmico, quando utilizado a partir de mecanismos generativos, os quais podem alterar as relações definidas, em tempo real; e estático, quando o mapeamento é explícito e pré-definido, não se alterando ao longo de sua execução.

Por fim, de maneira geral, o mapeamento em um sistema interativo define o controle que o público tem sob o sistema implementado, podendo ser divido em três categorias [Kirke and Miranda, 2012]:

- Sistemas onde o público tem controle direto sob os parâmetros do som, como por exemplo tempo, dinâmica, articulação e timbre. Este tipo de sistema geralmente é mais flexível, por permitir ao público, alterar parâmetros bastante específicos, podendo gerar diversos tipos de resultados;

- Sistemas onde o público possui acesso a modelos já definidos. O controle expressivo muitas vezes requer a manipulação simultânea de muitos parâmetros, o que pode ser um empecilho para usuários leigos. Desta forma, a utilização de modelos pré-definidos possibilita ao usuário manipular um conjunto de parâmetros de baixo nível, à partir de parâmetros com maior nível de abstração, como por exemplo a alteração de um determinado conjunto de notas pré-definidas referentes ao som a ser sintetizado. O controle direto e a utilização de modelos podem ser utilizados em conjunto. O usuário pode controlar parâmetros de baixo nível, como o tempo, o qual será suavizado à partir de um modelo, que leva em consideração predições sobre qual seria o tempo esperado pela música.

- Sistemas onde o público possui acesso a um nível mais alto, à partir de descritores semânticos como por exemplo expressões emocionais como agressivo, alegre ou triste. Neste caso uma emoção alegre poderia ser descrita com um andamento mais rápido, som intenso e timbre colorido, já uma emoção triste poderia descrever o contrário dos parâmetros citados. A vantagem da utilização deste tipo de sistema, é a capacidade de tornar o controle mais intuitivo além de exigir menos habilidade por parte do usuário.

3.4 Comunicação entre os componentes

A participação do público em performances digitais geralmente é dada através de dispositivos móveis ou computadores, conectados em rede. Uma das formas em se prover este tipo de comunicação é através do uso de uma rede local. A utilização de uma conexão local em performances digitais força o público a se juntar em um ambiente fechado, além de proporcionar uma interação mais rica entre o público por meio da aplicação desenvolvida. Em casos que a aplicação desenvolvida necessita de Internet para ser acessada, o uso de uma rede local pode ser vantajoso visto que a Internet pode ser fornecida através da rede criada pelos proponentes da apresentação artística.

Além da estrutura da rede em si, é necessário utilizar algum tipo de protocolo para estabelecer a comunicação entre as aplicações. Um protocolo que tem sido bastante utilizado para criação de IMD's colaborativos / interativos é o *Open Sound Control* (OSC), que é "um protocolo para comunicação entre computadores, sintetizadores de som e outros dispositivos multimídia que é otimizado para tecnologia de rede moderna" [Freed, 1997]. Wright [2005] o define como "um formato de mensagem binária que pode ser transportada por qualquer tecnologia de rede de propósito geral", podendo ser utilizada em conjunto com protocolos como o *Transmission Control Protocol* (TCP) ou *User Datagram Protocol* (UDP), via Ethernet ou conexões wifi.

Uma das vantagens em se utilizar o protocolo OSC é referente às mensagens transmitidas através dele, as quais não possuem um padrão sob o que o sintetizador deve implementar. Desta forma, o processo criativo durante a implementação de um IMD torna-se ainda mais artístico, quando utilizado via OSC, visto que o programador fica encarregado de decidir quais parâmetros devem estar acessíveis, quais seus respectivos nomes e como eles devem estar organizados. Esta flexibilidade forneceu suporte para a criação de diversos softwares e hardwares voltados para criação musical. Apesar de ser uma vantagem, apesar do protocolo OSC fornecer bastante liberdade ao programador, a sua má utilização pode acabar resultando na definição de nomes confusos para certos parâmetros ou até mesmo na utilização de parâmetros desnecessários.

Wright [2005] exemplifica alguns exemplos de implementações do protocolo OSC em: linguagens de programação como C, Java, Javascript, PHP, Python, dentre outras; sistemas interativos para síntese sonora e linguagens para processamento como ChucK [Wang et al., 2003], Max/MSP [Bevilacqua et al., 2005], Pd [Puckette, 1996] e Supercollider [McCartney, 2002]; e hardware para captura de gestos a partir de sensores como EtherSense [Fléty et al., 2004] e Teabox [Allison and Place, 2005]. Exemplos de performances que utilizam OSC podem ser visualizados em performances digitais musicais onde participantes produzem sons através de IMD's como em "Without Borders" [Gimenes et al., 2016] e "The Space Between Us" [Eaton et al., 2014], citados na subseção 2.5, ou até mesmo em orquestras como "The Meta-Orchestra" [Harris, 2004], composta

por um conjunto de Macintoshes que sintetizam sons através do SuperCollider e trocam parâmetros entre si através de OSC.

Em performances digitais musicais que utilizam IMD's como forma de interação, a cada nota tocada o músico fornece uma entrada para o sistema a qual é mapeada para sintetizar determinado som. Este tipo de sistema geralmente não necessita da confirmação de que os dados foram enviados corretamente, porém é necessário garantir que a latência entre o ato de fornecer a entrada e ouvir o som produzido deva ser a mais baixa possível. Por possuir baixa latência e ser fácil de se implementar, o protocolo OSC acabou se tornando o protocolo de uso geral para composição digital / musical.

3.5 Revisando o modelo

De acordo com o modelo proposto, temos então três componentes principais: Entrada; Mapeamento; e Saída. A Figura 9 demonstra de maneira resumida, a função de cada um.

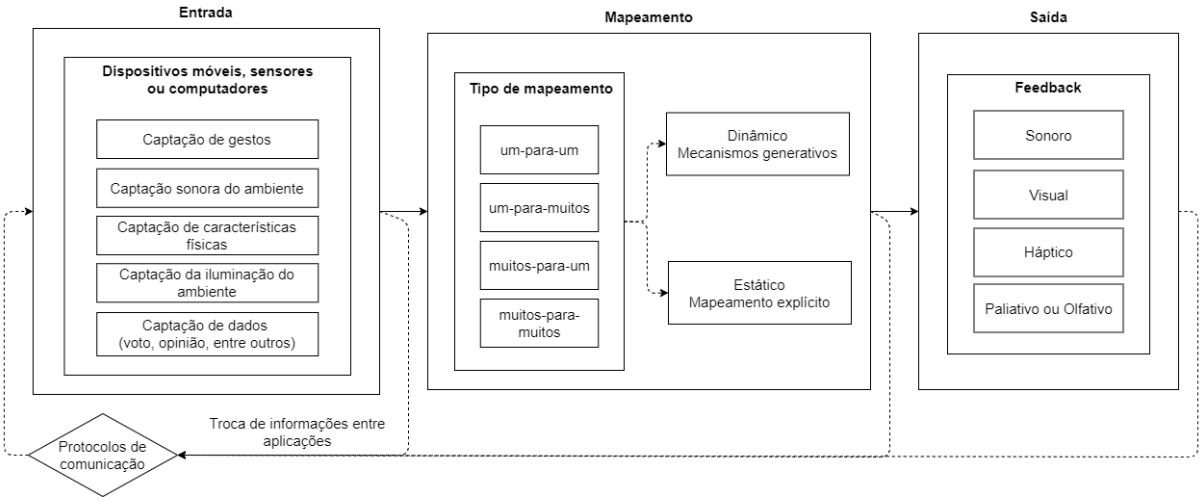

Figura 9: Arquitetura conceitual de interação do público.

O componente de entrada é totalmente dependente da forma de interação a ser proporcionada em uma performance, visto que ele fica responsável por definir qual tipo de captação será utilizado. Dentre as mais diversas formas de captações, temos: captação de gestos através de câmeras e *webcam's*, ou até mesmo através de sensores de movimento ou de presença para extrair imagens, vídeos ou até mesmo informações visuais sobre alguma movimentação no ambiente; captação de som através de gravadores ou microfones; captação de características físicas, a qual pode ser dada através de câmeras e *webcam's*, sensores térmicos ou mecânicos para extrair a temperatura, peso, altura, largura, envergadura, cor, dentre outras características do participante e / ou do ambiente; e captação

de dados, que pode ser dada de diversas formas, sejam elas através de mouse, teclado, botões, ou até mesmo via dispositivo móvel, através de uma interface / teclado virtual.

Já o componente de mapeamento, como visto na Seção 3.3 pode ser dado sob quatro formas, de acordo com a quantidade de entradas a serem mapeadas e de saídas. Em sistemas de voto onde o público vota para que um músico reproduza determinado som, por exemplo, geralmente o mapeamento é definido como "muitos-para-um" visto que todas as entradas fornecidas influenciam na saída, a qual define uma única ação a ser tomada. Caso os votos influenciassem, independentemente, cada músico de uma banda, o mapeamento poderia ser classificado como "muitos-para-muitos". Já em um IMD, por exemplo, o mapeamento poderia ser dado como "um-para-um" se cada botão da interface reproduzisse um único som, e como "um-para-vários" caso reproduzissem sons distintos sob diferentes fontes sonoras. O mapeamento criado pode ser estático, o qual não se modifica no decorrer da performance, ou dinâmico, o qual pode variar sob uma maneira pré-definida, aleatoriamente ou até mesmo sob técnicas de aprendizado de máquina.

Por fim, o componente de saída fica responsável por definir o tipo de *feedback* resultante da interação. Este *feedback* é diretamente ligado aos sentidos sensoriais humanos, os quais se passam por diferentes formas de percepção do participante no ambiente. Alguns tipos de feedbacks, como por exemplo os visuais, sonoros e hápticos são mais fáceis de serem implementados e proporcionados ao público. Já os feedbacks paliativos ou olfativos, geralmente são escolhas menos utilizadas, visto que o olfato e o gosto do participante são individuais, podendo influenciar negativamente a experiência de participantes durante a performance. Todos os componentes desta arquitetura podem realimentar um outro sistema de interação, através dos protocolos de comunicação.

A arquitetura apresentada na Figura 9 tem o objetivo de demonstrar as possibilidades de interações em performances digitais sob o ponto de vista de quais são os dados a serem coletados, como serão mapeados e quais os tipos de resultados eles podem proporcionar aos participantes, além das comunicações ou protocolos a serem utilizados entre os sistemas / aplicações.

A fim de simplificar ainda mais a proposta, uma arquitetura baseada em pipes-and-filters foi utilizada para representar os modelos de interação da performance e das instalações digitais criadas (Capítulo 4) e a arquitetura até então demonstrada (Figura 9) será utilizada para explicar como a interação com o público é dada em cada obra.

A Figura 10 demonstra um exemplo genérico de uma possível interação em uma performance ou instalação digital.

Na arquitetura proposta (Figura 10) cada tipo de interação será descrita como um fluxo o qual passa por filtros de hardware e de software. Os filtros de hardware são os dispositivos tecnológicos a serem utilizados na performance, podendo ser demonstrados

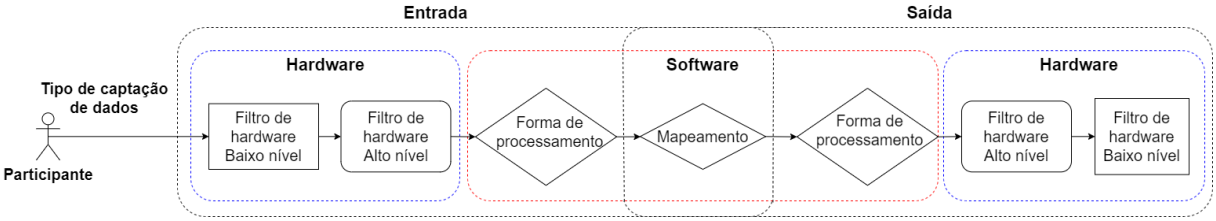

Figura 10: Proposta de arquitetura final de interação do público.

a mais baixo nível, como teclados, microfones, mouses e *webcam's* e também a um nível mais alto como computadores *desktop*, *notebooks* ou dispositivos móveis. Já os filtros de software englobam tanto o mapeamento quanto as formas de processamento dos dados. Desta forma temos, como exemplos de filtros de software, o processamento de som o qual envolve MIR e o processamento de imagem envolvendo visão computacional.

A utilização de uma arquitetura baseada em pipes-and-filters torna-se bastante promissora para avaliar os trabalhos desenvolvidos devido ao seu alto grau de modularidade, que demonstra de maneira simples como as interações ocorrem, de acordo com os aparatos tecnológicos utilizados e as técnicas de processamento dos dados coletados. A validação da arquitetura proposta será dada sob a performance "O Chaos das 5" (Seção 4.1) e sob as instalações "Persinoficação" (Seção 4.2) e "Leiamídia" (Seção 4.3).

4 Performances e Instalações Digitais Desenvolvidas

Este capítulo irá abordar as implementações e resultados alcançados a partir dos trabalhos desenvolvidos no contexto deste trabalho. Dentre eles temos a performance digital "O Chaos das 5", a qual foi apresentada em cidades como São Paulo, São João del-Rei, Belo Horizonte, Mariana e Conselheiro Lafaiete, a instalação tecnológica "Per(sino)ficação" também apresentada em São João del-Rei e a "Leiamídia" que devido ao COVID-19, ainda é inédita e não foi estreada.

Estes trabalhos serão analisados de acordo com a arquitetura proposta no Capítulo 3 deste trabalho, demonstrando as interações realizadas em cada trabalho, do ponto de vista técnico / tecnológico.

4.1 Espetáculo "O Chaos das 5"

O espetáculo "O Chaos das 5"[1] é uma performance criada pelos laboratórios ALICE e Ecolab do Gtrans (Grupo de pesquisa Transdisciplinar) da Universidade Federal de São João del-Rei (UFSJ) que envolve a participação do público em sua execução utilizando para isto dispositivos móveis. Da mesma maneira que um *Happening* de Kaprow [Ribeiro, 2010], este espetáculo cria um ambiente motivado pela preocupação com o espaço e o direcionamento do público no mesmo, além de atribuir responsabilidades ao espectador para a consolidação da peça.

Apesar de haver instruções para o público, como em um *happening*, o espetáculo se trata de uma performance onde os participantes da audiência podem decidir como agir e interagir com o espetáculo podendo ir de espectador a compositor, músico e participante da performance corporal, sendo convocados para irem além da simples contemplação do que está sendo apresentado [Santos, 2008].

A fim de proporcionar uma experiência imersiva para o público, o espetáculo conta com três camadas de informação:

Camada Gestual: caracterizada pela junção de gestos e improvisos a partir de 5 artistas interagindo fisicamente com o público no espaço do espetáculo.

Camada visual: que conta com projetores e aplicações para criação de imagens em tempo real, desenvolvidas especialmente para esta peça, apresentando uma estética que

[1] https://www.youtube.com/watch?v=85olXggE4EY

coloca o computador em cena, abrindo a "Caixa Preta"[2] e expondo a máquina em seu fazer imagético.

Camada musical: contando com músicos/integrantes do grupo, os quais ficam localizados ao redor do espaço, sendo responsáveis por criar a base sonora a partir de IMD's e guitarras elétricas processadas e o próprio público que produz música a partir de dispositivos móveis / IMD's.

Na camada gestual, os performers se colocam entre os membros da audiência dificultando a distinção entre quem faz parte do público e quem são os artistas. Esta distinção se torna mais clara no decorrer do espetáculo, especialmente em um momento em cena onde os performers tiram parte de seu figurino e convidam o público a interagir, pintando seus corpos com canetas e tinta (Figura 11).

Figura 11: Performance realizada pelos artistas no espetáculo O Chaos das 5.

A camada visual remete ao paradigma da pintura por meio da ação [Ribeiro, 2010], sendo que neste espetáculo, a pintura é digital. As ações feitas para criar as imagens projetadas são apresentadas ao público e o código do software é projetado em conjunto com seu resultado visual, como em uma performance de *live coding* [Collins et al., 2003]. Além disso, dois programadores ficam encarregados de modificar o código e as projeções visuais em tempo real (Figura 12).

Há também uma aplicação que é disponibilizada para os membros da audiência, permitindo que os mesmos tirem fotos da performance, as quais serão utilizadas nas projeções. A camada visual conta ainda com *webcam's* e com processamento de imagens

[2] A filosofia da caixa preta é um texto provocativo, que visa desacomodar clichês e criar inquietude, sobretudo naqueles que pretendem produzir imagens com um mínimo de consciência crítica.

Figura 12: Projeções visuais d' "Chaos das 5".

em tempo real, as quais capturam os artistas e o público, completando a configuração visual d' "O Chaos das 5".

Por fim, a camada musical conta com 2 tipos de músicos: 5 integrantes do grupo, os quais tocam diversos instrumentos eletrônicos, criados em JavaScript / HTML5 e PureData, utilizando como controladores mouses, teclados, joysticks e um pequeno controlador MIDI, além de guitarras elétricas com processamento de áudio através do PureData e GuitarRig; e o público, que toca 3 diferentes instrumentos eletrônicos em seus dispositivos móveis, disponibilizados através de uma rede wi-fi local.

4.1.1 Formas de interação de "O Chaos das 5"

O pré início de nosso espetáculo copia o modelo utilizado por diversos artistas, como Gilberto Mendes em sua peça Santos Footbal Music [Chagas, 1992], ou seja, com ajuda dos performers, o público vai aprendendo a interagir cada vez mais com a performance. Este pré início conta com a projeção de um cronômetro, implementado através do PureData, que fica responsável por marcar uma contagem regressiva para o início do espetáculo (Figura 13).

Esta projeção indica aos participantes como acessar a rede sem fio e o website da performance. Inicialmente este website foi implementado através da linguagem de programação JSP (Java Server Pages) a qual permite unir códigos em Java, Javascript e HTML. Além de ser executado em um servidor Apache Tomcat, com o auxílio de um roteador comum/doméstico e uma máquina com 8gb de RAM com um processador i5 de quinta geração. Durante as primeiras apresentações, a aplicação mostrou-se bastante

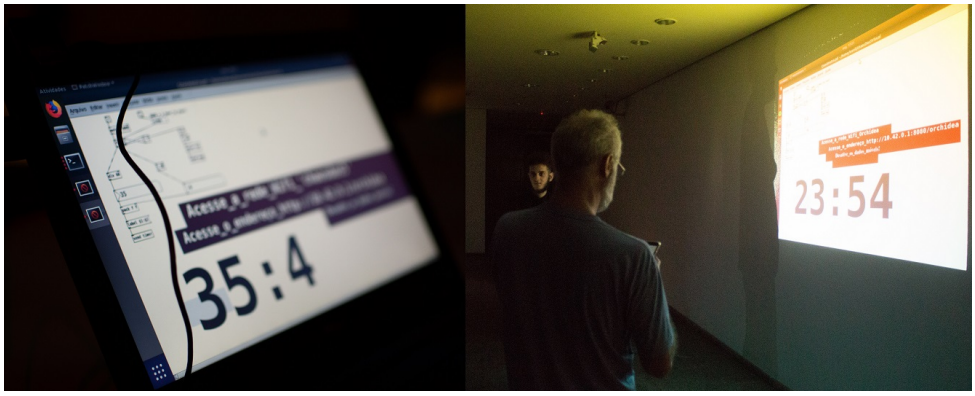

Figura 13: Pré início do espetáculo "O Chaos das 5".

lenta, mesmo com poucos acessos de usuários. Consequentemente, a aplicação necessitou de ser reprogramada, porém desta vez, através da linguagem Javascript / HTML com o auxílio do servidor Apache nativo do Linux, sem o uso do Java. Por fim, a aplicação passou a funcionar como esperado, atendendo à baixa demanda, com baixa latência.

Durante a tarefa de acessar o site proposto, os artistas começam a interagir com o público oferecendo xícaras de chá como em uma performance teatral, além de auxiliar as pessoas a utilizarem seus dispositivos móveis de maneira correta. Este tipo de informação é bastante importante pois certa parte do público pode ter dificuldades com estas questões tecnológicas [Hindle, 2013].

Ao acessar o website, o participante entra em uma espécie de jogo em que ele precisa recuperar 3 senhas que ficam expostas ao redor do ambiente performático. A interação nesta etapa da performance é totalmente física / digital, visto que o público é "forçado" a se locomover pelo ambiente até achar todas as senhas escondidas, as quais são utilizadas nos dispositivos móveis. Após achar as senhas, o nome do usuário / participante é requisitado. Estes dados são armazenados em um servidor local, e por serem dados simples, contendo apenas os nomes dos participantes, eles acabaram sendo gravados em um arquivo de texto, sem o uso de um banco de dados. Estes nomes são utilizados para a preparação dos créditos finais, os quais englobam os nomes de todos os usuários que acessaram a aplicação e consequentemente influenciaram no espetáculo, também sendo reconhecidos no final da performance como parte do elenco.

Após inserir o nome no sistema, o usuário fica aguardando em uma página estática que possui um contador sincronizado com o cronômetro. Quando o contador / cronômetro zera, todas as aplicações do lado do cliente enviam uma requisição de redirecionamento de página, através da linguagem PHP, para que os IMD's implementados sejam liberados para o público. Ao todo, 3 tipos de IMD's são disponibilizados (Figura 14).

Estes instrumentos fornecem ao público a capacidade de influenciar diretamente na performance, visto que eles também fazem parte da paisagem sonora. Para o desen-

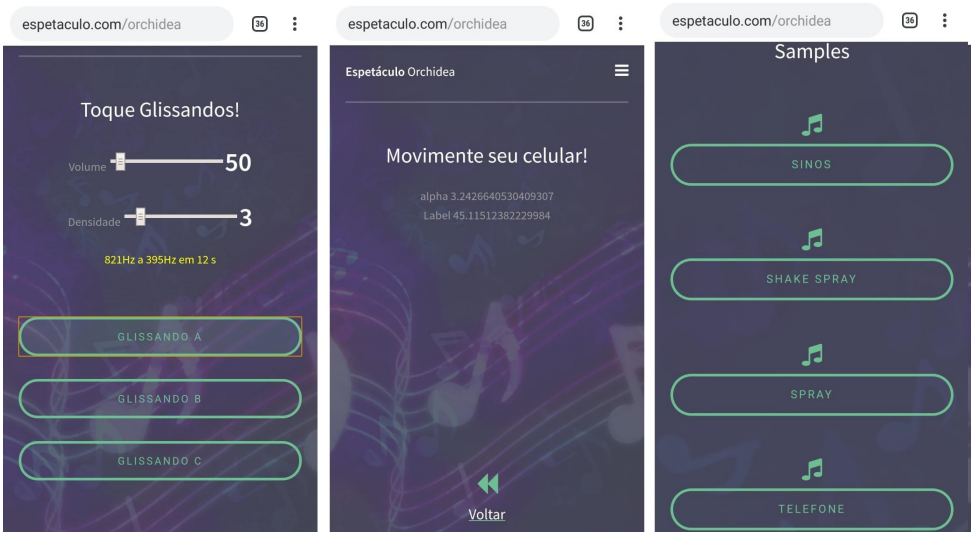

Figura 14: IMD's do "O Chaos das 5"

volvimento desta aplicação optou-se por utilizar a linguagem HTML / Javascript que, devido à sua portabilidade, o público pôde acessá-la via *browser* sem a necessidade de instalar um aplicativo, os quais poderiam causar problemas referentes à espaço de armazenamento e acesso à internet. Os instrumentos foram implementados visando simplicidade e usabilidade para que eles pudessem ser tocados por leigos em música, evitando com isto, explicações técnicas exaustivas e desnecessárias sobre como executar o IMD [Bin et al., 2016, Lee and Freeman, 2013].

Assim que os instrumentos são disponibilizados ao público e o cronômetro zera, a primeira etapa dentre as três existentes na performance se inicia. Esta etapa possui uma duração de aproximadamente 9 minutos, nos quais é construído um contínuo e lento aumento de tensão. A camada sonora passa a ser composta por sons sintéticos, com a predominância de um material musical formado por glissandos, os quais foram inspirados pela *Metastasis*, de Iannis Xenakis, representando a vertigem da queda livre. O público fica encarregado por configurar e tocar diferentes tipos de glissandos (primeiro instrumento da Figura 14), e posteriormente, no ápice desta etapa, tocar um som melódico / cromático, sintetizado a partir do acelerômetro dos dispositivos móveis, fazendo com o que diferentes movimentações sintetizem diferentes melodias (segundo instrumento da Figura 14). O gesto final desta parte é constituído por uma interrupção abrupta liderada pelos integrantes do grupo que silenciam simultaneamente e sincronizados com a projeção da imagem da Alice[3] pedindo silêncio a todos.

Já na segunda etapa, a qual possui duração de 10 minutos, o som do ambiente passa a simular uma imersão na paisagem audiovisual da vida cotidiana nas grandes cidades

[3] Personagem principal da obra "Alice no país das maravilhas" de Lewis Carroll. Este personagem cai numa toca de coelho e é transportado para um lugar fantástico povoado por criaturas peculiares e antropomórficas, revelando uma lógica do absurdo, que é uma característica dos sonhos [Carroll, 2018].

ressaltando sons concretos, formando uma paisagem sonora tensa, irregular e agressiva. Ocorre então um diálogo sonoro em 5 camadas: o som da cena urbana projetada em vídeo; um improviso com amostras disparadas e manipuladas em um instrumento criado para o espetáculo no PureData e controlado por joystick's; 2 guitarras elétricas fazendo intervenções pontuais e improvisadas; uma camada com o disparo de múltiplos samples a partir dos dispositivos móveis do público, representando sons de sirenes, sinos, automóveis, buzinas, carros, entre outros (terceiro instrumento da Figura 14).

Por fim, a terceira etapa se baseia em um ambiente de calma e surrealidade que é evocado por meio de uma citação distorcida do V movimento do *Quarteto para o fim dos tempos*, de *Olivier Messiaen*. Nesta etapa, o público não participa da camada sonora.

4.1.2 Avaliação da participação do público

Analisando a interação do público sob a arquitetura proposta no Capítulo 3 pelo ponto de vista tecnológico, na performance "O Chaos das 5" existem 3 tipos diferentes de interação, dentre elas (Figura 15): fornecimento de IMD's ao público; captação de senhas, fotos e nomes dos participantes; e a captação de gestos do público/performer.

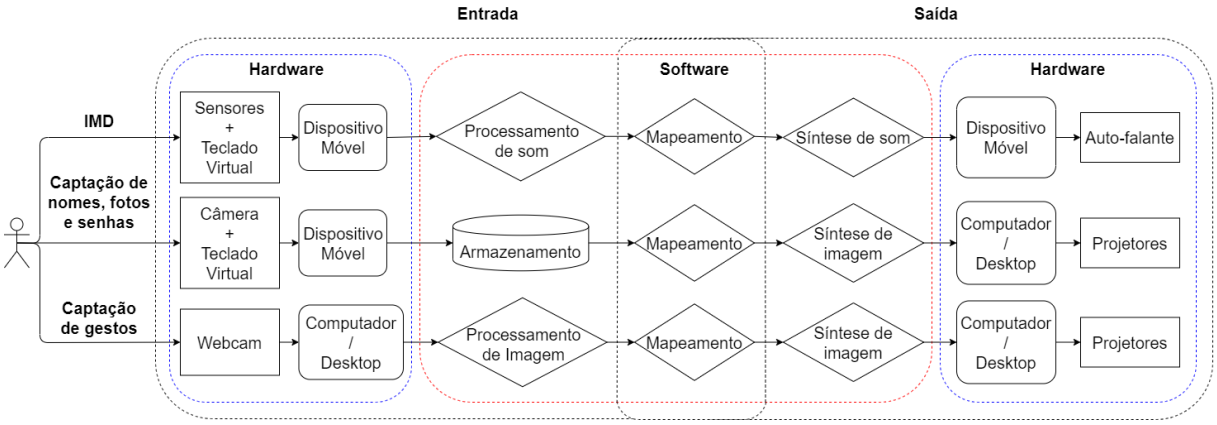

Figura 15: "O Chaos das 5" sob a arquitetura proposta

Em relação aos IMD's, do ponto de vista do hardware utilizado, a primeira interação do participante com o sistema é dada a partir de sensores, como o acelerômetro, e também a partir do próprio teclado virtual, o qual é utilizado para colectar informações como qual IMD será escolhido e quais serão suas respectivas configurações. Estes dados são processados nos dispositivos móveis, os quais utilizam um servidor hospedado em um computador / desktop apenas para disponibilizar os IMD's, o que justifica a não representação do desktop na arquitetura. As informações captadas pelo hardware são utilizadas na camada de software a qual ficará responsável por processar o som do IMD, mapeá-lo e depois sintetizá-lo. Neste caso, o mapeamento é classificado como um-para-um, visto que cada glissando, sample ou som produzido pelo acelerômetro do dispositivo móvel, gera

apenas uma síntese sonora. O som sintetizado é então reproduzido nos alto-falantes dos dispositivos móveis, representados na saída da arquitetura.

Os nomes e as senhas também são captados a partir do teclado virtual dos dispositivos móveis, e as fotos a partir da câmera do celular. As senhas são mapeadas diretamente, onde uma senha redireciona o usuário para uma página específica, proporcionando um feedback inteiramente visual. Já os nomes e as fotos são armazenados localmente.

Por fim, a captação de gestos na arquitetura é dada a partir de uma *webcam* ligada a um computador / desktop, o qual irá processar a imagem e aplicar diferentes efeitos sob um mapeamento muitos-para-muitos, visto que um gesto captado pode resultar em diferentes tipos de efeitos e vice-versa. E por fim, o feedback proporcionado a partir da captação de gestos acaba sendo unicamente visual. Todos os mapeamentos são estáticos, visto que eles não se modificam no decorrer da performance.

Seguindo as métricas demonstradas na subseção 2.5, podemos avaliar que a participação do público em nosso espetáculo possui um certo **grau de liberdade de controle** pois os instrumentos poderiam ser acessados a qualquer momento, em qualquer etapa. No entanto, em nossa primeira apresentação, tal grau de liberdade causou um pouco de desconforto devido à utilização dos instrumentos em momentos inesperados da performance, gerando ruídos indesejáveis durante o espetáculo. Por esta razão, decidimos diminuir esta liberdade para permitir que o usuário pudesse escolher entre apenas alguns instrumentos em uma determinada etapa, bloqueando os demais para que não fossem utilizados fora de hora.

A **versatilidade do sistema** foi dada de acordo com o esperado tendo sido identificado apenas problemas em relação à capacidade da rede. Diversos trabalhos relacionados utilizam a participação do público via telefones celulares em que o celular é caracterizado como um Instrumento Musical Digital (IMD), como é o caso dos trabalhos "*12*", do Radius Ensemble [Egozy and Lee, 2018], SWARMED [Hindle, 2013] e CROWDIN C[LOUD] [Junior et al., 2016]. Este tipo de aplicação geralmente é controlado por um servidor, o que pode causar problemas referentes à escalabilidade, caso o sistema seja acessado por uma quantidade de usuários que vai além do suportado. A escalabilidade de nosso servidor prejudicou a versatilidade do sistema e ainda é um desafio a ser superado. Possivelmente para solucionar esta questão da escalabilidade, pudemos notar que em boa parte dos trabalhos relacionados limitou-se a quantidade de participantes do público, a fim de evitar problemas deste tipo.

Quanto à **transparência na interação com o público**, o sistema implementado acabou sendo bastante adequado. Isto ocorreu pois os celulares foram utilizados como instrumentos e cada instrumento tinha um som muito característico, o que permitia aos participantes uma relação bastante causal entre o gesto e o som. Notamos que foi necessário apenas que as instruções sobre como se conectar na rede e como utilizar a apli-

cação fossem informadas corretamente ao público, pois a falta de instruções pode causar desconforto no público que deseja participar da peça mas que não sabe como proceder para realizar esta tarefa.

A distribuição da interação do público foi alcançada por meio da existência de mais de um instrumento para cada momento da peça. Quanto ao **foco sob a performance**, este foi um ponto bastante crítico em nosso sistema. Parte do público se distraiu tentando tocar e não conseguiu prestar atenção na performance.

Em relação ao **design sonoro**, proposto por Mamedes et al. [2015], toda a performance foi desenvolvida visando 3 cenas distintas e bem definidas, onde o ambiente sonoro se modifica de acordo com a cena, aprimorando a experiência estética dos participantes durante a performance: primeira etapa composta por sons sintéticos, com a predominância de um material musical formado por glissandos; segunda etapa realizando uma imersão na paisagem audiovisual da vida cotidiana nas grandes cidades ressaltando sons urbanos, formando uma paisagem sonora tensa, irregular e agressiva; e a terceira etapa simulando um ambiente de calma e surrealidade através de uma citação distorcida do V movimento do Quarteto para o fim dos tempos, de Olivier Messiaen.

Além disso, por utilizarem seus dispositivos móveis, os quais são bastante eficazes para o desenvolvimento de IMD's, do ponto de vista sonoro por parte do público, as condições técnicas foram bastante favoráveis para imersão dos participantes na camada sonora da performance. Por fim, durante boa parte da performance o público ficou acessando seus dispositivos móveis para produzir algum tipo de som, o que trouxe uma certa utilidade efetiva e prática aos IMD's desenvolvidos.

4.1.3 Lições aprendidas

A ideia de infraestrutura local, inspirada no trabalho de Hindle [2013], a partir de um servidor web, um ponto de acesso sem fio e um servidor DNS causou vantagens e desvantagens durante as apresentações do "Chaos das 5".

Diferente de performances como "*12*" [Egozy and Lee, 2018], no "Chaos das 5" não havia um palco principal e a participação do público não era limitada, visto que cada participante tinha liberdade de fazer o que bem quisesse, a qualquer momento. Como nossa atuação aconteceu em um espaço aberto, era impossível saber previamente o número de participantes. Isto gerou um problema, já que roteadores sem fio possuem um número limitado de clientes conectados e esta informação normalmente não é fornecida no manual e não é fácil testar a limitação de desempenho deste tipo de equipamento de rede. Além disso, se o dispositivo atinge o limite de clientes conectados, outros clientes que tentam se conectar continuam enviando mensagens para o dispositivo, fazendo com que ele fique bastante instável. A primeira lição aprendida é a de que dispositivos de rede possuem

limitações e é importante conhecê-las antes, visto que durante a primeira apresentação, foi utilizado um roteador doméstico, o qual não suportou o público, que não passava de 20 participantes.

A aplicação do "O Chaos das 5" foi disponibilizada tanto em um servidor local, quanto em um servidor hospedado na Universidade Federal de São João del-Rei, o qual poderia ser acessado pela internet. Quando o limite de conexões locais foi atingido, alguns membros do público tentaram acessar os instrumentos via 3G. Como a URL da aplicação local não era a mesma da aplicação em nuvem, o esforço de informar ao público sobre como acessar os instrumentos teve de ser realizado novamente. Ao final da apresentação, vários participantes pediram para acessar o servidor para experimentar os instrumentos, visto que durante a performance eles não conseguiram acessá-los. Ou seja, a URL do servidor local, deveria ser a mesma URL utilizada para o servidor em nuvem e isto permitiria que a performance continuasse mesmo após o seu término, com os usuários conectados em nosso servidor externo.

Outro comportamento inesperado aconteceu devido à nossa configuração de DNS. Configuramos o servidor DNS para resolver qualquer endereço em nosso servidor. Assim, mesmo que um participante tenha digitado o endereço errado, ao estar conectado em nossa rede local, ele seria obrigatoriamente redirecionado para os instrumentos. Inicialmente pareceu ser uma boa ideia, mas não percebemos que todo o tráfego de rede dos dispositivos participantes foi redirecionado diretamente para o nosso servidor. Os participantes tentaram utilizar a Internet e aplicativos como Netflix, Youtube, Whatsapp e Facebook, todos sendo respondidos pelo servidor DNS hospedado na mesma máquina em que os instrumentos eram disponibilizados. Isso criou um grande desperdício de processamento e um gargalo para a solicitação real, o que trouxe muita instabilidade ao nosso sistema. Neste caso, como a máquina utilizada tinha um poder de processamento baixo, a utilização do DNS acabou prejudicando ainda mais a fluidez do sistema.

Outra lição foi aprendida sobre a utilização do espaço. Não queríamos definir onde o público estaria e onde a performance aconteceria. Tudo que tínhamos definido foi onde colocar músicos, projetores e codificadores ao vivo com *notebook's* para as realizar as projeções ao vivo. Quando a performance começou e ligamos os projetores, esse arranjo acabou definindo o espaço da audiência. Obviamente, as pessoas evitavam ficar na frente das luzes dos projetores e em pouco tempo tínhamos as fronteiras indesejáveis entre o espaço público e o espaço de atuação.

Todas estas lições/experiências vivenciadas nos deu algumas dicas do que mais pode ser explorado em apresentações futuras. Percebemos que também podíamos explorar as luzes do celular mudando a cor de fundo do nosso aplicativo web para criar uma penumbra de iluminação das cenas, principalmente na terceira parte da performance. Pode ser interessante também usar o flash do celular para criar um efeito de luz estroboscópica

com a ajuda do público ou até mesmo utilizar feedback's ou interações hápticas a partir de diferentes vibrações nos dispositivos móveis.

Por fim, quatro aspectos que devem ser levados em consideração no desenvolvimento de uma performance voltada para a satisfação do público: na criação de um IMD, a interface utilizada deve ser simples e amigável para que o público fique satisfeito e consiga acessar as funcionalidades do sistema de maneira correta [Lee and Freeman, 2013]; a utilização de um portal captive facilita e simplifica o processo de redirecionamento do público para o IMD criado [Hindle, 2013] ou até mesmo a utilização de um DNS, caso a máquina tenha um alto poder de processamento; a interação do público com a performance, a partir de votos capazes de modificar aspectos sonoros, parece não ser uma boa ideia em relação à satisfação do público [Lee and Freeman, 2013]; além disso, em uma performance a explicação técnica do IMD utilizada pode acabar sendo uma ação exaustiva e desnecessária [Bin et al., 2016].

4.1.4 Análise Geral

"O Chaos das 5" foi a primeira tentativa de criar uma performance digital pelo nosso grupo de pesquisa. As possibilidades de criação de instrumentos com o webaudio nos motivaram a utilizar os celulares do público como alto-falantes difundidos que participam ativamente da paisagem sonora da performance. Certamente, tivemos alguns problemas técnicos para enfrentar e encontramos várias soluções aprendendo diferentes lições, conforme apresentado na subseção acima.

Devido a todos esses problemas, em especial os problemas causados pelo número de participantes, certamente algumas pessoas não conseguiram se conectar aos instrumentos e isso resultou em um efeito colateral interessante. Em algum momento da apresentação, as pessoas não sabiam se deviam tentar usar o celular e participar da apresentação, ou apenas assistir, prestando atenção no que estava acontecendo. Resultou em uma metáfora interessante da vida contemporânea, onde a ansiedade de estar conectado o tempo todo às vezes nos impede de observar o que está acontecendo ao nosso redor.

Na performance do espetáculo, conseguimos nos apropriar dos celulares do público. Por cerca de meia hora, as pessoas se mantiveram desconectadas da realidade externa ao espetáculo mesmo tendo em mãos a principal ferramenta de comunicação da atualidade. Isto permitiu que a imersão no espetáculo fosse ainda maior pois, utilizar o celular para tocar impediu, por exemplo, que as pessoas tentassem fotografar o espetáculo ao invés de assisti-lo. A chance do protagonismo e o acesso à ferramenta foi a chave para trazer esta outra experiência para o público. Esperamos, por fim, que esta experiência sirva como metáfora de como esta ferramenta pode servir para as pessoas agirem de maneira ativa nos tempos atuais.

A participação do público foi uma experiência social interessante e, de alguma forma, em um nível inesperado para nós. Alguns membros da audiência relataram que gostariam de tirar fotos do programa, mas não puderam, pois estavam usando o celular para reproduzir o IMD. Houveram participantes que solicitaram carregadores de celular, com medo de ficar sem bateria antes do final da apresentação e também houveram reclamações sobre não conseguir utilizar a Internet durante a performance. A experiência de ter o dispositivo móvel "coletado" por um tempo certamente teve um impacto social em boa parte do público.

A criação da performance envolveu 20 pessoas de diferentes áreas, conhecimentos e competências. Além disso, pessoas com diferentes níveis de experiência em apresentações ao vivo. Mesmo assim, as diferenças aqui foram somadas para tornar viável a criação da performance. A integração de uma grande equipe interdisciplinar e heterogênea foi uma experiência incrível para todos os participantes.

A participação do público completa o cenário transdisciplinar da obra e nos permite trazer ainda outras experiências para o criação artística assim como tratar o conhecimento e o desconhecido durante a performance musical. Se por um lado o espetáculo utiliza referências já estabelecidas, como a "Alice no país das maravilhas", por outro apresenta uma roupagem performática, atual e tecnológica, com assuntos atuais e pertinentes, como a repressão, o autoritarismo, o território e o cotidiano urbano.

4.2 Instalação Digital Per(sino)ficação

A cultura do sino é uma tradição secular que estava e ainda está fortemente ligada às atividades religiosas e sociais das antigas aldeias do período colonial brasileiro. Os sinos tocando, criados a partir de diferentes padrões rítmicos e timbrais, são objetos semióticos que estabelecem uma linguagem capaz de transmitir vários tipos de mensagens para a população local [Dangelo and Brasileiro].

Em São João del-Rei, Minas Gerais, a tradição única de sinos presente na composição da paisagem sonora da cidade e a forte relação afetiva e identitária de parte da população com esses instrumentos sonoros, acabou fazendo com o que a cidade de São João del-Rei passasse a ser conhecida como "a cidade dos sinos" ou "a cidade onde os sinos falam" [Morandi, 2014]. Essa afeição pelos sinos produz, em certa escala, uma personificação deles, tornando-os algo mais que objetos inanimados [Rosa et al., 2006].

Cada período do calendário católico é marcado por diferentes ritmos, onde cada igreja utiliza uma quantidade diferente de sinos, variando de um a três deles que podem ser combinados pelos compositores para criar todos os diferentes padrões presentes na cidade. Com base nessa tradição e como forma de enfatizar essa conexão de maneira artística

contemporânea, foi desenvolvido uma instalação digital chamada de Per(sino)ficação[4].

4.2.1 A instalação

Nesta instalação, o público é convidado a entrar em uma cabine com o objetivo de representar um sino da cidade. Na cabine, algumas projeções de diferentes torres e sinos das igrejas da cidade são projetadas. Até então, uma paisagem sonora calma da cidade é o único som reproduzido. Ao entrar na cabine, são captados a altura e largura do participante que ao apertar um botão, aciona a síntese sonora de um determinado sino com diferentes padrões e ritmos, completando a paisagem sonora.

A natureza da composição varia de acordo com as características do visitante. Estas características são captadas através de uma *webcam* que utiliza algoritmos de Visão Computacional, que em seguida, mapeia as informações coletadas para criação de um novo sino, a partir da síntese sonora. Vale ressaltar que o participante, de acordo com o gesto realizado, pode mudar drasticamente o timbre gerado.

Por fim, a junção dos sinos sintetizados com uma paisagem sonora de fundo, acaba simulando um ambiente urbano na instalação, criando uma atmosfera mais realista com base no cotidiano das pessoas [Schafer, 1993], aprimorando ainda mais a interação do público com a instalação.

4.2.2 Aspectos Técnicos

Este trabalho é baseado em uma aplicação desenvolvida em python, a qual fica responsável por coletar dados da audiência, via openCV, e enviá-los, via protocolo Open Sound Control (OSC), ao PureData que por sua vez, gera samples de sinos, os quais complementam a paisagem sonora (Figura 16).

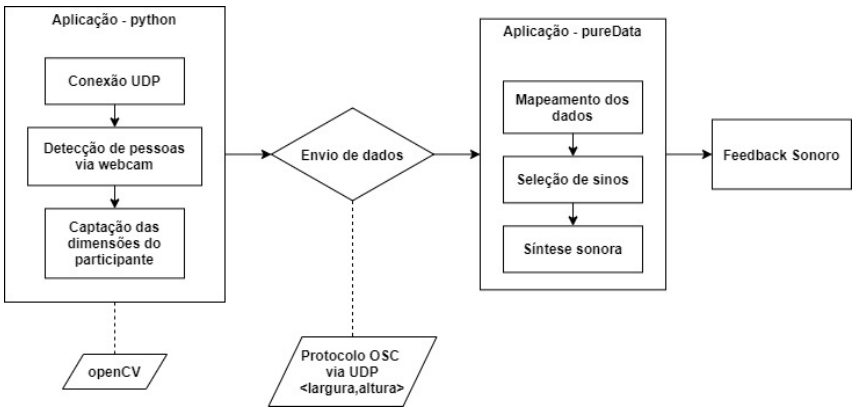

Figura 16: Fluxograma da instalação Per(sino)ficação.

[4] https://www.youtube.com/watch?v=M3VWWvjHl3k&feature=youtu.be&ab_channel=F%C3%A1bioPassos

Inicialmente, uma conexão UDP sob a porta 9001 é estabelecida localmente através da aplicação em python, para possibilitar o envio de dados para a aplicação desenvolvida no PureData. Logo após esta tarefa, através da biblioteca openCV, a *webcam* fica responsável pela constante captação e detecção de pessoas, além de coletar suas respectivas dimensões, sendo elas a altura e largura, até que um botão seja pressionado.

Quando o botão é acionado pelo participante, seus dados são enviados, através do protocolo OSC, para o PureData, que os recebe através do patch demonstrado na Figura 17.

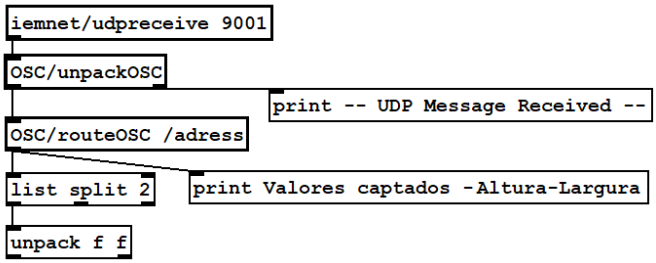

Figura 17: Captação dos dados do participante através do PureData.

O valor referente à altura do participante é utilizado para escolher qual ritmo será tocado. Ao todo são três ritmos distintos, onde cada um engloba 5 repetições de um sino gerado através da síntese aditiva[5], porém com diferentes *delay's*. Além da escolha do ritmo, a altura, em conjunto com a largura, também fica responsável por selecionar a frequência e a duração do sino a ser gerado, onde a altura é utilizada para frequência, e a largura para duração. Este patch[6] possui um conjunto de osciladores que realizam a síntese aditiva para simular um sino, criado por Jean-Claude Risset, o qual pode ser alterado em relação à sua frequência, duração e amplitude. Após o toque dos sinos, o usuário completa sua experiência com a instalação e o sistema fica aguardando até que o botão seja acionado novamente.

4.2.3 Avaliação da instalação

Por possuir apenas uma forma de interação, o Per(sino)ficação sob a arquitetura proposta acaba sendo um pouco mais simples, quando comparado com o "O Chaos das 5". Inicialmente, o participante fica encarregado de acionar um botão a partir do teclado do computador / *notebook* fornecido (Figura 18).

Logo após esta interação, a imagem do participante começa a ser coletada a partir da *webcam* a qual detecta a presença de pessoas no ambiente. A imagem é então processada na camada de software, a qual fica responsável por extrair as características físicas do

[5] A síntese aditiva é uma técnica em que sonoridades complexas são unificadas a partir de osciladores que precisam ser utilizados para compor um som.

[6] http://msp.ucsd.edu/techniques/latest/book-html/node71.html

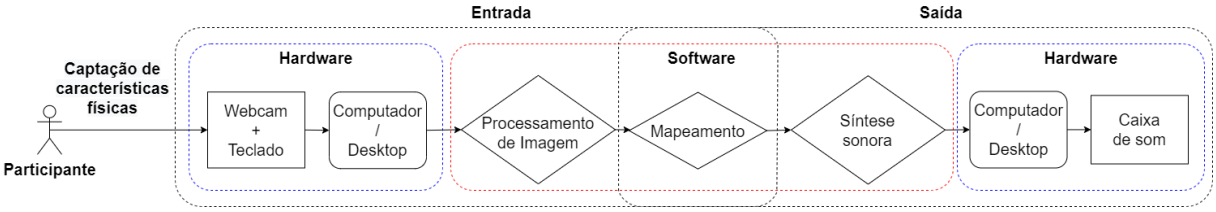

Figura 18: Per(sino)ficação sob a arquitetura proposta.

participante. Estas características são então mapeadas através do mapeamento um-para-muitos, visto que a largura influencia na duração e a altura influencia tanto no ritmo quanto na frequência do sample de sino a ser sintetizado. Logo após o mapeamento, o som é sintetizado e reproduzido nas caixas de som do *notebook*.

Este trabalho foi apresentado no 17th Brazilian Symposium on Computer Music (SBCM) em 2019, na cidade de São João del-Rei. A primeira execução deste trabalho acabou fornecendo um feedback positivo, relacionado tanto à interação do público com a máquina, quanto seu respectivo engajamento com a aplicação (Figura 19).

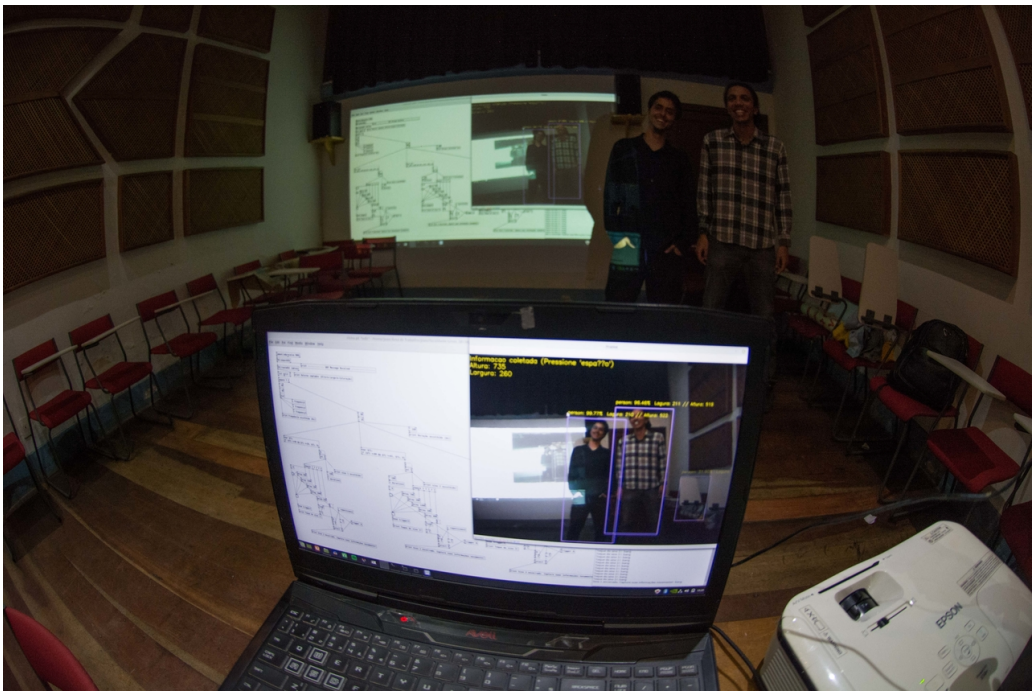

Figura 19: Apresentação da instalação Per(sino)ficação (SBCM 2019).

Devido ao fato de utilizar uma conexão, a qual é estabelecida localmente sem o uso de roteadores ou switches, além de não lidar com problemas de escalabilidade, problemas referentes à rede, como os citados na subseção 4.1.2, foram evitados. Na performance "O Chaos das 5", além dos problemas referentes à rede, houveram problemas de compatibilidade, visto que a aplicação foi desenvolvida para funcionar em dispositivos móveis, os quais operam sob diferentes navegadores. Já na per(sino)ficação, a aplicação foi desen-

volvida para ser executada especificamente em apenas um computador, o que facilitou a etapa de testes e garantiu sua fluidez durante a apresentação da instalação.

Após a apresentação da instalação, algumas propostas para aprimorar a interação do público foram definidas, dentre elas: aumentar a liberdade do público sob os gestos realizados, definindo um mapeamento muitos-para-muitos para captar o gesto realizado, e não a altura e largura do participante quando o botão for acionado, podendo ser implementado até mesmo através de redes neurais; fazer com o que o participante consiga influenciar não apenas na paisagem sonora, mas também nas imagens projetadas; elaborar timbres que remetam à sinos e padrões rítmicos reais presentes na cidade de São João del-Rei, os quais podem ser utilizados para fornecer um feedback ao participante, contendo informações históricas e musicais sobre o sino gerado.

A interação do público com a instalação foi promissora, dado que os participantes interagiram com a máquina através de gestos, os quais eram captados na forma de diferentes alturas e larguras, gerando sons de sinos que por sua vez, os engajava a realizar outros gestos, com o objetivo de observar a versatilidade do sistema em gerar tipos distintos de sinos ou apenas apreciar a paisagem sonora gerada (Figura 20).

Figura 20: Captação de características físicas do público em Per(sino)ficação (SBCM).

Em relação às métricas de avaliação propostas por Mazzanti et al. [2014], pode-se dizer que a **liberdade do participante** na instalação foi bastante limitada, visto que na Per(sino)ficação só há um tipo de interação. O sistema desenvolvido e sua respectiva forma de interação trouxe aos participantes um certo conforto, visto que eles se movimentavam, realizavam diferentes gestos e já recebiam o feedback sonoro em tempo real, sem necessitar

de configurar algo ou tocar determinados botões, como ocorre na utilização de IMD's. Por prover um tipo de interação individual, a interação dos participantes com o sistema foi bastante clara e simples de se entender, além de ser dada em apenas um local, visto que para interagir com o sistema o participante precisava ficar em frente à *webcam* durante toda a experiência.

O **foco** da instalação foi dado sob dois locais, englobando a área em que o participante realizava seus gestos em frente ao *notebook* e o projetor que demonstrava a progressão da interação. Em Per(sino)ficação existem dois tipos de participantes, dentre eles: o participante ativo, o qual realiza determinados gestos; e o passivo que apenas observa os gestos do participante ativo, podendo influenciá-lo ou não em sua performance, porém recebendo o mesmo tipo de feedback, seja ele sonoro através de sinos sintetizados ou visual através do projetor.

Em relação ao **design sonoro**, apesar de ser uma instalação que possui um sistema claro de interação, que utiliza as condições técnicas favoráveis para uma síntese sonora efetiva, além de ser de fácil utilização, no Per(sino)ficação a experiência estética acaba ficando a desejar, visto que os sinos sintetizados ocorrem sob determinadas frequências as quais não se remetem à sinos reais. Estes sons poderiam ser utilizados em conjunto com uma breve explicação histórica de cada um, melhorando bastante a experiência do participante sob o sentido estético da instalação.

4.3 Instalação Digital Leiamídia

A disputa política brasileira tem se acirrado nos últimos tempos e dentro deste acirramento, o espaço midiático se tornou um ponto central de debate que envolve tanto direita quanto esquerda. É comum encontrarmos críticas dentro de ambos os espectros políticos relacionados à parcialidade da mídia quando uma notícia é passada ao vivo no jornal da TV. Mas será possível ler uma notícia de maneira a não transmitir um sentimento? Na instalação tecnológica Leiamídia[7], o participante é convidado a ser âncora de um telejornal, se passando por um jornalista, o qual fica responsável por informar os telespectadores sobre uma das últimas notícias publicadas no site do G1 sobre política. Uma rede neural convolucional é então utilizada para extrair o quanto de alegria, tristeza ou medo, o participante está demonstrando ao ler um texto sobre determinado assunto. E por fim, de acordo com a classificação realizada, são aplicados diferentes filtros para contextualizar o ambiente em que o participante está imerso com a emoção captada.

[7] https://www.youtube.com/watch?v=YxoEqYJBJpA&t=8s

4.3.1 Funcionamento do sistema

Inicialmente, o participante entra na cabine a qual fica responsável por mostrar a interface inicial do sistema, contendo as últimas notícias publicadas no site do G1, sobre política (Figura 21).

Figura 21: Interface inicial da Instalação Leiamídia.

Todas as notícias são coletadas e atualizadas segundo um *delay*, e ao selecionar uma notícia, o sistema apresenta ao participante a tela secundária, onde as interações começam a ser realizadas (Figura 22).

Figura 22: Exemplo de interação da instalação Leiamídia.

Na parte esquerda da tela, são apresentadas as informações da notícia selecionada,

contendo seu respectivo título e conteúdo. Já na direita, existem dois blocos: na parte superior, uma *webcam* começa a gravar a imagem do participante além de mostrar uma legenda com as últimas notícias, enfatizando ainda mais a ideia de que o participante esteja atuando como um jornalista; e na parte inferior, são impressas as últimas predições realizadas, demonstrando a porcentagem de cada emoção expressada pelo participante.

Após cada predição, são aplicados 3 tipos de filtros de acordo com a maior porcentagem da emoção captada, dentre eles: filtro de medo, onde o brilho da imagem é definido de acordo com a posição do mouse e influencia diretamente nos valores RGB de cada *pixel*, trazendo uma sensação de um ambiente fechado, obscuro e fóbico; filtro de alegria, onde todos os valores RGB dos *pixels* que estão abaixo de um determinado limitante são multiplicados por 4, fazendo com o que as sombras passem a ser representadas por cores mais vivas e fortes, trazendo a sensação de um ambiente mais alegre; e o filtro de tristeza, onde os tons vermelhos e verdes são atenuados e os *pixels* da imagem são agrupados em janelas as quais utilizam a média dos *pixels* de cada grupo, fazendo com o que a imagem passe a ter um efeito bastante parecido com o utilizado em noticiários onde o participante opta por esconder o seu rosto, trazendo talvez, um sentimento de tristeza (Figura 23).

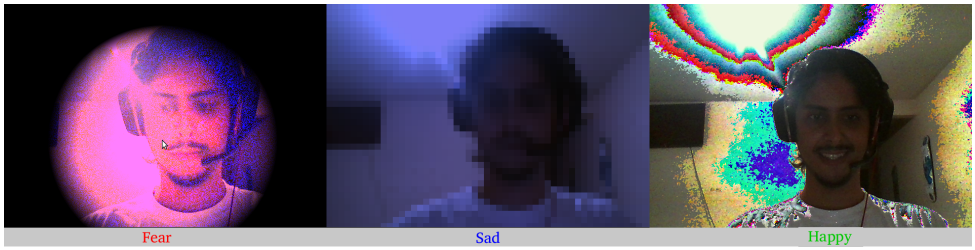

Figura 23: Diferentes efeitos visuais proporcionados pela classificação de sentimentos (Leiamídia).

4.3.2 Aspectos de implementação

Resumidamente, o Leiamídia[8] é composto por 4 módulos principais (Figura 24):

- CNN_emotions: trata-se de uma rede neural convolucional implementada em python, através da biblioteca TensorFlow com o framework jupyter-notebook, o qual é bastante utilizado em aplicações de aprendizado de máquina. Esta rede é composta por 3 camadas convolucionais, seguidas de uma camada densa. Além disso, ela foi treinada a partir do *Toronto Emotional Speech Set* (TESS [Lok, 2019]) para classificar 3 emoções, sendo elas: alegria, tristeza ou medo.

- Notices_crawler: este crawler[9] foi implementado em python, através da biblioteca scrapy e da NewsPlease. Este módulo é responsável por coletar as últimas notícias

[8] https://github.com/joaotaraujo/leia_midia

[9] Crawlers são basicamente uma técnicas de extração de dados utilizada para coletar dados de web sites.

sobre política contidas no site do G1 e é atualizado de acordo com um *delay* que pode ser definido pelo usuário. As notícias coletadas são então armazenadas em um arquivo de texto, o qual será utilizado pelo módulo Processing_app.

- Osc_server: o servidor osc é responsável por aguardar até que o Processing_app envie uma requisição de predição a ele, através do protocolo OSC, contendo o nome do áudio a ser classificado. Após receber uma requisição, este servidor carrega o modelo da rede neural, já treinado, classifica o áudio a partir dele e envia o resultado da classificação de volta para o processing_app.

- Processing_app: esta é a aplicação principal, a qual é responsável por mostrar a interface ao usuário, comunicar com o notices_crawler para atualizar as notícias e com o servidor OSC para enviar as requisições de predição.

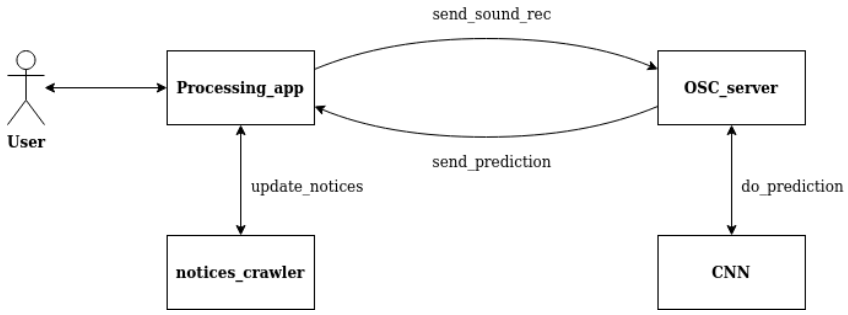

Figura 24: Fluxograma de funcionamento da aplicação utilizada no "Leiamídia".

Ao iniciar a implementação foi necessário decidir como implementar o front-end / interface do sistema, dado as infinitas possibilidades de linguagens de programação, IDE's e API's existentes. De acordo com as performances descritas na Subseção 2.5, vimos que a utilização de aplicativos para dispositivos móveis pode não ser uma boa opção, tanto por questões de portabilidade quanto pelo esforço necessário para instalá-lo.

Tanto a interface quanto o processamento do áudio em conjunto com o classificador poderiam ser implementados em HTML, CSS e javascript. Hoje em dia, o TensorFlow, por exemplo, permite a implementação de redes neurais através do javascript, permitindo a criação / implementação de classificadores web automáticos para diversos fins. Além disso, o fato de utilizar todo o sistema sob apenas um módulo / linguagem poderia melhorar o tempo de resposta da interface com o servidor osc responsável pelas classificações, visto que neste caso, o tempo gasto pela comunicação OSC poderia ser melhor aproveitado.

Porém, para aumentar a flexibilidade do sistema, fazendo com o que ele não seja necessariamente atrelado à apenas uma tecnologia / linguagem de programação, além de reforçar as infinitas possibilidades de criação de arte existentes ao se implementar um sistema com diferentes módulos, surgiu a proposta de utilizar aplicações distintas, as quais se comunicam pelo protocolo OSC.

Desta forma, a linguagem de programação chamada Processing[10] foi utilizada para implementar o front-end do Leiamídia. Por possuir um escopo bastante semelhante ao arduíno, um programa interativo em Processing possui duas funções principais: *setup()*, responsável por definir as configurações iniciais do sistema; e *draw()*, a qual possui toda a lógica do sistema, sendo executada sob *looping* infinito. Esta linguagem acabou sendo bastante utilizada em escolas e universidades por estudantes, artistas, designers e pesquisadores, tanto para fins educacionais quanto criativos no campo da arte digital.

Como o conteúdo das notícias é grande, a comunicação do processing_app com o notices_crawler foi dada através da leitura e escrita de um arquivo de texto compartilhado, visto que trocar todas estas informações via OSC poderia influenciar bastante na fluidez do sistema. Para extrair as notícias do G1, foi utilizado a biblioteca scrapy, em python. Esta biblioteca permite extrair, processar e armazenar dados coletados de web sites. Através do scrapy, o *crawler* fica então responsável por coletar as URL's contidas no G1 [11].

Para extrair as url's, foi necessário um estudo sob a estrutura do site, visto que para coletá-las é preciso informar o nome do seletor / filtro, o qual define o tipo de estrutura / dado que o crawler irá coletar. Para coletar os links das últimas notícias, o crawler buscou por elementos do tipo "<a> ::attr(href)", o qual irá coletar o atributo href de cada elemento com a tag <a> contido no site. Para facilitar a coleta do conteúdo das notícias, foi utilizado a biblioteca NewsPlease, também em python, a qual facilita bastante a extração de informações sobre notícias na web. A partir de uma simples função, fornecendo o link de cada notícia foi possível coletar seu respectivo título, descrição, autores, data de publicação, link da imagem e o conteúdo principal. Estas informações são coletadas e armazenadas em um arquivo de texto, o qual é carregado pelo processing_app.

Além das funcionalidades referentes às notícias, temos o processamento do áudio gravado quando o participante começar a ler uma notícia. O osc_server, implementado em python, fica então aguardando uma requisição de predição, que será enviada via OSC, de tempo em tempo, pelo processing_app, no formato "nome_do_audio.wav". Assim que a mensagem chega no osc_server, ele carrega o modelo da rede neural convolucional já treinado, armazena o resultado em uma string com o padrão "emocaoClassificada-%deMedo-%deAlegria-%deTristeza" e a envia de volta para o processing_app. O processing_app informa ao participante qual a porcentagem de cada emoção captada e aplica 3 tipos de filtros citados anteriormente. Este ciclo se repete e o sistema é reconfigurado quando o participante clica para voltar à pagina inicial das notícias.

Em relação ao classificador, foi necessário dividir sua implementação em 3 etapas: pré-processamento dos dados; treino da rede neural convolucional; e validação do

[10] O processing foi criado em 2001 e é baseado na linguagem Java, sendo assim uma linguagem de mais alto nível, que consegue facilitar a programação de diversos tipos de animações.

[11] https://g1.globo.com/politica/

classificador.

Inicialmente, os dados foram coletados a partir da base de dados TESS. Esta base de dados é composta por um conjunto de 200 palavras ditas na frase "say the word is _____" por duas atrizes, retratando 7 emoções: raiva; nojo; medo; felicidade; surpresa; tristeza; e neutralidade. Porém, neste trabalho, foram utilizados 600 samples para treino e teste, sendo 200 para cada uma das 3 emoções: medo; tristeza; e alegria. E para validação foram gravados mais 20 samples com cada emoção.

Ao ser digitalizado, o áudio passa a ser representado por um *array* contendo a sua respectiva amplitude no tempo. Como os áudios contidos na base de dados TESS possuíam diferentes durações, foi necessário colocar todos os *arrays* no mesmo tamanho. Para isto, o tamanho do maior áudio foi coletado e todos os áudios menores, como o Audio2 da Figura 25, foram complementados com 0's a partir do max_pad_len.

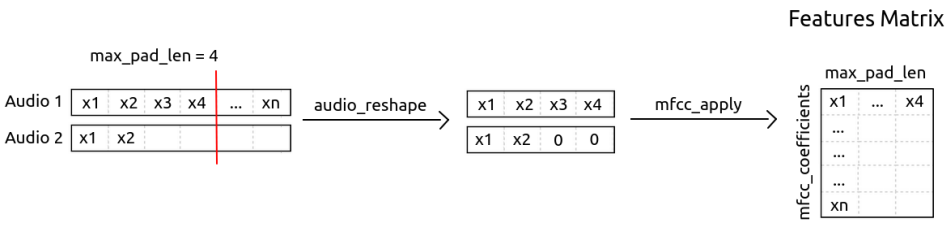

Figura 25: Exemplo abstrato da extração de características de um áudio digitalizado.

Durante a etapa de validação, caso o áudio gravado tenha uma duração maior do que o max_pad_len, como o Audio1 na Figura 25, ele será cortado e analisado até o valor do max_pad_len.

A fim de diminuir a dimensionalidade dos dados e também fazer com o que os dados fiquem mais ricos em informação, foi aplicado o descritor de áudio MFCC, abordado na seção 2.7.2, para extrair 40 coeficientes MFCC. Este descritor geralmente é utilizado para detectar timbres, e foi escolhido pelo fato de que a variação do timbre de uma pessoa pode estar fortemente relacionado a um tipo de emoção, podendo assim, ser classificada automaticamente.

Como saída temos então uma matriz com max_pad_len linhas e 40 colunas. Como as redes neurais convolucionais trabalham com dados bidimensionais, a matriz de features representada na Figura 25 já pode ser processada e classificada. Porém, para obter uma melhor visualização dos dados, sob a matriz de features de cada áudio foi calculado sua respectiva média e desvio padrão, e em seguida plotados em um gráfico (Figura 26).

Os dados coletados pela base de dados TESS foram representados por pequenos círculos e os dados gravados para validação por cruzes grandes. Durante a gravação dos dados de validação, notou-se que a qualidade do microfone influenciava bastante, visto

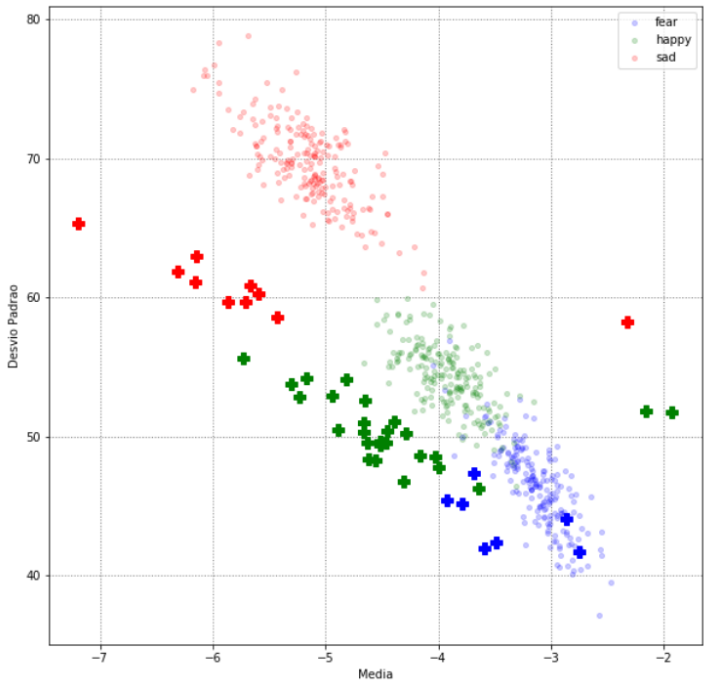

Figura 26: Gráfico dos dados de treino, teste e validação (média x desvio padrão).

que ao utilizar um microfone de qualidade baixa, os dados praticamente não seguiam um padrão, dada uma emoção. Ou seja, a existência de ruídos no som pode influenciar drasticamente no treinamento da rede, fazendo com o que ela aprenda a classificar apenas dados ruidosos e inconsistentes. Por fim, ao utilizar um microfone com melhor qualidade, os ruídos foram praticamente removidos e ao plotar o gráfico, notou-se que os áudios seguiam um padrão e poderiam ser classificados automaticamente.

Após o pré-processamento dos dados foi necessário definir a estrutura e os parâmetros da rede neural. Esta etapa geralmente é feita através de diversos testes variando parâmetros como o número de épocas, batch_size, window_size, quantidade de camadas, tipo do modelo e métrica para avaliação do erro. Por fim, os valores dos parâmetros que tiveram melhor resultado foram: 30 épocas; batch_size de tamanho 64; window_size com tamanho 2; 3 camadas convolucionais seguidas de 1 camada densa; modelo sequencial; e entropia cruzada como função de perda.

Com estes parâmetros, a CNN comportou de maneira satisfatória, onde a acurácia dos dados de teste foi aumentando gradativamente ao longo das épocas, ao mesmo passo em que a perda foi diminuindo, chegando bem próximo de zero (Figura 27).

Por fim, após o treinamento, a rede é salva em um arquivo no padrão *Hierarchical Data Format 5* (HDF5), o qual pode ser carregado para que a CNN comece a realizar predições praticamente em tempo real, visto que ela já possui suas determinadas conexões e não precisa ser treinada após cada classificação.

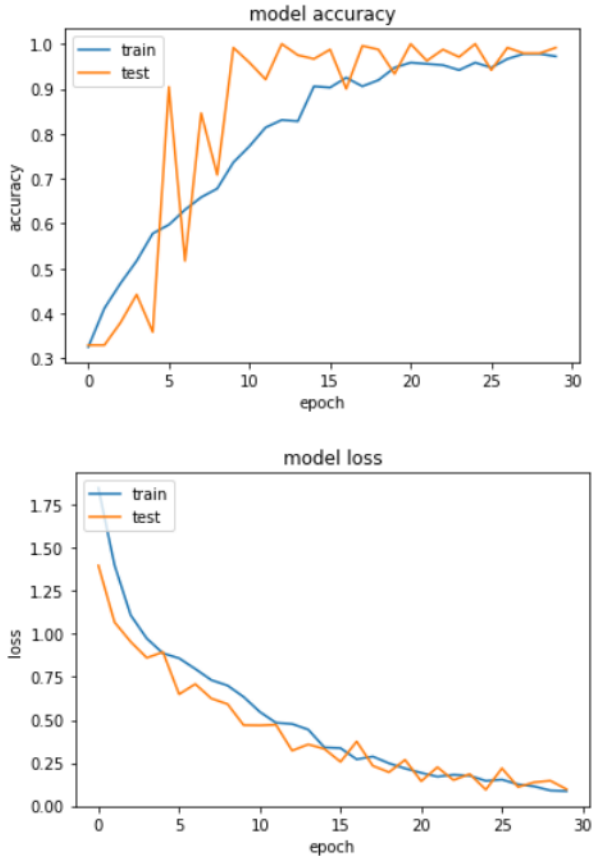

Figura 27: Resultado do treinamento da CNN.

4.3.3 Avaliação da Instalação

A Figura 28 demonstra o funcionamento do Leiamídia a partir da arquitetura proposta no capítulo 3.

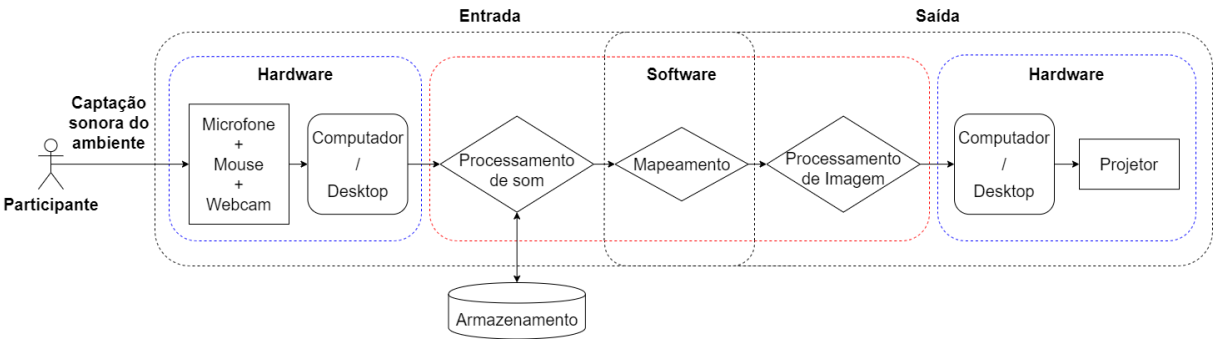

Figura 28: Leiamídia sob a arquitetura de interação proposta.

Inicialmente, o participante interage com a interface do sistema utilizando o mouse para selecionar a notícia a ser lida. A partir deste momento o microfone do computador passa a gravar todo o som reproduzido no ambiente e a *webcam* passa a gravar a imagem do participante. O som captado é então armazenado localmente e pré-processado para ser

classificado pela CNN. Após a classificação, 4 informações são geradas: qual a emoção do áudio captado; porcentagem de medo; porcentagem de alegria; e porcentagem de tristeza. Sob estas informações, o mapeamento é realizado para aplicar efeitos sob a imagem coletada na *webcam*, o qual é definido como muitos-para-um visto que as 4 informações irão influenciar apenas na imagem do participante projetada.

Devido ao COVID-19, esta instalação não pôde ser apresentada, porém é possível avaliar o sistema desenvolvido sob algumas das métricas citadas na subseção 2.5. A **liberdade de controle** proporcionada pelo "Leiamídia" é bastante semelhante à Per(sino)ficação, visto que o participante pode interagir apenas com o computador. Porém em "Leiamídia" existem diferentes interações, como por exemplo a interação sob o efeito de medo que pode ser dada tanto pela voz do participante, quanto pelo mouse que vai definir qual a área da imagem será mostrada. As possibilidades de performance em Leiamídia são bastante simples, visto que o participante pode apenas escolher uma notícia e realizar diferentes tipos de vozes, recebendo apenas um tipo de feedback, que é visual e engloba poucos tipos de efeitos, deixando o sistema bastante previsível. O **foco** e a distribuição do público em "Leiamídia" é totalmente semelhante à Per(sino)ficação, onde o foco é dado no projetor ou na área em frente à *webcam* e os participantes passam a ser definidos como ativo para quem utiliza o sistema e passivo para quem observa a experiência proporcionada.

Por fim, em relação ao **design sonoro**, é possível analisar o som que o participante reproduz. Em "Leiamídia" o principal objetivo é observar se realmente é possível à máquina, classificar qual a emoção de uma pessoa ao ler uma notícia sobre determinado assunto, ou à pessoa, enganar a máquina, forjando determinadas emoções. A utilização de um computador, uma *webcam* e um projetor para classificar a voz de um participante e demonstrar o resultado a ele, acaba sendo mais do que necessária, levando em consideração as mínimas condições técnicas que devem ser utilizadas para este fim.

5 Considerações Finais

Até então, no campo da arte digital, diversos trabalhos artísticos têm sido criados na forma de páginas web interativas, quadros pintados por máquinas, instalações ou performances que captam a interação do público para sintetizar imagens e sons, dentre outros. A discussão sobre o que é arte é bastante antiga e pertinente, mas, independentemente do resultado desta discussão, existem artistas criando obras digitais e existe um determinado público que admira este tipo de obra.

Além de todo o sentido estético e de toda experiência que um trabalho artístico pode proporcionar, o mercado voltado para esta área tem crescido bastante nos últimos anos, o que acaba engajando cada vez mais, pesquisadores a desenvolverem técnicas que forneçam a possibilidade de criação de funcionalidades cada vez mais expressivas, inovadoras e interativas. Através da área de machine learning, por exemplo, é possível criar uma infinidade de sistemas preditivos automáticos os quais podem fazer com o que a máquina se aproxime cada vez mais ao comportamento de um ser humano, possibilitando a criação de diversas formas de interação.

A performance digital "O Chaos das 5" foi o primeiro trabalho desenvolvido no contexto deste mestrado. Ao todo foram mais de 6 apresentações em diferentes cidades, o que acabou influenciando no processo criativo da obra. Todas as vezes em que a performance foi apresentada em uma cidade, a montagem do cenário foi diferente e precisou ser repensada levando em consideração sentidos estéticos, visto que nesta performance não há um palco principal e neste caso, há a intenção de fazer com o que o público se assemelhe, ao máximo, à um performer.

Esta performance conseguiu de fato mostrar a todos os envolvidos, o potencial que a arte digital possui. Durante as apresentações notou-se a existência de um público que realmente aprecia novas experiências e está aberto a conhecer novas possibilidades de utilização de tecnologia em ambientes performáticos / artísticos. Logo após "O Chaos das 5", criou-se a instalação digital "Per(sino)ficação".

Por se tratar de uma instalação a qual é dada em um ambiente fechado, a interação do público com a obra foi dada em menor escala, se comparada a'"O Chaos das 5". Um aspecto interessante da "Per(sino)ficação" está relacionado ao funcionamento da aplicação desenvolvida, visto que ela utiliza uma forma de comunicação entre diferentes algoritmos, a qual pode ser utilizada para desenvolver diferentes funcionalidades sob diversos programas.

Partindo da ideia da comunicação entre diferentes algoritmos, surgiu a possibilidade de se utilizar uma rede neural para processamento de áudio. Atualmente existem

diversas bibliotecas que auxiliam bastante a implementação de diferentes tipos de redes. Desta forma, surgiu a instalação "Leiamídia", a qual demonstrou explicitamente o poder que as redes neurais possuem para criar modelos preditivos automáticos bastante condizentes com a realidade, fazendo com o que seja possível criar formas de interação que aproximem, simplifiquem e tornem ainda mais natural a forma com o que o participante / público se comunica com a máquina. Apesar de já estar totalmente implementada, devido ao COVID-19 esta instalação ainda não pôde ser apresentada.

Em relação à arquitetura criada, o objetivo foi o de formalizar, da maneira mais simples possível, as formas de interação do público com a arte digital, os aparatos tecnológicos e as técnicas utilizadas, de maneira que todas estas questões fossem evidenciados sob um padrão, o qual facilita a distinção de diferentes interações de diferentes obras artísticas. Devido a este objetivo, a arquitetura pipes-and-filters foi utilizada como base para descrição de cada aparato ou técnica utilizado.

Durante o período deste mestrado, foram publicados sete trabalhos relacionados às obras artísticas desenvolvidas, dentre eles:

- "A participação do público em performances artísticas por meio do uso de dispositivos móveis" publicado e apresentado no V Seminário de Artes Digitais, em Belo Horizonte [Araújo and Schiavoni, 2019];
- "Mediação tecnológica e ubiquidade na participação do público em espetáculos de arte sonora" publicado e apresentado na I Conferência Internacional de Pesquisa em Sonoridades, em Florianópolis [Araújo et al., 2019];
- "A technical approach of the audience participation in the performance O Chaos das 5" publicado e apresentado no XVII Simpósio Brasileiro de Computação Musical, em São João del-Rei [Araújo et al., 2019];
- "ALICE: Arts Lab in Interfaces, Computers, and Everything Else - Research report (2019)" publicado e apresentado no XVII Simpósio Brasileiro de Computação Musical, em São João del-Rei [Schiavoni et al., 2019];
- "Per(sino)ficação" publicado e apresentado no XVII Simpósio Brasileiro de Computação Musical, em São João del-Rei [Carvalho et al., 2019];
- "As lições aprendidas com a Orchidea" publicado na revista DEBATES [Schiavoni et al., 2020];
- "Using technology and games to support the audience participation in artistical performance" submetido para a revista PerMusi [Araújo et al., 2021] e ainda no prelo.

Houve também outros trabalhos desenvolvidos no decorrer desta pesquisa:

- "Ha dou ken music: different mappings to play music with joysticks" publicado durante o NIME (International Conference on New Interfaces for Musical Expression, 2019) [Rocha et al., 2019a];
- "Ha Dou Ken Music: Mapping a joysticks as a musical controller", publicado no XVII Simpósio Brasileiro de Computação MusicalRocha et al. [2019b];
- "Harmonia: uma ferramenta de aprendizagem musical" publicado durante a ANPPOM (Associação Nacional de Pesquisa e Pós-Graduação em Música, 2019) [Constante et al., 2019];
- "Harmonia: a MuseScore's plugin to teach music", publicado no XVII Simpósio Brasileiro de Computação Musical **?**
- "Os sinos na era da ubiquidade: a relação entre as tecnologias e o patrimônio histórico" publicado no livro "Poderes do som: políticas, escutas e identidades" [Araújo et al., 2020].

Dentre os trabalhos futuros, temos:

- "O Chaos das 5": criação de um sistema gameficado baseado em *QR Codes*, onde ao acessar a aplicação, o público é induzido a caminhar sobre determinados locais, durante a peça, para coletar *QR Codes* que irão fornecer informações sobre o que fazer, tendo como recompensa final a liberação dos IMD's; a implementação de uma galeria de fotos que pode ser acessada a qualquer momento, a qual possui todas as imagens captadas pelos próprios participantes; a utilização dos dispositivos móveis como parte do sistema de iluminação do espetáculo utilizando, para isto, a lanterna e o flash da câmera;
- "Per(sino)ficação": utilização de timbres mais semelhantes aos utilizados na cidade de São João del-Rei; melhoria no mapeamento das características físicas, utilizando talvez um mapeamento muitos-para-muitos e não muitos-para-um; captação de gestos e não apenas valores como a altura e largura; síntese sonora em tempo real a fim de aumentar a expressividade, tornando o sistema mais dinâmico; síntese de diferentes imagens, as quais também podem sofrer um certo tipo de distorção, de acordo com as informações captadas pela *webcam*;
- "Leiamídia": utilização do dataset completo, abrangendo todas os 8 tipos de emoções fornecidos; a criação de efeitos mais expressivos, os quais poderiam utilizar até mesmo reconhecimento facial em conjunto com a classificação sonora; e a utilização

de mais descritores de áudio para aprimorar ainda mais a acurácia da rede neural utilizada.

Por fim, o autor gostaria de agradecer ao CNPq pelo fomento da bolsa de Pós-Graduação, que tornou este trabalho possível.

Referências

J. T. Allison and T. A. Place. Teabox: a sensor data interface system. In *Proceedings of the 2005 conference on New interfaces for musical expression*, pages 56–59, 2005. Citado na página 51.

X. Amatriain, J. Massaguer, D. Garcia, and I. Mosquera. The clam annotator: A cross-platform audio descriptors editing tool. In *ISMIR*, pages 426–429, 2005. Citado na página 31.

R. A. P. d. Andrade et al. *Lygia Clark: a obra é o seu ato dos Casulos ao Caminhando*. [sn], 2003. Citado na página 18.

M. E. D. Antar. *O clownprovisadorlivre: um estudo sobre interação e performance na livre improvisação musical*. PhD thesis, Universidade de São Paulo, 2016. Citado na página 45.

J. Araújo, A. de Paulo, I. S. Junior, F. L. Schiavoni, M. C. F. Canito, and R. A. Costa. A technical approach of the audience participation in the performance'o chaos das 5'. In *Anais do XVII Simpósio Brasileiro de Computação Musical*, pages 28–34. SBC, 2019. Citado na página 80.

J. T. Araújo, F. L. Schiavoni, I. d. O. S. Junior, R. de Vieira, and M. Almeida. Using technology and games to support the audience participation in artistical performance. *PerMusi, Scholarly Music Journal*, xx(40), 2021. Citado na página 80.

J. T. Araújo and F. L. Schiavoni. A participação do público em performances artísticas por meio do uso de dispositivos móveis. In *Anais do 5o. Seminário de Artes Digitais 2019*, pages 203–211, Belo Horizonte, 2019. Citado na página 80.

J. T. Araújo, F. L. Schiavoni, and F. dos Passos Carvalho. Per(sino)ficação. In *Anais do I Conferencia Internacional de Pesquisa em Sonoridades*, pages 148–150, 2019. Citado na página 80.

J. T. Araújo, F. L. Schiavoni, and F. dos Passos Carvalho. *Poderes do Som: políticas, escutas e identidades*. Insular Livros, 2020. Citado na página 81.

P. Atzori and K. Woolford. Extended-body: Interview with stelarc. *CTheory*, pages 9–6, 1995. Citado na página 20.

S. S. Ballard. Nam june paik, cybernetics and machines at play. 2013. Citado na página 18.

C. Bardiot, M. KUO, L. LIPPARD, and B. O'DOHERTY. *9 Evenings Reconsidered: Art, Theatre, and Engineering, 1966*. MIT List Visual Arts Center, Cambridge, MA, 2006. Citado na página 20.

E. Berdahl and D. Huber. The haptic hand. In E. Berdahl and J. Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 303–306, Baton Rouge, Louisiana, USA, May 2015. Louisiana State University. doi: 10.5281/zenodo.1179022. URL http://www.nime.org/proceedings/2015/nime2015_281.pdf. Citado na página 47.

F. Bevilacqua, R. Müller, and N. Schnell. Mnm: a max/msp mapping toolbox. 2005. Citado na página 51.

S. A. Bin, N. Bryan-Kinns, and A. P. McPherson. Skip the pre-concert demo: How technical familiarity and musical style affect audience response. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 200–205, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University. ISBN 978-1-925455-13-7. URL http://www.nime.org/proceedings/2016/nime2016_paper0041.pdf. Citado 4 vezes nas páginas 25, 46, 59 e 64.

W. Brent. *Physical and perceptual aspects of percussive timbre*. PhD thesis, UC San Diego, 2010. Citado na página 33.

P. M. Brossier. *Automatic annotation of musical audio for interactive applications*. PhD thesis, University of London, 2006. Citado na página 31.

B. E. Bürdek. *Design: History, theory and practice of product design*. Walter de Gruyter, 2005. Citado na página 28.

A. C. Caetano. Interfaces hápticas. *VENTURELLI, Suzete (org.)*, 7:10–15, 2008. Citado na página 47.

J. Cage. *M: Writings' 67-'72*, volume 635. Wesleyan University Press, 1973. Citado na página 18.

B. Calais. A inteligência artificial invade o mundo da arte. https://forbes.com.br/forbeslife/2020/01/a-inteligencia-artificial-invade-o-mundo-da-arte/, 2020. 2020-05-25. Citado 2 vezes nas páginas 38 e 39.

C. Cannam, C. Landone, and M. Sandler. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1467–1468. ACM, 2010. Citado na página 31.

L. Carroll. *Alice Adventures in Wonderland & Through the Looking-Glass*. CreateSpace Independent Publishing Platform, 2018. ISBN 1503250210. URL https://www.amazon.com/Alice-Adventures-Wonderland-Through-Looking-Glass/dp/1503250210?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1503250210. Citado na página 59.

F. Carvalho, F. L. Schiavoni, and J. Araújo. Per (sino) ficação. In *Anais do XVII Simpósio Brasileiro de Computação Musical*, pages 225–226. SBC, 2019. Citado 2 vezes nas páginas 49 e 80.

P. Chagas. A invenção do jogo:"santos football music"de gilberto mendes. *Revista Música*, 3(1):70–81, 1992. Citado 2 vezes nas páginas 19 e 57.

E. Clarke. Empirical methods in the study of performance. *Empirical musicology: Aims, methods, prospects*, pages 77–102, 2004. Citado na página 16.

S. Clinch. Smartphones and pervasive public displays. *IEEE Pervasive Computing*, 12 (1):92–95, 2013. Citado na página 24.

N. Collins, A. McLean, J. Rohrhuber, and A. Ward. Live coding in laptop performance. *Organised sound*, 8(3):321–330, 2003. Citado 2 vezes nas páginas 44 e 56.

R. T. Constante, J. T. Araújo, and F. L. Schiavoni. Harmonia: uma ferramenta de aprendizagem musical. In *XXIX Congresso da Anppom-Pelotas/RS*, 2019. Citado na página 81.

L. Dahl, J. Herrera, and C. Wilkerson. Tweetdreams : Making music with the audience and the world using real-time twitter data. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 272–275, Oslo, Norway, 2011. URL http://www.nime.org/proceedings/2011/nime2011_272.pdf. Citado 3 vezes nas páginas 26, 45 e 46.

A. G. D. Dangelo and V. B. Brasileiro. Sentinelas sonoras de são joão del rei. *Belo Horizonte: Estúdio*, 43. Citado na página 65.

A. C. Danto. Marcel duchamp e o fim do gosto: uma defesa da arte contemporânea. *ARS (São Paulo)*, 6(12):15–28, 2008. Citado na página 17.

S. Dixon. Live tracking of musical performances using on-line time warping. In *Proceedings of the 8th International Conference on Digital Audio Effects*, volume 92, page 97. Citeseer, 2005. Citado na página 47.

S. Dixon. *Digital performance: a history of new media in theater, dance, performance art, and installation*. MIT press, 2015. Citado na página 20.

J. Eaton, W. Jin, and E. Miranda. The space between us: a live performance with musical score generated via affective correlates measured in eeg of one performer and an audience member. In *NIME'14 International Conference on New Interfaces for Musical Expression*, pages 593–596, 2014. Citado 4 vezes nas páginas 27, 45, 46 e 51.

E. Egozy and E. Y. Lee. *12*: Mobile phone-based audience participation in a chamber music performance. In T. M. Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 7–12, Blacksburg, Virginia, USA, June 2018. Virginia Tech. URL http://www.nime.org/proceedings/2018/nime2018_paper0002.pdf. Citado 3 vezes nas páginas 24, 61 e 62.

A. R. Fachini and M. R. Heinen. Aplicação de mfcc para modelar sons de instrumentos musicais. *11th proceedings of brazilian congress on computational intelligence*, 2016. Citado na página 32.

E. Fléty, N. Leroy, J.-C. Ravarini, and F. Bevilacqua. Versatile sensor acquisition system utilizing network technology. 2004. Citado na página 51.

J. T. Foote. Content-based retrieval of music and audio. In *Multimedia Storage and Archiving Systems II*, volume 3229, pages 138–148. International Society for Optics and Photonics, 1997. Citado na página 32.

A. Freed. Open sound control: A new protocol for communicating with sound synthesizers. In *International Computer Music Conference (ICMC)*, 1997. Citado na página 51.

A. Friberg and G. U. Battel. Structural communication. *The science and psychology of music performance: Creative strategies for teaching and learning*, pages 199–218, 2002. Citado na página 47.

M. Gimenes, P.-E. Largeron, and E. Miranda. Frontiers: Expanding musical imagination with audience participation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 350–354, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University. ISBN 978-1-925455-13-7. URL http://www.nime.org/proceedings/2016/nime2016_paper0068.pdf. Citado 4 vezes nas páginas 25, 45, 46 e 51.

F. Gonzaga and M. Fernandes. Inteligência artificial é o próximo passo na evolução da arte? https://arte.estadao.com.br/focas/estadaoqr/materia/inteligencia-artificial-e-o-proximo-passo-na-evolucao-da-arte, 2019. 2020-05-25. Citado 2 vezes nas páginas 39 e 40.

H. R. Guimarães. Recuperação de informações musicais: Uma abordagem utilizando deep learning, 2018. Citado na página 36.

P. Hamel and D. Eck. Learning features from music audio with deep belief networks. In *ISMIR*, volume 10, pages 339–344. Utrecht, The Netherlands, 2010. Citado 2 vezes nas páginas 31 e 32.

Y. Harris. The meta-orchestra: research by practice in group multi-disciplinary electronic arts. *Organised Sound*, 9(3):283–294, 2004. Citado na página 51.

A. Hindle. SWarmed: Captive portals, mobile devices, and audience participation in multi-user music performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 174–179, Daejeon, Republic of Korea, May 2013. Graduate School of Culture Technology, KAIST. URL `http://nime.org/proceedings/2013/nime2013_62.pdf`. Citado 8 vezes nas páginas 24, 43, 45, 46, 58, 61, 62 e 64.

M. Hirabayashi and K. Eshima. Sense of space: the audience participation music performance with high-frequency sound id. In *NIME*, pages 58–60, 2015. Citado 3 vezes nas páginas 25, 45 e 46.

J. Hodgson. *Understanding Records: A Field Guide To Recording Practice*. Continuum, 2010. Citado na página 30.

O. Hödl, G. Fitzpatrick, F. Kayali, and S. Holland. Design implications for technology-mediated audience participation in live music. 2017. Citado 2 vezes nas páginas 14 e 44.

M. D. Hoffman, D. M. Blei, and P. R. Cook. Content-based musical similarity computation using the hierarchical dirichlet process. In *ISMIR*, pages 349–354, 2008. Citado na página 32.

B. V. Hout, L. Giacolini, B. Hengeveld, M. Funk, and J. Frens. Experio: a design for novel audience participation in club settings. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 46–49, London, United Kingdom, 2014. Goldsmiths, University of London. URL `http://www.nime.org/proceedings/2014/nime2014_481.pdf`. Citado 3 vezes nas páginas 26, 45 e 46.

A. Hunt, M. M. Wanderley, and R. Kirk. Towards a model for instrumental mapping in expert musical interaction. In *ICMC*. Citeseer, 2000. Citado 4 vezes nas páginas 41, 48, 49 e 50.

A. D. C. Junior, S. W. Lee, and G. Essl. Understanding cloud support for the audience participation concert performance of crowd in c[loud]. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 176–181, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University. ISBN 978-1-925455-13-7. URL `http://www.nime.org/proceedings/2016/nime2016_paper0037.pdf`. Citado 4 vezes nas páginas 25, 45, 46 e 61.

G. D. Juraszek et al. Reconhecimento de produtos por imagem utilizando palavras visuais e redes neurais convolucionais. 2014. Citado na página 35.

H. Katayose and K. Okudaira. ifp: a music interface using an expressive performance template. In *International Conference on Entertainment Computing*, pages 529–540. Springer, 2004. Citado na página 47.

D. Kaufman. Dá pra fazer arte com inteligência artificial? https://epocanegocios.globo.com/colunas/IAgora/noticia/2019/08/da-pra-fazer-arte-com-inteligencia-artificial.html, 2019. 2020-05-25. Citado na página 38.

A. Kirke and E. R. Miranda. *Guide to computing for expressive music performance*. Springer Science & Business Media, 2012. Citado na página 50.

G. Kogan. Machine learning for artists. https://medium.com/@genekogan/machine-learning-for-artists-e93d20fdb097#.xc9qk7ca7, 2016. 2020-05-17. Citado na página 37.

M. Kraiser, C. G. Guimaraes, M. B. Mendonca, et al. Expressão nas performances com mediação tecnológica. 2011. Citado na página 21.

D. Kuhn, A. Fagundes, A. Souza, and F. C. Souza. Application of neural networks for recognition of musical instrument images. 09 2019. Citado na página 35.

O. Lartillot, P. Toiviainen, and T. Eerola. A matlab toolbox for music information retrieval. In *Data analysis, machine learning and applications*, pages 261–268. Springer, 2008. Citado na página 32.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. Citado 2 vezes nas páginas 35 e 37.

S. W. Lee and J. Freeman. echobo : Audience participation using the mobile music instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 450–455, Daejeon, Republic of Korea, May 2013. Graduate School of Culture Technology, KAIST. URL `http://nime.org/proceedings/2013/nime2013_291.pdf`. Citado 5 vezes nas páginas 24, 45, 46, 59 e 64.

G. Levin et al. Dialtones-a telesymphony, 2001. Citado na página 20.

M. Levy and M. Sandler. Lightweight measures for timbral similarity of musical audio. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pages 27–36. ACM, 2006. Citado na página 32.

R. Lins and C. S. Novaes. *Aplicação para Aumentar a Interação entre Músico e Público em Apresentações Musicais ao Vivo em Bares do Recife*. PhD thesis, Universidade Federal de Pernambuco, 2017. Citado 3 vezes nas páginas 27, 45 e 46.

B. Logan and A. Salomon. A music similarity function based on signal analysis. In *ICME*, pages 22–25, 2001. Citado na página 32.

B. Logan et al. Mel frequency cepstral coefficients for music modeling. In *ISMIR*, volume 270, pages 1–11, 2000. Citado na página 32.

E. J. Lok. Toronto emotional speech set (tess), 2019. URL `https://www.kaggle.com/ejlok1/toronto-emotional-speech-set-tess`. [DataSet]. Citado na página 72.

P. S. Lopes, E. L. Lasmar, R. L. Rosa, and D. Z. Rodríguez. O uso da rede neural convolucional como classificador de emoções em um sistema de recomendação de música. In *Anais do XIV Simpósio Brasileiro de Sistemas de Informação*, pages 324–317. SBC, 2018. Citado na página 35.

M. Loureiro, H. Yehia, H. DE P AULA, T. Campolina, and D. MO TA. Content analysis of note transitions in music performance. In *Proceedings of the 6th Sound and Music Computing Conference (SMC 2009). Porto, Portugal*, pages 355–359, 2009. Citado na página 16.

G. Madison, F. Gouyon, F. Ullén, and K. Hörnström. Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of experimental psychology: human perception and performance*, 37(5):1578, 2011. Citado na página 31.

T. N. Magalhães. Análise do bending e do vibrato na guitarra elétrica a partir dos descritores de expressividade da ferramenta Expan. Master's thesis, Escola de Música da Universidade Federal de Minas Gerais, Universidade Federal de Minas Gerais, 2015. Citado na página 31.

C. R. Mamedes et al. Design sonoro e interação em instalações audiovisuais. 2015. Citado 5 vezes nas páginas 21, 22, 28, 29 e 62.

M. Mancini, R. Bresin, and C. Pelachaud. A virtual head driven by music expressivity. *IEEE transactions on audio, speech, and language processing*, 15(6):1833–1841, 2007. Citado na página 47.

M. I. Mandel. Svm-based audio classification, tagging, and similarity submissions. *online Proc. of the 7th Annual Music Information Retrieval Evaluation eX-change (MIREX-2010)*, 2010. Citado na página 32.

M. I. Mandel and D. P. Ellis. Song-level features and support vector machines for music classification. *Proc. 6th Int. Conf. Music Information Retrieval*, 2005. Citado na página 32.

E. Martineli. *Extração de conhecimento de redes neurais artificiais.* PhD thesis, Universidade de São Paulo, 1999. Citado na página 34.

M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559, 2003. Citado na página 35.

D. Mazzanti, V. Zappi, D. G. Caldwell, and A. Brogni. Augmented stage for participatory performances. In *NIME*, pages 29–34, 2014. Citado 2 vezes nas páginas 28 e 69.

J. McCartney. Rethinking the computer music language: Supercollider. *Computer Music Journal*, 26(4):61–68, 2002. Citado na página 51.

B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, volume 8, pages 18–25, 2015. Citado na página 31.

C. McKay and I. Fujinaga. jmir: Tools for automatic music classification. In *ICMC*, 2009. Citado na página 31.

C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga. Ace: A framework for optimizing music classification. In *ISMIR*, pages 42–49, 2005. Citado na página 31.

E. A. G. d. Melo et al. Arthron: uma ferramenta para gerenciamento e transmissão de mídia sem performances artístico-tecnológicas. 2010. Citado 2 vezes nas páginas 20 e 21.

T. Morandi. Sentinela sonora. https://www.youtube.com/watch?v=9xPTkbXPAfo, 2014. URL https://www.youtube.com/watch?v=9xPTkbXPAfo. Citado na página 65.

A. Mulder, S. Fels, and K. Mase. Empty-handed gesture analysis in max/fts. In *Proceedings of the AIMI international workshop on Kansei-the technology of emotion, Antonio Camurri (ed.)*, pages 87–90, 1997. Citado na página 50.

F. Pachet and J.-J. Aucouturier. Improving timbre similarity: How high is the sky. *Journal of negative results in speech and audio sciences*, 1(1):1–13, 2004. Citado na página 32.

E. Pampalk. Computational models of music similarity and their application in music information retrieval. *PhD thesis, Technischen Universitat Wien*, 2006. Citado na página 32.

M. V. Paolo Prandoni. *Signal Processing for Communications (Communication and Information Sciences)*. EPFL Press, 2008. Citado na página 29.

E. L. B. Patrício. *Instrumentos musicais digitais - uma abordagem composicional*. PhD thesis, Dissertação (Mestrado)—UFPA-Universidade Federal Do Paraná, Departamento De . . . , 2010. Citado na página 41.

L. F. Patsko. Tutorial: Aplicações, funcionamento e utilização de sensores. *Maxwell Bohr: Instrumentação Eletrônica*, 2006. Citado na página 42.

J. Pearson. Uma obra de arte gerada por ia foi vendida num leilão por us$ 432 mil. https://www.vice.com/pt_br/article/43ez3b/uma-obra-de-arte-gerada-por-ia-foi-vendida-num-leilao-por-usdollar-432-mil, 2018. 2020-05-25. Citado na página 38.

A. S. Pires. *Métodos de segmentação musical baseados em descritores sonoros*. PhD thesis, Universidade de São Paulo, 2011. Citado na página 30.

J. Plaza. Arte e interatividade: autor-obra-recepção. *ARS (São Paulo)*, 1(2):09–29, 2003. Citado na página 23.

M. S. Puckette. Pure data. In *International Computer Music Conference.*, volume 1997, pages 224–227, San Francisco, 1996. International Computer Music Association. Citado na página 51.

D. Pye. Content-based methods for the management of digital music. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 4, pages 2437–2440. IEEE, 2000. Citado na página 32.

J. H. Reiss. *From margin to center: The spaces of installation art*. Mit Press, 2001. Citado na página 21.

D. S. Ribeiro and G. Mendes. Pintura artística. https://medium.com/@genekogan/machine-learning-for-artists-e93d20fdb097#.xc9qk7ca7, 2016. 2021-06-01. Citado na página 38.

F. C. Ribeiro. Action painting, happening e performance art: da ação como fator significante à ação como obra nas artes visuais. *Visualidades*, 8(2), 2010. Citado 2 vezes nas páginas 55 e 56.

C. Rocha. O imaterial e a arte interativa. *DOMINGUES, Diana. VENTURELLI, Suzete.(org.). Criação e poéticas digitais. Caxias do Sul: Edusc*, pages 27–31, 2005. Citado 2 vezes nas páginas 17 e 45.

G. L. Rocha, J. T. Araújo, and F. L. Schiavoni. Ha dou ken music: Different mappings to play music with joysticks. In *NIME*, pages 77–78, 2019a. Citado na página 81.

G. L. Rocha, J. a. Teixeira, and F. Schiavoni. Ha dou ken music: Mapping a joysticks as a musical controller. In F. Schiavoni, T. Tavares, R. Constante, and R. Rossi, editors, *Proceedings of the 17th Brazilian Symposium on Computer Music*, pages 69–75, São João del-Rei - MG - Brazil, Sept. 2019b. Sociedade Brasileira de Computação. Citado na página 81.

J. B. S. Rosa, R. Magalhães, and A. Oliveira. Entoados. https://www.youtube.com/watch?v=sBswNbF3HTc&t=279s, 2006. URL https://www.youtube.com/watch?v=sBswNbF3HTc&t=279s. Citado na página 65.

J. B. Rovan, M. M. Wanderley, S. Dubnov, and P. Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Kansei, The Technology of Emotion. Proceedings of the AIMI International Workshop*, pages 68–73. Genoa: Associazione di Informatica Musicale Italiana, October, 1997. Citado na página 41.

W. Salomão. *Hélio Oiticica: Qual é o parangolé?–E outros escritos*. Editora Companhia das Letras, 2015. Citado na página 18.

J. M. P. Santos. Breve histórico da "performance art" no brasil e no mundo. *Revista Ohun*, 4(4):1–32, 2008. Citado 2 vezes nas páginas 17 e 55.

R. M. Schafer. *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993. Citado na página 66.

F. L. Schiavoni, A. Gomes, A. de Paulo, C. de Souza, F. Carvalho, F. Resende, G. L. Rocha, I. S. Junior, I. Andrade, J. Araújo, et al. Alice: Arts lab in interfaces, computers, and everything else-research report (2019). In *Anais do XVII Simpósio Brasileiro de Computação Musical*, pages 157–164. SBC, 2019. Citado na página 80.

F. L. Schiavoni, J. T. Araújo, I. d. O. S. Junior, and I. de Melo Freitas. As lições aprendidas com a orchidea. *DEBATES-Cadernos do Programa de Pós-Graduação em Música*, (23), 2020. Citado na página 80.

K. Seyerlehner, G. Widmer, and P. Knees. Frame level audio similarity-a codebook approach. In *Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)*, 2008. Citado na página 32.

I. E. Simurra. Análise musical assistida por descritores de Áudio: um estudo de caso da obra reflexões de jônatas manzolli, 2015. Citado na página 30.

M. Sogabe. Instalações interativas mediadas pela tecnologia digital: análise e produção. *ARS (São Paulo)*, 9(18):60–73, 2011. Citado 2 vezes nas páginas 22 e 23.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. Citado na página 37.

B. Taylor. A history of the audience as a speaker array. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 481–486, Copenhagen, Denmark, 2017. Aalborg University Copenhagen. URL http://www.nime.org/proceedings/2017/nime2017_paper0091.pdf. Citado 2 vezes nas páginas 17 e 20.

E. C. F. Teixeira. Análise quantitativa da expressividade musical com base em medidas acústicas e do gesto físico. Master's thesis, Engenharia Elétrica / Universidade Federal de Minas Gerais, Universidade Federal de Minas Gerais, 2010. Citado na página 16.

S. D. Thorn and B. Lahey. A haptic-feedback shoulder rest for the hybrid violin. In *Proceedings of the 2019 International Computer Music Conference*, pages 1–6, 2019. Citado na página 47.

E. d. S. Torres, L. d. J. Silva, R. M. Santos, H. T. Macedo, and L. N. Matos. Redes neurais convolucionais no reconhecimento de fala em português para jogos em plataformas móveis. *Proceedings of SBGames*, 2016. Citado na página 35.

D. Trueman, P. R. Cook, S. Smallwood, and G. Wang. Plork: the princeton laptop orchestra, year 1. In *ICMC*, 2006. Citado na página 44.

L. Turchet and M. Barthet. Demo of interactions between a performer playing a smart mandolin and audience members using musical haptic wearables. In T. M. Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 82–83, Blacksburg, Virginia, USA, June 2018. Virginia Tech. URL http://www.nime.org/proceedings/2018/nime2018_paper0019.pdf. Citado 3 vezes nas páginas 27, 45 e 47.

T. Ungvary and M. Kieslinger. *Creative and interpretative processmilieu for live-computermusic with the Sentograph.* na, 1998. Citado na página 49.

A. C. G. Vargas, A. Paes, and C. N. Vasconcelos. Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres. In *Proceedings of the XXIX Conference on Graphics, Patterns and Images*, volume 1, 2016. Citado 2 vezes nas páginas 35 e 37.

M. M. Wanderley, N. Schnell, and J. Rovan. Escher-modeling and performing composed instruments in real-time. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, volume 2, pages 1080–1084. IEEE, 1998. Citado na página 49.

G. Wang, P. R. Cook, et al. Chuck: A concurrent, on-the-fly, audio programming language. In *ICMC*, 2003. Citado 2 vezes nas páginas 44 e 51.

A. Ward, J. Rohrhuber, F. Olofsson, A. McLean, D. Griffiths, N. Collins, and A. Alexander. Live algorithm programming and a temporary organisation for its promotion. In *Proceedings of the README Software Art Conference*, volume 289, page 290, 2004. Citado na página 44.

N. Weitzner, J. Freeman, S. Garrett, and Y.-L. Chen. massmobile-an audience participation framework. In *NIME*, volume 12, pages 21–23, 2012. Citado 4 vezes nas páginas 26, 43, 45 e 46.

A. Williamon. *Musical excellence: Strategies and techniques to enhance performance*. Oxford University Press, 2004. Citado na página 17.

M. Wright. Open sound control: an enabling technology for musical networking. *Organised Sound*, 10(3):193, 2005. Citado na página 51.

L. Zhang, Y. Wu, and M. Barthet. A web application for audience participation in live music performance: The open symphony use case. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 170–175, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University. ISBN 978-1-925455-13-7. URL http://www.nime.org/proceedings/2016/nime2016_paper0036.pdf. Citado 3 vezes nas páginas 26, 45 e 46.

Y. Zhang, Y. Li, D. Chin, and G. Xia. Adaptive multimodal music learning via interactive-haptic instrument. *arXiv preprint arXiv:1906.01197*, 2019. Citado na página 47.

U. Zölzer. *Digital Audio Signal Processing*. John WIley and Sons, 2008. Citado na página 30.