



14º Simpósio Brasileiro de Computação Musical

14th Brazilian Symposium on Computer Music

31 de outubro a 02 de novembro de 2013

**Escola de Música de Brasília (BEM)
Brasília, DF**

14º Simpósio Brasileiro de Computação Musical (SBCM 2013)

31 de Outubro a 02 de Novembro de 2013

Escola de Música de Brasília

Brasília, DF, Brasil

ANAIIS

ISBN 978-85-7669-286-7

Editora

Sociedade Brasileira de Computação – SBC

Organizadores

Edilson Feredá (UCB)

Fernando William Cruz (UnB)

Giordano Cabral (UFRPE)

Damián Keller (UFAC)

Realização e Organização

Mestrado em Gestão do Conhecimento e Tecnologia da Informação (MGCTI/UCB)

Laboratório de Computação Musical e Áudio (LCMA/UnB Gama)

Promoção

Comissão Especial de Computação Musical – CECM

Sociedade Brasileira de Computação – SBC

Este volume contém os artigos apresentados no 14º Simpósio Brasileiro de Computação Musical (SBCM 2013). A inclusão nesta publicação não necessariamente constitui endosso pelos editores e/ou organizadores.

A fonte e os direitos da SBC devem ser devidamente referenciados. As cópias não devem ser utilizadas de nenhum modo que implique o endosso da SBC. Cópias da obra não podem ser colocadas à venda sem a autorização expressa da SBC.

Permissão para fazer cópias impressas ou digitais de todo ou parte deste trabalho para uso pessoal ou acadêmico é concedido sem taxas desde que cópias não sejam feitas ou distribuídas para renda ou vantagem comercial e que cópias contenham esta observação e citação completa na primeira página.

Sociedade Brasileira de Computação
CNPJ nº 29.532.264/0001-78
Inscrição Estadual isenta
CCM nº 18115128

<http://www.sbc.org.br>
Av. Bento Gonçalves, 9500 – Setor 4 – Sala 116 – Prédio 43424 – Agronomia
CEP 91501-970 – Porto Alegre – RS, Brasil

Prefácio

Desde 1994, quando da sua 1ª edição, o SBCM vem firmando-se como importante espaço integrador entre as comunidades científicas de Computação, Música, Engenharias e Tecnologias Sonoras no Brasil. Em sua trajetória têm-se explorado sistematicamente as múltiplas relações entre música e tecnologia e discutido novos rumos de pesquisa e desenvolvimentos nesta extensa área denominada Computação Musical.

A Computação Musical é uma área eminentemente interdisciplinar e que engloba problemas computacionais interessantes e relevantes (por exemplo, de representação de conhecimento musical, de controle a tempo real de dispositivos, de interação com usuários possivelmente leigos em informática, para enumerar apenas alguns) subjacentes ao universo musical, que já é fascinante por si só. É uma área emergente no Brasil, mas já possui um reconhecimento da comunidade acadêmica de informática, uma vez que a Sociedade Brasileira de Computação (SBC) possui uma Comissão Especial de Computação Musical, que organiza um evento periódico.

As edições anteriores do SBCM aconteceram em: *(i)* 1994 - Caxambu, MG; *(ii)* 1995 - Canela, RS; *(iii)* 1996 - Recife, PE; *(iv)* 1997 - Brasília, DF; *(v)* 1998 - Belo Horizonte, MG; *(vi)* 1999 - Rio de Janeiro, RJ; *(vii)* 2000 - Curitiba, PR; *(viii)* 2001 - Fortaleza, CE; *(ix)* 2003 - Campinas, SP; *(x)* 2005 - Belo Horizonte, MG; *(xi)* 2007 - São Paulo, SP; *(xii)* 2009 - Recife, PE e *(xiii)* 2011 - Vitória, ES.

A 14ª edição do SBCM ocorreu em Brasília e buscou promover um diálogo da comunidade acadêmico-científica com os setores da sociedade interessados na aplicação e no desenvolvimento de Tecnologias da Informação e Comunicação relacionadas com a produção e o consumo de artefatos musicais.

Estes Anais reúnem os trabalhos técnico-científicos aprovados.

Brasília, novembro de 2013.

Os Organizadores do SBCM 2013.

Comitê de Programa SBCM 2013

Organizadores do 14º SBCM – Simpósio Brasileiro de Computação Musical

Edilson Ferneda (UCB) e Fernando William Cruz (UnB Gama)

Coordenador do Comitê Técnico do 14º SBCM

Giordano Cabral (UFRPE)

Membros do Comitê de Programa Técnico:

Charles Nichols (Virginia Tech)

Edilson Ferneda (Universidade Católica de Brasília – UCB)

Evandro Miletto (Instituto Federal do Rio Grande do Sul - Campus Porto Alegre)

Fabien Gouyon (INESC Porto)

Fabio Kon (IME – USP)

Fernando William Cruz (Universidade de Brasília)

Flávio Soares Corrêa da Silva (Universidade de São Paulo)

Geber Ramalho (Universidade Federal de Pernambuco)

Giordano Cabral (UFRPE)

Hugo de Paula (PUC Minas)

Ian Whalley (University of Waikato)

Jean Bresson (Ircam)

Jean-Pierre Briot (LIP6 - CNRS - Univ. Paris 6 & PUC-Rio)

Luis Jure (Universidad de la República)

Luiz Naveda (Ghent University)

Marcelo Gimenes (Universidade Estadual de Campinas)

Marcelo Pimenta (Universidade Federal do Rio Grande do Sul)

Marcelo Queiroz (University of São Paulo)

Marcio Brandao (Universidade de Brasília)

Martín Rocamora (Universidad de la República)

Sever Tipei (University of Illinois School of Music)

Tiago Tavares (Universidade Estadual de Campinas)

Coordenador do Comitê Musical do 14º SBCM

Damián Keller (UFAC)

Membros do Comitê de Programa Musical:

Barry Truax (Simon Fraser University)

Camilo Rueda (Pontificia Universidad Javeriana)

Carlos Reynoso (Universidad de Buenos Aires)

Damián Keller (Universidade Federal do Acre, NAP)

Fernando Iazzetta (Universidade de São Paulo)

Jônatas Manzolli (Universidade Estadual de Campinas)

José Fornari (Universidade Estadual de Campinas)
Marcelo Gimenes (Universidade Estadual de Campinas)
Marcelo Wanderley (McGill University)
Maria Helena de Lima (Universidade Federal do Rio Grande do Sul, CAp)
Rodolfo Coelho de Souza (Universidade de São Paulo, Ribeirão Preto)
Rodrigo Ciccelli Velloso (Universidade Federal do Rio de Janeiro)
Rogério Costa (Universidade de São Paulo)
Victor Lazzarini (National University of Ireland, Maynooth)

Outras coordenações relacionadas

Comitê Técnico Local

Ana Paula Bernardi Silva (UCB)

Márcio da Costa Pereira Brandão (UnB)

Coordenador da Mostra de Tecnologia Musical

Geber Lisboa Ramalho (UFPE, Brasil)

Coordenadores de Seções

Leandro Costalonga (UFES)

José Fornari (Unicamp)

Marcelo Pimenta (UFRGS)

14º Simpósio Brasileiro de Computação Musical (SBCM 2013)

COMITÊ GESTOR DA COMISSÃO ESPECIAL DE COMPUTAÇÃO MUSICAL

Dámian Keller (UFAC, Rio Branco-AC)

Edilson Feredá (UCB, Brasília-DF)

Fábio Kon (USP, São Paulo-SP)

Giordano Cabral (UFRPE, Recife-PE)

José Fornari (Unicamp, Campinas-SP)

Leandro Costalonga (FAESA, Vitória-ES)

Marcelo S. Pimenta (UFRGS, Porto Alegre-RS), presidente da CECM

Sumário

Artigos musicais

- *Marulho TransOceânico: performance musical entre dois continentes* 01
Alexandre Fenerich (UFJF), Giuliano Obici (USP), Flavio Schiavoni (USP)
- *Creating an Ecologically Modeled Performance Through the Remote Manipulation of Multiple Soundscapes* 13
Luzilei Aliel (UNICAMP), José Fornari (UNICAMP)
- *Avaliando Interfaces Gestuais Para Prática de Instrumentos Virtuais de Percussão* 23
Jaziel Vitalino Souza e Silva (UFPE), Filipe Calegario (UFPE), Giordano Cabral (UFRPE), Geber Ramalho (UFPE)
- *GTSpat: um sistema interativo de síntese e espacialização sonora* 35
Gabriel Rimoldi (UNICAMP), Adolfo Maia Jr. (UNICAMP)
- *Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade* 47
Damián Keller (UFAC), Floriano Pinheiro da Silva (UFAC), Edemilson Ferreira da Silva (UFAC), Victor Lazzarini (National University of Ireland), Marcelo Soares Pimenta (UFRGS)

Artigos Técnicos

- *A Preliminary Evaluation of Leap Motion as Controller of New Digital Music Instruments*..... 59
Eduardo Silva (UFPE), Jader Abreu (UPE), Janiel Almeida (UFPE), Veronica Teichrieb (UFPE), Geber Ramalho (UFPE)
- *A Hybrid Approach to Recommend Partners in Collaborative Musical Environments* 71
Guilherme Testa (UFRGS), Marcelo Pimenta (UFRGS), Damian Keller (UFAC, NAP), Victor Lazzarini (National University of Ireland)
- *Simulation of a Giannini True Reverber vacuum-tube guitar amplifier by using wave digital filters* 83
Thomaz Oliveira (UNICAMP), Gilmar Barreto (UNICAMP), Alexander Pasqual (UFMG)
- *Interactive Free Improvisation Using Time-domain Extrapolation of Textural Features* 95
Stéphan Schaub (UNICAMP), Tiago Tavares (UNICAMP), Adriano Monteiro (UNICAMP)
- *Implementação Orientada a FPGAs de um Oscilador Digital Multiplexado*..... 105
Mateus Negrelli (UnB), Marcio Brandão (UnB)
- *Detecção de Movimentos Auxiliares de Clarinetistas Através de Visão Computacional*..... 116
Rodrigo Seger (UFPR), Alessandro Koerich (PUCPR), Marcelo Wanderley (McGill University)
- *Impacto Regional do SBCM: Caso de estudo em Vitória-ES* 127
Leandro Costalonga (UFES), Marcus V. Neves (UFES)
- *O conceito de música experimental no ciberespaço: uma pesquisa etnográfica em comunidades virtuais de música* 134
Ramiro Galas (UnB)

- *Influência do Gênero Musical na Aquisição de Músicas Digitais* 142
Irapuru Florido (PUCPR), Maria Alexandra V. C. da Cunha (PUCPR), Daniel F. Sampaio (PUCPR), Miriam S. Mazo (PUCPR)
- *Avaliando Interfaces Gestuais Para Prática de Instrumentos Virtuais de Percussão* ... 154
Jaziel Souza (UFPE), Filipe Calegario (UFPE), Geber Ramalho (UFPE), Giordano Cabral (UFRPE)

Pôsteres técnicos

- *Organização Automática de Coleções de Partituras Musicais Utilizando Agrupamento Hierárquico* 166
Lucas Martiniano (UTFPR-CP), Andre Kashiwabara (UTFPR-CP), Carlos Silla (UTFPR).
- *Classificação Automática de Gêneros Musicais Utilizando Cifras* 178
Ricardo Corassa (UTFPR), Ricardo Corassa (UTFPR), Carlos Silla (UTFPR)
- *FPGA implementation of a digital audio reverberation structure* 189
Diego C. do Nascimento (UFRN), Valentin Roda (UFRN), José Alberto N. de Oliveira (UFRN)
- *FFT benchmark on Android devices: Java versus JNI* 193
Antonio de Carvalho Jr (USP), Marcelo Queiroz (USP), Andre Bianchi (USP), Max Júnior (USP)
- *Sound processors for live performance* 197
Eduardo Afonso (USP), Ruy Cunha Jr. (USP), Regis Faria (USP)
- *Uma roupa computacional para apoiar a educação musical no contexto da sala de aula* 201
Thiago Santos (UFRJ)
- *Videoaula no ensino coletivo de instrumentos musicais: um relato de experiência* 213
Juliana Silva (IFB)
- *Estudo interdisciplinar de obras para flauta solo* 217
Marcos Sampaio (UFBA), Guilherme Bertissolo (UFBA), Lucas Robatto (UFBA), José Martinez Neto (UFBA), Alisson da Silva (UFBA)

Pôsteres musicais

- *Quattro Pezzi per Orchestra de Scelsi sob a Ótica da Análise Musical Assistida por Computador* 221
Ivan Eiji Simurra (UNICAMP), Igor Leão Maia (UNICAMP), Jônatas Manzolli (UNICAMP)
- *Composição de um Tema Musical utilizando a abordagem do Lindenmayer System e da Geometria Fractal* 225
Dirceu de F. P. Melo (IFBA)

Amostra de protótipos

- *Using drawings and a guitar pedal to create one-man-band music: the Illusio* 230
Jerônimo Barbosa (UFPE), Filipe Calegario (UFPE), Veronica Teichrieb (UFPE), Geber Ramalho (UFPE), Giordano Cabral (UFRPE)

Marulho TransOceânico: performance musical entre dois continentes

Alexandre Fenerich¹, Giuliano Obici², Flávio Schiavoni³

¹Universidade Federal de Juiz de Fora
Instituto de Artes e Design
Juiz de Fora - MG - Brazil

alexandre.fenerich@ufjf.edu.br

²Escola de comunicação e Arte
Universidade de São Paulo
São Paulo - SP - Brazil
Technische Universität Berlin
Berlin - Germany

giuliano@usp.br

³Instituto de Matemática e Estatística
Universidade de São Paulo
São Paulo - SP - Brazil

fls@ime.usp.br

Abstract. *This paper is about aspects of audio performances that uses internet, a research from the work Ocean Seanoises (2012), from Duo N-1 (N-1.art.br), a musical piece for two computers located in different cities and audio transmission by internet. The main points of the article are: the performance's presentation, a contextualization of noise in art and a contextualization of the piece in the context of processual composition. The results present the transmission's sound marks in the listening process as well as the changing of roles of author/performer/listener in works that use the sound recording as a performatic resource.*

Resumo. *Este texto trata de aspectos envolvendo performance de áudio via internet a partir o trabalho Marulho TransOceânico (2012) do Duo N-1, (n-1.art.br), peça musical para dois computadores localizados em cidades diferentes e transmissão de áudio via internet. Os pontos principais deste artigo são: apresentação da performance, a contextualização do ruído na arte e uma contextualização a partir de músicas processuais. Os resultados evidenciam aspectos da marca sonora que a transmissão engendra no processo da escuta bem como as alternância de papéis entre autor/performer/ouvinte em trabalhos que utilizam a gravação como recurso performático.*

1. Introdução

A possibilidade de cooperar utilizando redes computacionais pode trazer desafios para técnicos e artistas. Concertos remotos utilizando Internet tendem a gerar resultados insatisfatórios e uma certa frustração quando se tem como objetivo simular uma performance de concerto nos moldes tradicionais. Tal frustração pode ser ainda

maior caso a expectativa seja utilizar conexões domésticas para a transmissão de áudio.

Um dos obstáculos encontrados nesse tipo de performance via Internet está na dificuldade técnica em garantir a integridade do sinal de áudio durante a transmissão. O áudio digital em seu formato PCM 32 bits com taxa de amostragem de 48kHz, por exemplo, consome uma largura de banda de 192kB/s. As conexões domésticas possuem largura de banda suficiente para recebimento (*download*) em tal velocidade mas não temos garantia de envio (*upload*) em tal velocidade.

Esta limitação das conexões domésticas levaram as ferramentas de comunicação de rede a adotar algumas técnicas para diminuir a quantidade de dados em uma transmissão de voz. Entre estas técnicas está a diminuição do tamanho de amostras para 16 bits ou menos, diminuição da taxa de amostragem, ou a compressão com perdas eliminando parte do sinal.

Além da largura de banda nem sempre ser suficiente para a transmissão de áudio, a velocidade de transmissão pode variar de acordo com o desempenho da rede. Isto pode causar latência e variação da latência (“jitter”) além de causar perdas de pacotes. As ferramentas mencionadas são concebidas também para automatizar configurações que corrijam o sistema caso haja pacotes perdidos ou atrasados.

Tanto a compressão do dado para o envio quanto a recuperação do sinal no caso de falhas na transmissão fazem de ferramentas comuns para a comunicação em rede uma alternativa pouco atrativa para performances musicais convencionais. Mesmo assim, artistas propõem trabalhos colaborativos utilizando tais características embutidas nessas ferramentas baseando-se em outros fatores estéticos que não o áudio com sua máxima qualidade¹.

É neste contexto que este artigo apresenta o processo de Marulho TransOceânico (2012) – obra musical do Duo N-1², a partir do registro de uma apresentação realizada em 30/08/2012 em João Pessoa e Berlim simultaneamente³. Nesta peça os resíduos sonoros decorrentes dos dispositivos de transmissão e recepção de áudio em rede foram explorados, de forma a tornar evidente ao ouvinte tais processos técnicos que estão embutidos na transmissão e comunicação sonora via protocolos digitais que nos cercam.

Apesar de ambos integrantes do Duo N-1 estarem vinculados a instituições de pesquisas (Universidade Federal de Juiz de Fora, Universidade de São Paulo e Technische Universität Berlin) esta performance utilizou-se de uma conexão de internet doméstica. Com isto, procurou-se manter uma investigação caseira descartando a possibilidade de conseguir recursos técnicos melhores através do vínculo institucional que os músicos mantém. Esta é uma característica do trabalho do Duo que se mantém aqui: a utilização de tecnologias ou técnicas acessíveis aos seus membros - os quais não possuem formação técnica. Por esta razão, utilizou-se o Skype como plataforma de comunicação apesar das evidentes desvantagens que o software apresenta no que tange à qualidade do áudio transmitido, já que foi concebido para a comuni-

¹Cascone [Cascone, 2000] apresenta tendências musicais contemporâneas como a Glitch Music, que explora ruídos e defeitos gerados em meio digital para a geração sonoridades características deste sistema. Alva Noto e Sakamoto são exemplos de artistas que exploram atualmente estas características [Nakahodo, 2012]. Artistas como Nicolas Collins [Duguid, 1985] e Yasonao Tone [LaBelle, 2006] vêm explorando estas sonoridades desde os anos 80. Além disso, o coletivo “Net vs. Net” realiza performances distribuídas que se aproveitam criativamente da latência e dos atrasos decorrentes das transmissões via rede [Tomiyoshi, 2013, 3]

²<http://giulianobici.com/n-1/mar-trans.html> acessado 11.08.2013

³O registro pode ser escutado em https://soundcloud.com/n-1-1/marulho_transoceanico

cação verbal e vocálica, mas vantagens expressivas no que tange sua disseminação e estabilidade em transmissões P2P [Tomiyoshi, 2013, 30].

Partido deste pressuposto, iniciamos esta discussão apresentando referências estéticas e musicais de autores que ajudam a pensar este campo de trabalho. Posteriormente discutiremos como os recursos de transmissão caseira em rede foram apropriados nesta peça de modo a explorar faturas sonoras não previstas no projeto destas tecnologias de transmissão. Finalmente, discutiremos a partir do experimento proposto pela obra, o ruído das transmissões sonoras em rede enquanto potencialidade criativa.

O presente artigo é organizado da seguinte forma: na seção 2 apresentaremos os conceitos envolvidos neste artigo, na seção 3 apresentamos o desenvolvimento da peça, na seção 4 o resultado obtido e na seção 5 apresentamos conclusões e trabalhos futuros.

2. Conceitos envolvidos

2.1. Da entropia da voz: ecos de Alvin Lucier

“Creio que estou tentando ajudar as pessoas a segurar conchas sobre seus ouvidos e escutarem o oceano novamente”⁴
(Alvin Lucier)

Alvin Lucier, no capítulo ‘Tape Recorders’ do livro ‘Music 109: Notes on Experimental Music’ [Lucier and Ashley, 2012] escreve sobre peças que se utilizam da gravação ao vivo em performance musical. Nele Lucier aponta que em tais peças se evidencia um outro modo de composição, como se a música passasse a ser composta por si mesma durante sua execução, cabendo ao compositor/intérprete apenas enquadrar o seu processo generativo - como escreve sobre ‘Came Out’, de Steve Reich:

O que era incrível é que você podia escutar a peça musical sendo composta por si mesma. (...) Uma vez que se inicia o processo de Came out, o compositor não dá nenhum passo a não ser o de separar e multiplicar as imagens [sonoras] de alguma forma em duas faixas de gravação. O compositor não decide o que acontece de momento a momento. [Lucier and Ashley, 2012, p.104].

Seguindo os comentários sobre o impacto que foi a descoberta do recurso da defasagem na música advindo do trabalho de Reich, relata ainda Lucier:

Ele sincronizou os tapes e os tocou em dois gravadores. Após deixá-los correr um pouco, eles passaram gradualmente a se separar. Dois gravadores nunca possuem exatamente a mesma velocidade, mesmo dentre os melhores que você possa encontrar. Um deles é mais veloz, e qualquer que seja o material gravado nos tapes, irá se mover à frente do outro. O loop se move fora de fase. Você passa a escutar a fase ligeiramente à frente de sua irmã gêmea. Os dois sons idênticos ligeiramente fora de sincronia geram um terceiro som. Ao deixar os gravadores tocando com sua própria velocidade, sem tratar a variação de velocidade como um defeito, Reich descobriu a fase na música. Foi uma grande descoberta. Foi como se um milagre tivesse acontecido. [Lucier and Ashley, 2012]

⁴“I guess I’m trying to help people hold shells up to their ears and listen to the ocean again”[Lucier et al., 1995, p.306]

Lucier descreve em detalhes a música como um processo que se desdobra a partir de um mecanismo, de uma célula ou de um embrião, sem a intervenção da vontade do compositor/intérprete ao manipular o material após ela iniciada.

O mesmo processo se evidencia em ‘I am sitting in a room’ (Lucier, 1969), onde a voz gravada faz soar a arquitetura a partir de reiteradas repetições/gravações, articulando ritmicamente a acústica da sala. “Se Lucier constrói uma arquitetura, é uma arquitetura imbuída com a problemática de ter um corpo” [LaBelle, 2006, p. 132] - uma arquitetura expressa sonoramente, carregada de sua voz. Nesta peça, as características acústicas do lugar da difusão da primeira versão gravada e de todas as subsequentes vão se tornando evidentes para a escuta na forma de um acorde, cujas notas correspondem às frequências das ondas estacionárias da sala. A referência causal da voz (sua identidade) vai sendo gradualmente perdida:

A imagem sonora inicial é a de uma voz emitida em uma sala com uma certa acústica (...). No final da peça, escutamos essencialmente a ressonância da sala vagamente articulada pelas flutuações de amplitude da voz. Neste caso, nossa percepção daquilo que é o objeto sonoro e daquilo que é o espaço acústico no qual ele é projetado passam a confluir. No início da peça, podemos afirmar sem dúvida alguma que o objeto sonoro é a voz. No final da peça, o objeto sonoro é uma entidade claramente mais ‘abstrata’, cujas características derivam da acústica da sala. [Wishart and Emmerson, 1996, p. 158]

Na peça ‘Came Out’ de Reich o processo se dá pela defasagem simples; em ‘I am Sitting In a Room’, pela saturação do sinal que leva a uma acentuação da ressonância da sala. Em ambos os casos, há um desencadeamento que independe da decisão dos compositores, os quais apenas deram início e criaram as condições para que ocorressem.

Nas peças descritas acima, uma vez iniciadas, os ‘performers’ passam de músicos a espectadores, vivenciando a mesma experiência da plateia. Da mesma forma, o compositor não pode mais intervir na execução do seu projeto, pois caso contrário seu gesto ficaria impregnado na peça. Os papéis aqui se fundem da figura do compositor/performer a ouvinte/público. Nessa situação, o processo disparado tende a desdobrar por si próprio um percurso sonoro.

Além disso, os autores não podem prever exatamente o resultado, e sua atitude passa a contar com um afastamento do controle momentâneo em função da surpresa da escuta. Ao contrário do objeto musical tradicional - impregnado de qualidades sonoras dadas por uma performance ou por decisões composicionais ponto a ponto - tais peças não sofrem ação alguma, no seu decorrer, de um gesto ou de um quebra - de uma ação composicional forte. Ganham assim uma certa autonomia ao se comportar de modo descolado de qualquer agenciamento gestual ou qualquer ação com o fim de direcionar a um sentido específico.

2.2. Marulho Oceânico

Marulho Oceânico é uma série de peças do Duo N-1 que partem de algumas das ideias que apresentamos acima e que possuem o mar como temática comum. Entretanto, dentre as várias versões, Marulho Transoceânico é a que mais se aproxima das práticas de música processual mencionadas. Os “marulhos” vêm adquirindo diferentes versões durante a trajetória do duo, num processo de reinvenção a cada versão, acumulando camadas de sentidos distintos e diálogos entre as próprias peças.

Até o momento foram criadas quatro versões: 1a versão - CD Jardim das Gambiarras Chinesas (2009)⁵, 2a. Audio e video live electronics (2009-2010)⁶, 3a. CD nmenosvideo - Limiares (2012)⁷, 4a. TransOceânico (2012)⁸. As versões exploram, sonora e visualmente, aspectos como imensidão, imersão, ruído, movimentos e ondulações, bem como a cor azul.

3. Descrição do trabalho

Marulho TransOceânico (MtO), quarta versão da uma série de peças Marulho Oceânico, é uma performance para voz falada, computadores e transmissão de áudio via internet. Nesta versão, o prefixo “trans” do título refere-se ao fato da performance ser realizada via internet (*streaming*) com cada integrante localizado em cidades diferentes, separados por um oceano.

A performance surgiu em 2012 com objetivo de desenvolver uma plataforma/set/instrumento de forma a explorar a conectividade como suporte performático. Utilizou-se de recursos técnicos disponíveis aplicados aos objetos do cotidiano e programas de computador (uma gambioluteria⁹): buscou-se encontrar uma forma de trabalhar à distância usando computadores pessoais, programas acessíveis e livres e uma conexão doméstica de internet. Manteve-se assim uma autonomia no processo de criação ao não se necessitar de equipamentos de ponta para a realização do trabalho – postura estética que norteia o trabalho do Duo N-1 [Fenerich and Obici, 2011].

Assumindo tal situação como campo investigativo, surgiram algumas perguntas: qual noção de espaço para uma performance realizada simultaneamente em dois lugares do globo? Como performatizar o próprio streaming e os elementos implícitos na transmissão? Qual sonoridade pode ser traduzida por este meio?

3.1. Metodologia

Diante de tais questionamentos partiu-se para os experimentos, os quais convergiram para um exercício simples de gravação e envio do mesmo sinal de áudio repetidas vezes via streaming (Skype) e troca de comandos de mensagem via TCP (Transport Control Protocol) através do objeto net send do programa Pure Data. A programação-partitura decorrente destas experimentações segue um procedimento semelhante à peça “I am sitting in a room” de Alvin Lucier.

O princípio da performance de Lucier, que faz emergir as ressonâncias naturais da sala em que é apresentada, serviu de modelo para a seguinte especulação, que moveu o trabalho: em procedimento semelhante, qual seria o espaço de uma transmissão de áudio distribuída? Pode-se dizer que o mote era simular um procedimento parecido ao “estou sentado em uma sala” para “estamos sentados em dois lugares do globo terrestre distantes um do outro e partilhamos a mesma música”. Como este espaço partilhado soaria? Diante deste cenário o Duo prosseguiu no exercício de configuração-criação-instrumento para a realização da peça. O setup para sua realização e a performance em si vem descrita abaixo.

⁵http://giulianobici.com/n-1/cd_jardim.html

⁶<http://giulianobici.com/n-1/mar.html>

⁷http://giulianobici.com/n-1/cd_n-video.html

⁸<http://giulianobici.com/n-1/mar-trans.html>

⁹Neologismo para descrever processo que mistura gambiarra e luteria de instrumentos e apropriação de materiais no set/plataforma/instrumento do Duo N-1. Vide descrição da performance Marulho TransOceânico 2012 - <http://dezdimensoes.wordpress.com/2012/08/28/dimensao-sonora/>

3.2. Da plataforma/setup/instrumento

Um primeiro passo para a realização do setup foi a configuração de rede. Por se tratar de uma conexão doméstica, foi necessário configurar um redirecionamento de portas para permitir que mensagens externas não requisitadas alcancem uma máquina da rede Interna. Esta configuração foi feita no modem / roteador localizado na Alemanha por duas razões: a) termos o acesso de configuração a este dispositivo e b) tal endereço tornou-se acessível tanto para a apresentação em João Pessoa quanto para os ensaios do Duo que foram realizados entre Berlim e outras cidades do Brasil.

Nesta configuração definimos um IP fixo na rede interna para a máquina que seria utilizada na performance e roteamos as portas de comunicação TCP utilizadas para os dados de controle para o IP desta máquina. Com isto garantimos a comunicação utilizando ambiente doméstico.

Um segundo passo foi o entendimento de que necessitaríamos de três softwares para realizar o trabalho, em cada um dos computadores: Skype, que realizou a transmissão de áudio, PureData, que gravou os áudios, enviou os comandos de início e fim dos processos e tocou os áudios gravados, e Soundflower, que realizou o roteamento de áudio entre os softwares anteriores. Além disso, percebeu-se que seriam necessários 2 tipos de conexão de rede: via skype transitava o áudio e via Pure Data, através dos objetos natsend/netreceive, mensagens de controle que seriam utilizadas para sincronizar os patches entre ambas as pontas da transmissão. Faz-se notar aqui que o Skype não depende da configuração supra citada pois o mesmo utiliza-se de outras técnicas para permitir a conexão doméstica entre pares.

Enviou-se então, a partir de uma das máquinas, as mensagens com o formato OSC de “/gravar” e “/reproduzir”. Com isto uma das máquinas assumiu a responsabilidade de sincronização de ambas e a troca de mensagem para este controle foi feita pelo redirecionamento de portas pré-configurado.

O patch para a gravação em ambas as máquinas recebia o sinal de áudio do skype redirecionado pelo soundflower. No primeiro momento grava-se em ambas este sinal de entrada por 22 segundos. Feita esta gravação ela é automaticamente tocada pelo Pure Data com a saída de áudio deste programa direcionada para o Skype via soundflower. Ao mesmo tempo, em outro buffer, o patch recebe o sinal provindo da reprodução da gravação recém-executada na outra ponta da transmissão. Ao final deste par reprodução/gravação, o processo é automaticamente reiniciado.

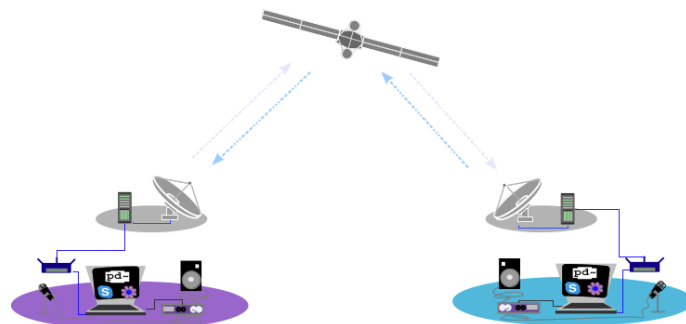


Figura 1: Plataforma / set / instrumento

3.3. Da partitura

A partitura da peça segue uma lógica de alternância, proveniente da própria lógica do setup: a peça inicia com cada performer (A e B) lendo um texto (parte 1). A voz

do executante A é transmitida e gravada pelo set do executante B simultaneamente com gravação da voz de B pelo set de A. Na parte 2 inicia-se a gravação do streaming ao vivo. A partir da terceira parte os executantes param de ler o texto. Inicia-se a execução das duas gravações, que são transmitidas e novamente gravadas pelas duas pontas da transmissão, seguindo uma programação que automatiza e repete o processo. A gravação tocada é sempre o resultado que foi gravado da transmissão anterior, conforme o esquema:

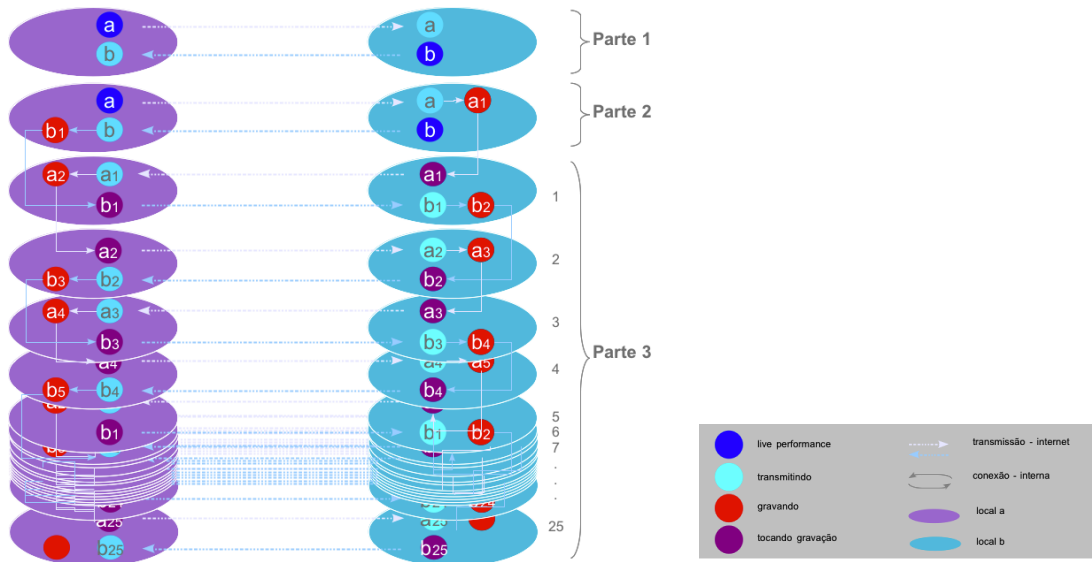


Figura 2: Esquema / partitura + legenda

O processo é repetido 25 vezes e o resultado acumula tanto o material inicial gravado quanto as sucessivas regravações de sua transmissão. Diante desse processo, o que se evidencia é o próprio extrato das camadas de transmissão, uma espécie de marca perceptiva dos resíduos sonoros envolvidos na transmissão gerados pela repetição e sua acumulação temporal no material gravado.

4. Resultado: Análise do MtO a partir do registro da performance.

Aqui apresentaremos um registro em áudio que é a gravação total de apenas um dos lados (em João Pessoa), de modo que, no seu início, apenas a voz de um dos executantes (que está sendo transmitida) é escutada. Analisaremos o sinal de áudio da peça ao longo dos seus 12 minutos de execução.

4.1. Sonoridade predominante

Do ponto de vista da sonoridade predominante, observa-se, primeiramente, que há um corte de frequências em aproximadamente 12300Hz, tanto no início, em que há apenas uma voz (fig. 3a), quanto em trechos de maior densidade (fig. 3b). Isto se deve não ao sistema de gravação, mas à compressão utilizada para a transmissão, pois alguns “lapsos” ocorrem de modo a, por vezes, ultrapassar este limiar.

Além disso, há um ruído considerável já no início do áudio, escutado enquanto um hiss nos primeiros instantes de silêncio (fig. 3a), e um nível alto incomum nas regiões aguda e superaguda (acima de 4000 Hz – figs. 3a e 3b).

Estas características de tradução de frequências e de amplitudes são refletidas nas particularidades da transmissão via Skype, cuja compactação utiliza-se do

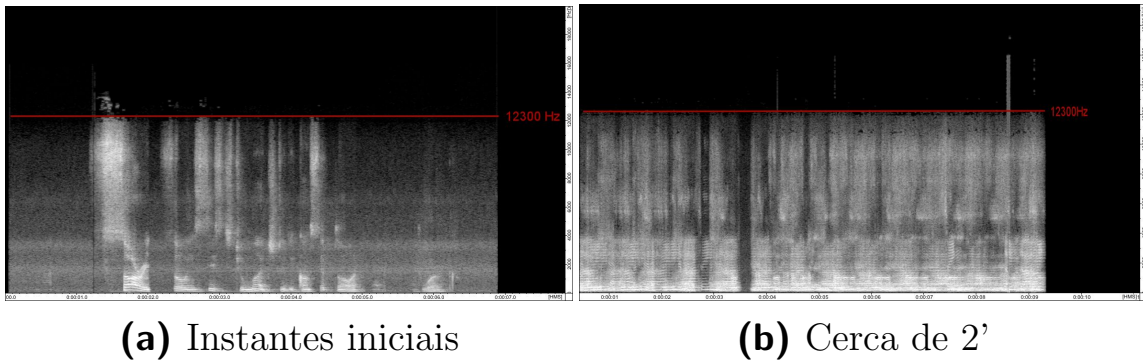


Figura 3: Espectrograma da peça

codec Silk a fim de transmitir os dados. Este codec reduz o 'sample rate' e o 'bit rate' do áudio de entrada para uma banda de frequências e amplitudes consideradas suficientes para a inteligibilidade da voz falada. Dependendo da banda de transmissão, o 'sample rate' pode ser reduzido a 24, 16 ou 8 KHz, e o 'bit rate' de 40 a 6 Kbps[Jensen et al., 2010].

Neste codec, o áudio passa por um filtro passa alta (High pass filter) a fim de excluir frequências graves consideradas desnecessárias para a voz falada. Isto significa que há uma alta compressão do áudio no que tange a tradução do espectro de frequências e, sobretudo pela baixa taxa de kbps, que há um achatamento das nuances de definição de amplitudes.

Somado a isso, o codec Silk foi concebido para reconhecer padrões vocais, excluindo bandas de frequência que não trafegam por esta zona[Jensen et al., 2010]. A sonoridade é portanto simplificada, aproximando-se à de um rádio antigo, com predominância de frequências médio-agudas e poucas nuances; além disso, um material sonoro muito distinto do espectro da voz não será (em tese) transmitido.

Essa característica da ferramenta se evidencia em trechos da peça onde ainda é possível reconhecer a voz falada ao longo da gravação. Isto nos indica blocos atrasados na transmissão que a ferramenta soma ao sinal. A presença de tais blocos nos levam a questionar o que a ferramenta reproduziu no momento em que o mesmo deveria ter sido reproduzido. Mecanismos de recuperação de falha podem ter a) reproduzido o bloco anterior, b) interpolado os valores entre os 2 blocos, c) adicionado conteúdo qualquer. Qualquer uma destas soluções não conseguiria recompor o sinal ausente mas servem para manter o sinal de áudio contínuo e evitar 'glitch'.

A reprodução deste sinal, corrompido sucessivas vezes, e a correção do sinal a cada reprodução leva a adição de muitos blocos de recuperação de perdas. Pelo registro da performance podemos notar que a faixa de frequência da voz falada diminui ao mesmo tempo que há uma ampliação em todas as frequências presentes no espectro do sinal (fig. 3b). Neste ponto, tanto a visualização do espectro quando a audição da peça nos mostram que este processo tende ao ruído branco devido a sucessivas adições de conteúdo randômico ao sinal [Ephremides and Thomas, 1974].

A reaplicação do codec com perda também pode contribuir para esta degeneração do sinal. A cada processo de gravação há uma reaplicação do codec, e novas componentes do sinal podem ser removidas até que o mesmo possua pouca ou nenhuma característica ao registro sonoro original.

4.2. Estágios da peça

Pensando agora num aspecto mais formal, uma escuta por inteiro deste registro evidencia o processo da peça ao longo dos seus quase 12 minutos de duração. Pode-se então definir alguns momentos de articulação de modo a evidenciar seus estágios – muito embora tais pontos, salvo no início e no fim, são arbitrários, pois a peça segue como um contínuo, não havendo articulações claras.

O primeiro destes estágios se dá em 0'22", quando claramente se interrompe a gravação do som externo (microfone) e se inicia o processo das regravações das transmissões. A partir daí aparecem 'loops' de trechos vocais, gradualmente defasados a cada entrada. Neste estágio, cada nova aparição permanece claramente destacável, sendo pouco a pouco ofuscadas, de modo que em cerca de 01'56" se tem uma sucessão de picos, e as vozes acumuladas parecem tão comprimidas que perdem totalmente seu timbre característico.

O processo de acumulação das vozes continua até cerca de 3'48", mas neste momento uma série de rápidas aparições de material mais recentemente gravado alterna-se com material mais antigo (um áudio comprimido que consiste na acumulação de vários extratos vocais), o qual tende a uma maior entropia (efeito que denominamos maré – Fig. 4).

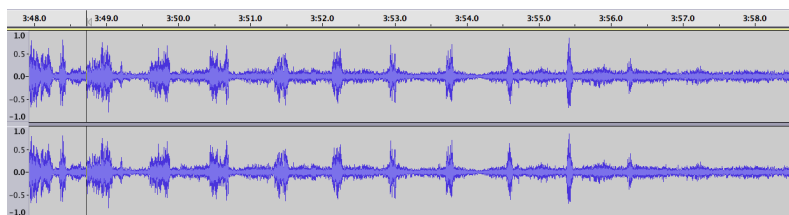


Figura 4: Alternância de material antigo com recente, em 3'49" (Maré)

O ruído gerado pelas sucessivas compactações (aqui chamado ruído de fundo), sobretudo a partir de 4'15", passa a predominar sobre o material vocálico, que vai progressivamente desaparecendo ou deixando de figurar no primeiro plano da escuta. Se tomarmos mais outros dois momentos, por exemplo em 6'36" e 10'37", notamos que o ruído de fundo vai se tornando progressivamente mais amplo em tessitura e amplitude, muito embora permaneça em ambos os pontos um resquício de material vocálico condensado e em 'loop'. Em alguns momentos, irrompe um material vocálico mais recente ou menos misturado, talvez decorrente de um ciclo que ainda continha elementos vocálicos dos estágios anteriores – 'efeito maré', que permeia toda a peça.

Um aspecto que fica bastante claro é que depois de certo número de repetições o processo de certa forma se estabiliza (a partir talvez de 4'23"). Após este ponto as repetições defasadas passam a não ser perceptíveis, o ruído de fundo vai lentamente tomando conta do espaço acústico e picos de amplitude passam a ser claramente periódicos, a cada 12' (fig. 5), coincidindo com o início de cada regravação (que tinha um ciclo de 12 segundos).

Tal periodicidade da gravação assemelha-se à técnica de síntese de cordas do algoritmo Karplus-Strong [Karplus and Strong, 1983]. Diferente deste algoritmo, a reprodução aqui de um buffer periodicamente não gera uma frequência mas um ritmo devido a seu tempo de reprodução ser muito elevado (12s).

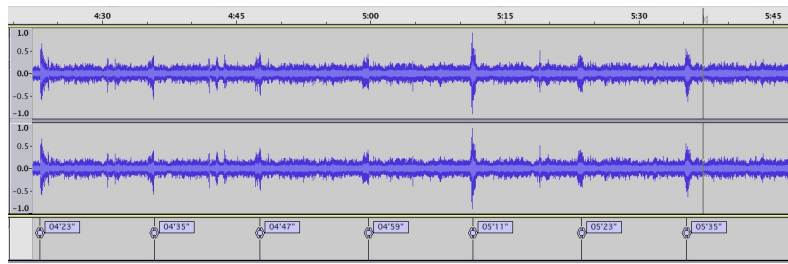


Figura 5: Picos de amplitude a cada 12', após 04'23''

Podemos então definir três grandes estágios para a obra: a gravação inicial, a aparição das vozes em ‘loops defasados’ e a progressiva entropia rumo à dissolução das vozes.

5. Conclusão

Neste artigo apresentamos a peça MtO do duo N-1 e uma discussão do resultado sonoro. No diálogo com a peça de Lucier, vale ressaltar que, diferentemente de ‘I am sitting in a room’, MtO trata da transmissão de uma conversação posta a soar em duas salas ‘simultaneamente’. Entretanto, as reiteradas gravações não acentuam o espaço onde cada performance acontece, mas a codificação da transmissão e o ruído causado pela aplicação repetida da codificação e transmissão. Por outro lado, da mesma forma que o espaço acústico da sala é vocalizado na peça de Lucier, o ruído resultante da acumulação, em MtO, passa inicialmente a ser articulado pelas vozes.

Em MtO, as reiteradas gravações das vozes que se perdem pouco a pouco no ruído de fundo, imperceptível ou ignorado, que paira numa transmissão via Skype, são decorrentes do acréscimo em amplitude do fundo da transmissão, causado pelo processo de retroalimentação do sinal de áudio. Não resta sequer a articulação rítmica da fala, como ocorre na obra de Lucier - sobrando apenas uma pulsação causada talvez tanto pelo vai-e-vem da transmissão ultramarina (o efeito maré) quanto pela retomada do loop da gravação.

No caso do MtO, diferentemente da peça de Lucier, o ruído não foi criado pela reverberação da sala mas pela compressão do sinal para o envio efetuado pelo codec do aplicativo Skype, assim como pela tentativa de recuperação de perdas de pacotes. O codec Silk efetua uma compressão dinâmica, pois varia as taxas de amostragem na medida em que o tamanho ou a velocidade da banda de transmissão mudam – e numa transmissão caseira esta medida parece percorrer um largo âmbito. Esta compactação dinâmica e a recuperação de falhas da rede tendem a trazer alguns ruídos inesperados – os quais passam despercebidos numa conversação normal, mas que ficam impregnados no áudio por conta do processo cumulativo proposto na performance.

Se por um lado, na transmissão por Skype, há uma sonoridade quase esquemática, suficiente apenas para traduzir os signos verbais (na estreita faixa de frequências e amplitudes que traduz), por outro há um campo de imprevisibilidade que traz curtos soluços (lapsos de transmissão), ruídos inesperados, pequenos cortes e trancos, resultantes talvez da transmissão e da compressão dinâmica. A ferramenta concebida preferencialmente para a transmissão verbal engendra, pelo mecanismo inventado para torná-la eficiente, uma variabilidade que traz, para a experiência da escuta desta peça, outros elementos sonoros além da voz a cada ciclo de repetição. Mesmo com sua sonoridade restrita, o processo de transformação do som vocálico

em ruído (no percurso da peça) gera interesse para a escuta pela constante variabilidade. Os lapsos sonoros que descrevemos anteriormente vão se acumulando ao longo dos ciclos de gravação e são acrescidos de outros lapsos, outras rítmicas e outras sonoridades.

É nesse sentido que, partindo de uma banda de transmissão caseira e de programas acessíveis, procurou-se explorar da transmissão aquilo que ela tem de imprevisível: os ruídos que ela porta, sua acumulação e latência, além da de-sincronia.

Este último aspecto revelou-se uma fatura interessante que não apareceria em uma suposta transmissão sem latência. Nos testes de sincronização entre os computadores, realizados em Juiz de Fora e Berlim, constatou-se que a mensagem-comando de início e fim do processo demorava alguns segundos para alcançar seu destino. Tal fato levou à dedução de que a simultaneidade, ou a noção de tempo real, neste caso, era uma impressão forjada pelo sistema de conversação contínua dado pelo Skype. Esta de-sincronia vem a gerar, assim, escutas distintas da peça em cada um dos lugares em que esta ocorre (e não apenas de uma versão a outra) – pois cada set e lugar terá uma acumulação diferente dos ruídos da transmissão. Desta maneira, por conta dos ruídos sonoros e temporais do ponto de vista da comunicação a peça ganha, a cada realização, duas versões distintas; seu próprio mecanismo gera uma escuta desdobrada em duas, separadas espacial e temporalmente.

Se, em tese, por se tratar de uma transmissão em meio digital, o ruído desta transmissão não deveria existir, o presente trabalho explicita justamente uma materialidade que se coloca aos ouvidos, dada talvez pela natureza da compressão e pelo próprio aparato midiático de transmissão. É sobre essa materialidade que trabalha: nessas camadas do meio digital que a peça se baseia, tentando tornar sensível o papel midiático do codec, como vem apontando Lev Manovitch [Manovich, 2010] em seus estudos do software no contexto das novas mídias, assim como Jonathan Sterne [Sterne, 2012] ao discutir a história do MP-3, mostrando a não neutralidade dos codecs e dos meios no universo digital.

“Formato denota toda uma gama de decisões que afetam o olhar, sentir, experiência e o funcionamento de um meio. Ele também define um conjunto de regras segundo as quais a tecnologia pode funcionar.”¹⁰

Trabalhos futuros incluem a possibilidade de gravação da peça em ambos os lados ou a utilização de outras ferramentas que possam trazer-lhe outras características.

6. Agradecimentos

Os autores gostariam de agradecer o apoio das agências de fomento CNPq (processo nº 141730/2010-2), FAPESP (processo nº 2008/08632-8), CAPES (processo nº BEX 1194/12-7) e DAAD (Deutsche Akademische Auslandsdienst) .

Referências

Cascone, K. (2000). The aesthetics of failure: ”post-digital”tendencies in contemporary computer music. *Computer Music Journal*, 24(4):12–18.

¹⁰“Format denotes a whole range of decisions that affect the look, feel, experience, and workings of a medium. It also names a set of rules according to which a technology can operate.”[Sterne, 2012, p.7]

- Duguid, B. (1985). Interview with nicolas collins. In *online: <http://media.hyperreal.org/zines/est/intervus/collins.html>*.
- Ephremides, A. and Thomas, J. B. (1974). On random processes linearly equivalent to white noise. *Inf. Sci.*, 7:133–156.
- Fenerich, A. and Obici, G. (2011). Jardim das gambiarras chinesas: uma prática de montagem musical e bricolagem tecnológica. In *Encontro Internacional de Música e Arte Sonora*, Juiz de Fora - MG.
- Jensen, S., Vos, K., and Soerensen, K. (2010). SILK Speech Codec. Technical Report draft-vos-silk-02.txt, IETF Secretariat, Fremont, CA, USA.
- Karplus, K. and Strong, A. (1983). Digital synthesis of plucked-string and drum timbres. *Computer Music J.*, 7(2):43–55.
- LaBelle, B. (2006). *Background Noise: Perspectives on Sound Art*. Bloomsbury Academic.
- Lucier, A. and Ashley, R. (2012). *Music 109: Notes on Experimental Music*. Wesleyan University Press.
- Lucier, A., Gronemeyer, G., and Oehlschlägel, R. (1995). *Reflexionen*. Edition MusikTexte. Ed. MusikTexte.
- Manovich, L. (2010). Estudo dos software. *FILE: Teoría digital : dez anos do FILE–Festival Internacional de Linguagem Eletrônica, 2010*, 10.
- Nakahodo, L. N. (2012). Pionier I00, de Alva Noto & Ryuichi Sakamoto: uma análise musical a partir do método de Philip Tagg. In *Encontro Internacional de Música e Arte Sonora 2012*.
- Sterne, J. (2012). *MP3 : the meaning of a format*. Durham: Duke University Press. Durham: Duke University Press.
- Tomiyoshi, M. M. (2013). Performances musicais distribuídas através de internet residencial. Mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo.
- Wishart, T. and Emmerson, S. (1996). *On Sonic Art*. Contemporary music studies. Harwood Academic Publishers.

Creating an Ecologically Modeled Performance Through the Remote Manipulation of Multiple Soundscapes

Luzilei Aliel, José Fornari (Tuti)

Núcleo Interdisciplinar de Comunicação Sonora (NICS)
Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 13.083 – 970 – Campinas – SP – Brasil

luzaliel@gmail.com, tutifornari@gmail.com

Abstract. *This article presents and discusses the process of creation and development of a sonic art installation entitled “Firefly Destination”; a multimodal interactive artwork developed in PureData (Pd). This is a sonic artwork that aims to explore the remote interaction of distinct soundscapes manipulated in real-time by a single performer. This also focuses on the main goals of ubiquitous music by tackling the interaction of multiple soundscapes and their relations with the performer and listeners. This work uses Skype – a well-known VoIP (voice over internet protocol) – to allow the interaction of performer and soundscapes as agents of a complex open system. Together they create an interactive meta-soundscape, which is a sonic environment formed by sonic features of distant and distinct soundscapes modulated by the performer. This category of artwork may present aspects of self-organization by emerging aesthetical meaning through the ecological modeling of remote soundscapes linked in a computational networked multimodal performance.*

1. Introduction

Nowadays, with the technological resources that we have, it is possible to ponder on how much geographic distances still separate us. It is known that communication and informational digital technologies enable the interaction virtually immediate (perceptually unnoticeable) of audio and image between individuals located in different parts of the planet. This may create a perceptual dystonia from the senses of hearing and sight, as they become virtually immediate, in contrast to other senses such as touch, smell and taste, that are still bound to the limits imposed by the individual's spatial location. Nevertheless, it may seem feasible that audiovisual interactivity is enough to, at certain extent, provide the urgent needs and basic purposes of human remote communication. In the past decade, technology also enabled the mass production of miniaturized and low-cost sensors (equipments that convert physics quantities into electronic signals), such as microphones, video cameras, accelerometers, gyroscopes, compasses, GPS (global positioning system) and so forth. They allow the acquisition of human location, movement, and gestural data, which can be transmitted remotely in real-time, seemingly instantaneously. The interactivity of these new information and communication systems seems to converge so to enable the exploration of multiple medias, such as: sound, gestures and visual information; as building blocks of a large multimodal experience that may ultimately promote the creation of a sonic environment that is not bounded into a single geographic location but spaciouly distributed.

The acquisition and virtually immediate transmission of multimodal data, allowed the creation of the project here presented, "Firefly Destination" or DP (from the Portuguese expression: “*Destino Pirlampo*”). DP explores acoustic possibilities that can be generated in the triadic interaction of sound, image and gesture. The advent of electronics in music contributed to the reduction of perceptual distance between its creation – given by the composer and performer – and its reception and interpretation – given by the end listeners. The immaterial registration of sound, through audio recording (being it analog or digital) made possible to indefinitely store an acoustic phenomenon which is intrinsically immaterial and fleeting. The advances of signal processing techniques enabled the manipulation, processing and reuse of recorded

sound as well as to collect and transmit these sounds in real-time, which broadened the exploration of new aesthetics in previously unthinkable paths.

However, musical performances are still geographically limited by the imperative need of the members of a musical group to be located in the same physical spot and there remain during all the duration of a performance. Technological advances are still not fully capable of make sure that a formal performance, for example, a classical quartet, could be formed and carried on by a group of musicians spread in remote locations while they play together. There have been several attempts on this matter, as the ones of a well-known series of videos available on www.youtube.com, titled "Virtual Choir", by Eric Whitacre. Yet, such attempts are still in their early stages of development. They still require a great deal of editing, post-production and still, have to abide by the inevitable delay of each and every online data communication.

As suggested, a major problem for remote computer interactivity in musical performances is the delay, also known by the name of latency. This is specially perceived in the interactions separated by large distances. Latency occurs due to the physical limit of the time interval that sound necessarily takes to travel between its acoustical source and reception. According to [Fornari 2010], this is one of the elements that may have contributed to the classical orchestra, during its evolution, to reach an specific limit in its physical size, in terms of the geographical spread of the musicians inside the orchestra, and consequently the size of stages specially built to fit such musical performances. This limit is approximately 34 meters. Sound takes about 1/10 of a second (100ms) to travel 34m. This interval is also the time lag where the auditory system begins to notice the delay in the sound communication, as reverb or even echo. This author named this concept as LIS (Limit of Interact Simultaneity). Any sound communication between musicians – who normally require a strong synchronization (as in the case of a classical music performance) – separated by a distance larger than LIS will tend to have synchronization problems due to the noticeable delay in their sonic cooperation.

Following the same principle, for the interactions between musicians through internet network, the occurrence of latency, disregarding the actual distance between parts, provides the same limit given by LIS. Consequently, similar problems of musical synchronization will be found even in remote interactions. According to [Barbosa 2010], even considering a hypothetical situation where someone had access to the highest possible speed of data processing, noticeable delay would still be observed. This author mentions that if someone consider the smallest possible length of a peer-to-peer connection between two individuals placed in the opposite sides of Earth, they would be separated by an distance of about 20,000 Km (corresponding to half of the distance of Earth circumference, which is about 40.000km). Even with the data transferred at the speed of light (around 300,000 Km/s) and counting with unrealistic unlimited bandwidth, bidirectional latency would still be present, which would last approximately 133.4ms, which is higher than LIS (100ms), as said before, the minimum tolerable threshold of sound synchronization. This points to the fact that there is no way to avoid the occurrence of latency in remote interactions. On that fact, there are recent researches in computer music of performances that support and take advantage of latency as another element of their artistic development and compositional exploration [Traub 2005]. Considering this context, the work here presented do not try to overcome latency but aims to minimize that, making sound communication dynamic and in real-time. DP creates an interactive sound model that incorporates the inexorable systemic latency of any remote communication. DP does not depend on the synchronization between their formant agents, once that the final result aimed y that is the creation of an interactive meta-soundscape; a soundscape made of multiple soundscapes, which is a form of sonic art that is not particularly dependent on agents synchronization, as normally happens in the case of formalized music.

The following sections will describe DP artwork installation and its relation with the concepts of eco-composition and soundscape.

2. The DP project

The main objective of DP project is to explore and promote the correlation and interactive manipulation of sound objects from remotely distant soundscapes through the transmission of its digital audio into a single centralizing and guiding performer. As mentioned before, its original name is in Portuguese language; "*Destino Pirlampo*" meaning: "Firefly Destination". For this reason, this project is here also referred by the initials "DP". The inspirational motif of DP is the comparison of a single centralizer performer in the role of a light source that attracts revolving sound objects like flying insects. It also borrows inspirational motivation from Lorenz's "butterfly effect" (as explained in section 5) where sonic meaning may emerge out of chaos. The poetic license taken here is based on the assumption that if a flying insect, when attracted to a source of light, eventually finds its own death, then a firefly, which has its own source of light, as meeting with a bigger source of light, may find its own self. The performer represents this light source that attracts and manipulates sound objects from distant soundscapes (the fireflies). Altogether they create a meta-soundscape, representing the realization of its own self essence; a transcending soundscape not restricted by a physical location but inhabiting the cyberspace; the immaterial landscape where this virtual soundscape takes place. The performer builds this new soundscape through multiple gestural content as exploring them in a ludic and intuitive manner, dynamically transforming the incoming sound content. This concept focuses on the correlation of remote audio together with visual aid that is created and mediated by digital audio from multiple agents.

Barreiro and Keller [2010] proposes that creative musical or sonic actions can fall into one of three categories: 1) individual activities without direct sonic results, 2) individual activities aiming for immediate audible results; 3) social interaction through the active participation of multiple users. This last category may imply that the interaction between individuals immersed in their soundscapes might foster a new type of action, the adjust or realization of themselves as agents in a new environment [Barreiro and Keller 2010]. In DP project, the environment can be appointed as a content originated from almost all possible sonic interactions. Much of the sound material essential for the construction of this soundscape comes from this interaction. This reciprocity occurs throughout cyberspace, where participants (in the role of agents) send real-time audio content that are processed and controlled in a single station, where the performer and the audience are located, thus providing the possibility of creating a meta-soundscape. These characteristics are aligned with the principles of ecological model, as defined by Keller [2004]. According to this author "Ecological models are the application of sound synthesis processes of interaction between agents and objects in the environment. Each iteration generates a sound event located in a location of space-time. The exchange between agents and objects changes the state of the organism and the material, adapting objects features to the needs of these agents, thus forcing them to change their behavior depending on the profile of its ecological niche. The sound is only one of the products of this interaction, which makes ecological modeling also multimodal [Keller 2004].

In the scope of ecological modelling, there are two types of related strategies for the interaction between composer and listener: 1) the application of artificial niches (the ones not found in nature) and 2) the exploration of natural affordances; natural channels of interaction common to humans in ecological niches [Barreiro and Keller 2010]. The concept of "sonic affordance" is characterized by relationships of sound interpretation between each sound object, each one having its own form of perception. Such forms of perception can arise from multiple factors, making this perceptual experience unique. The creational process of DP is related to the interaction between listener and composer allowing exchanges of information for the production of natural and artificial niches. In the brink of such exchanges the resulting sound may present original characteristics, while still being similar, which is a condition aligned to the fundamental principles of soundscapes and eco-composition, as following defined.

3. From Eco-composition to Soundscape

It might be possible to analyze electroacoustic music from three distinct perspectives: 1) the composer, 2) its sound material and 3) the creational process. Acousmatic music – which aims to seek and use fundamental aspects of sound, setting them apart from any reference or context other than sonic – is a direct descendant of Schaefferian thought and primarily focuses on the figure of the composer [Chion 2009]. This material is compiled and treated as the composer's aesthetics pleases. In a nutshell, acousmatic music can be defined by an aesthetic posture that is based on the composer understanding, in the role of a unquestionable designer of all compositional decisions [Barreiro and Keller 2010]. This sturdy posture of placing the composer in first place and so in absolute control of sonic creation contrasts with other approaches, such as the soundscape composition [Shafer 2001]. Unlike acousmatic music, soundscape composition focuses primarily on the sound material. This one began at Simon Fraser University, in Canada, where composer and educator Murray Schafer established the “World Soundscape Project”. Schafer coined the neologism “soundscape” following an analogy with a landscape of sounds, by mixing the words “sound” and “landscape”, thus allowing the inclusion of extra-musical sound material in contemporary musical creation [Keller 2004], as well as the classification of virtually any physical environment as an element of spontaneous creation of soundscapes [Schafer 1997]. Natural soundscapes are generated by multiple sound sources which can be represented as external agents of an open and complex system. Through the process of self-organization, sonic meaning might emerge from this system, in the form of perceptual and cognitive features that undoubtedly characterizes and distinguishes each and every soundscape, where acoustic information is never repeated but its cognition is always similar. For this reason, generating a soundscape, or rather composing it (or even composing a piece of music with sonic elements of natural soundscapes) is not a trivial task. Soundscapes are elusive sonic entities. As a matter of fact, once recorded, a soundscape actually ceases to exist. The recording of a soundscape is merely an immaterial registration of a sound phenomenon already over, that has occurred in the past, at the time when it was recorded, as a dynamic, cognitively similar and acoustically unique auditory phenomenon. However, the term “soundscape composition” usually refers to the type of composition using sound material recorded from actual soundscapes. One usual type of soundscape composition is named eco-composition. This is based on the organization of audio fragments of soundscapes in modular structures that form a musical piece. Thus eco-composition creates compositional structures using simple concepts such as modularity and the accumulation of sound material. By means of interactivity, the composer (or even the listeners) can also dynamically adapt the compositional structure of the piece, thereby creating a listening experience which is immersive for the listener that is also one of the composers of such sonic environment. These are definitive and essential characteristics of eco-composition, that focuses on the compositional process rather than in the sound material or the composer.

The work here introduced aims to extend the concept of eco-composition. This paper discusses the possibility of a dynamically intertwining soundscapes materials arriving simultaneously from different and remote locations. Instead of using recorded sounds, DP uses actual audio streaming straight from soundscapes dynamic generation. The DP project welcomes this contrasting mix of distinct soundscapes manipulated by a single performer that processes and blends all sound material to create a new soundscape which transcends the concept of “source bonding”, as defined in [Smalley 1990]. This one deals with the cognitive bias, or gestaltic association, that automatically associates sound objects to their respective source contexts. This referential linkage can be understood as an informational flow between sound object and sound source. In DP, this flow is handled by the performer that sometimes makes this association clear, while at other times, such association seems to break apart, giving away the referential link from sound and source. In the meta-soundscape resulting from DP dynamic compositional process, this informational flow enables the emergence of sonic meaning and therefore the understanding and appreciation of this artwork. This is necessarily given by a set of cognitive concepts in common between composer and listener, which permeates the artwork and its audience. Just as in verbal language, where the meaning

of words and structures can be understood (decoded), in a non-verbal sonic arts where it is essential that the listeners know the basics of the language in which the discourse is being developed [Denora 2000].

As seen, the interaction is part of the proposition of eco-composition, even collective interaction, where the figure of a single the centralizing composer becomes diffuse or less present. In such systems, materials, methods, format and audience are the integral parts of this compositional process, having the experience of this act as highly significant for the artistic endeavors [Keller 2004]. It is found in this method an open proposal, validating changes between agents and objects in a pre-determined space-time location. DP expands the concept of eco-composition, enabling communication between agents and physically separated objects, thus manipulating them and transcending the obstacle of a single space-time location. As said, DP aims to create a meta-soundscape through sound material acquired from distant and distinct natural soundscapes, where the performer, as the centralizing agent, is the handler of this new soundscape, created by the interactive modification of multiple dimensions of sonifications, thus approaching DP from the compositional processes of ubiquitous music.

4. Ubiquitous Music

In basic conceptual terms, ubiquitous music is made by the interaction of multiple users, handling a myriad of electronic informational technologies. This concept is related to the fusion between computing and natural environment, as proposed by Mark Weiser in late 1980s [Weiser 1991]. As a first condition, one can conceptualize ubiquitous music as coming from the previous context of ubiquitous computing, where the advent of mobile computing, independence and interconnection, devices heterogeneity, wireless infrastructure, pervasiveness, context-awareness, mobility and portability are complementary, at the end, all aiming to be incorporated to a cybernetic environment. These associations tend to allow a dynamic network of interactions between agents, objects and environments [Weiser and Radanovitsck 2011].

The term “ubiquitous”, as used in relation to computing, refers to the existence, availability and access of multiple individuals to a network of information integrated into their daily lives, as is the cyberspace. This same premise is applied in DP, where the sound information is sent in real time by several remote agents. This enables a huge range of sonic possibilities while keeping a reasonable dynamic handling. In order to establish a methodology for modeling ubiquitous music, Fornari [2010] mentions six possible categories: 1) the development of computational models for the acquisition of artistic gesture and its remote transmission, 2) the development of computational models of real-time sound synthesis, which are the core of virtual musical instruments; 3) the integration between these models of sound synthesis and gestural acquisition; 4) the real-time interactivity and intuitive exploration of sound material; 5) the experimentation of acousmatic extended musical techniques; 6) the application of ecological modeling.

In DP project, all areas mentioned above were met, as explained in section 6. Here, massive sonorities containing natural soundscapes are received, manipulated and synthesized by a single performer. These tend to be guided by a non-deterministic control, since such sound events often have a high degree of randomness and unpredictability. For example, if one of the agents is in a park, sending sound content for the DP performer, and suddenly a bird approaches and starts tweeting, this event will be unpredictable for the performer, as well as unlikely to be similarly repeated. The chaotic element that constantly permeates the production of DP refers to the concept of self-organization, which will be treated as follows.

5. From Butterfly Effect to Self-Organization

The butterfly effect is a famous term commonly used to refer to the theory developed by Edward Lorenz, in mid 1962, which is a mathematical model for a certain class of chaotic events. Lorenz, a meteorologist, found that in some circumstances very small variations in the parameters of certain mathematical models would turn to be impossible to forecast a predictable results, for instance, the provision of a particular climate state. The effects of these changes were both quantitative and qualitative. These observations prompted Lorenz to coin the following famous and nowadays memetic phrase: "the flapping of a butterfly's wings in one side of the planet can unleash a tornado on its other side"¹

Chaotic systems have at their core the principle of taken into consideration both order and disorder, considering them as valid the aspects of uncertainty. Naveira [1998] proposed that chaos theory should be understood as an alternative for the determinism of mathematical formalization, where the results are perfectly predictable; and the uncertainty of stochastic models, where the occurrence of events can not be predicted. It should be noted that the use of the term "chaos", as commonly associated disorder, is here misleading. In science, this term does not necessarily refers to disorder, but to "order masqueraded as randomness" [Naveira 1998]. This concept is used in DP. All material sent via network is foremost taken as valid input, emphasizing them not as disorder or mere error but as masked order guided by randomness. As explained in section 2, the very title of DP ("Firefly Destination") is based on this aesthetical premise, associating the flapping of the wings and the interaction of an insect also with the butterfly effect.

The feature of spontaneously organizing a chaotic stream of data into some form of meaningful information is here seen as a process of self-organization. This concept, in a way, intertwines and formalizes several other concepts nowadays in use, such as "networked organizations", "self-managing teams" and "learning organization" [Naveira 1998]. Relating music and self-organization, Mazolli [1996] presents a proposal to create cycles of sound material occurring through reflexion and feedback. These are: 1) the manner in which the composer begins the work; 2) the number of cycles of sonic interaction; 3) the musical form that gradually arises. Manzolli points out that self-organization in a compositional process tends to become an advantageous tool once it allows spontaneity to be part of the process, as one of the parameters within a complex system. Such possibilities allow the adaptation of interactive models to be incorporated as a way of boosting the composer's creative autonomy. Compositional systems based on interactive models, as defined by Mazzolli [1996], particularly refer to the aspect of interactivity, disregarding the magnitude or the duration of sonic events. What that matters is the compositional dynamics of its temporal merging. This happens in the musical field where composer and performer become parts of a single compositional agent, which enables interactive and improvised structures to spontaneously occur [Mazzolli 1996].

1 <http://blog.ap42.com/2011/08/03/the-butterfly-effect-variations-on-a-meme/>

6. Method and Results

The computational model of DP was developed in the programming environment of Puredata, also known as Pd (www.puredata.info). Pd is an open-source visual programming platform for real-time data control, analysis, transformation and synthesis. Pd has an extended version named Pd-extended, that comes with several external libraries aggregating capacities other than the ones already present in Pd. Pd-extended, as it is called, can handle several types of data, such as: symbolic; numerical scalar and vectorial; MIDI and OSC; digital audio, image, animation, and video; text files; network protocols; among others [Fornari, 2010]. For the computing development of DP, it was initially used a version of Pd-extended running in Linux operating system. However, there were technical issues in the communication of digital audio data with the computational model of DP that seems to only be possible to be solved by the migration of this model to another operating system, such as Windows. A computational model developed in Pd is called a "patch". For the DP project, 3 patches were built. As mentioned in section 4, this model covers all six areas of ubiquitous music, as further explained.

The first patch has visual characteristics which are modified dynamically in relation to their geometric shapes and colors of graphical objects, according to the received audio data and gestural data. This patch also contains the module of gestural data acquisition that acquire real-time data from the performer's gestures (see next figure). They are dynamically captured by the webcam of the machine (a common laptop) where this patch was running. The fact that this is a bodiless gestural interface, as described in [Fornari 2012], allowed the performer to explore intuitive control of the process while keeping the freedom of movements and gestural exploration. Next figure shows a detail of a typical performance of DP.

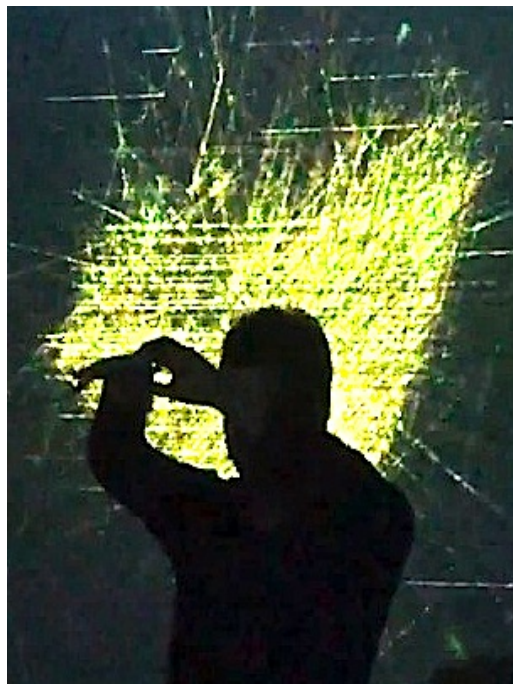


Figure 1: Detail of the performer gestures during a DP presentation, in front of the complex geometric structure (the green-yellow structure at the background) that is dynamically generated by the computer model (the first patch described above).

The second patch contains a subtractive sound synthesis model that emulates the typical sound of wind, by using a white-noise generator whose digital audio produced is processed by dynamic band-pass digital filters. This synthesis model aims to provide background sonification for the dynamically controlled digital audio streaming received from multiple soundscapes. This “wind-sound” synthesis is controlled by the intensity of gestures, captured by the bodiless interface of the first patch. Details of implementation of this gestural controlled synthesis can be found in [Fonari 2012]. The third patch is a model of digital audio acquisition that receives data from a VoIP (voice over internet protocol) software. In this experiment, the VoIP used was Skype (www.skype.com), a well-known cross-platform application software that allows real-time voice communication of multiple users connected throughout internet. The artistic proposal here presented uses audio data from multiple agents remotely located, realtime connected in conference-call through VoIP, where all audible information data comes as a single audio stream from this connection. This stream of audio, received by the third patch is mixed with the subtractive synthesis of the second patch, which is intuitively controlled in real time by the performer of DP, and also controls the realtime animation rendered on-the-fly by the first patch. Together these 3 patches create a multimedia experience that is immersive, dynamic and ubiquitous.

During the implementation of this project, it was found that digital audio communication between Skype and Pd was not the trivial task previously expected to be implement in Pd running on Linux. While researching and trying some freeware softwares applications, it was found a simple open-source application, developed only to run in Windows, that was able to successfully performed the required communication task. This is the application called Virtual Audio Cable (VAC); a free program able to transfer audio streams from one application to another. If there may be other available open-source applications with similar function, the authors tested only solutions for Linux and Windows, where they found only this one that were successfully able to recognize and to be recognized by Skype. In order to enable audio transfer from Skype to Pd through VAC, one should create two "virtual wires" in this softwares, which will feature input and output of multiple audio channels. Once these virtual cables are created, Skype and Pd will be instantaneously recognized and connected. The following figure shows a snapshot of the integration of these 3 pieces of software (Pd , Skype and VAC) running in Windows.

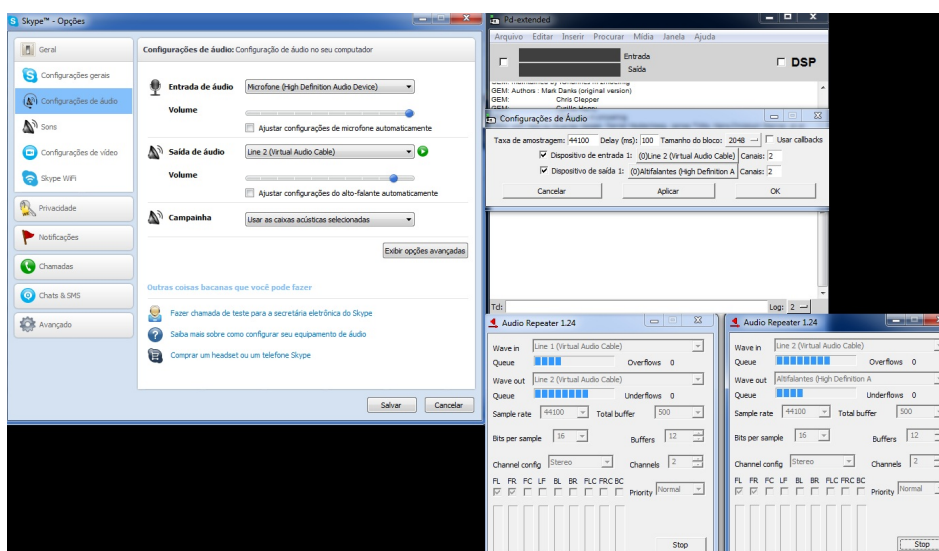


Figure 2: Software integration between: Pd, Skype and VAC; for the DP project.

7. Conclusions

This project aimed the artistic exploration of creating what is here referred by the coined name: “meta-soundscape”; a soundscape created by the audio data acquired and transformed in realtime from multiple soundscapes. This is done by overcoming physical limitations of locality throughout technological resources that allow the apparent reduction of distance between soundscapes remotely located by streaming audio realtime collected into a single location, where the artistic performance of DP takes place.

The fact that more and more individuals are connected through cyberspace, sharing so many types of data and forms of media with multiple users, makes artistic projects such as DP a “natural” consequence. This is a multimodal self-organized performance of social networks which is based in one interactive system that mediates several soundscapes remotely located. Audio data from multiple soundscapes are transmitted via VoIP, enabling the creation of a meta-soundscape that is dynamically manipulated by a single performer. As such, DP is not iconoclastic, but represents a continuation of artistic soundscape composition exploration. While preserving a centralized figure of a single performer, the compositional role is shared among all agents, acting as the producers of new sonic possibilities and meanings, expressed in the resulting meta-soundscape. One of this project premises is the advent of chaos as an enabler of a self-organizing process to generate meta-soundscapes. Through a simple association with the butterfly effect, this project also accepts and uses unpredictable sonic results while maintaining the typical soundscape self-similarity. Its computing development focused on the creation of a modular interface capable of taking multiple streams of sound patterns that can be controlled in real-time during the course of a single performance. This process was designed and implemented through an interconnection between the computational model of DP (compounded of 3 Pd patches), a VoIP (that handles audio data from the soundscapes to stream into Pd), and VAC (that transfer audio data, bridging the VoIP and Pd). This ensures that all initial goals of DP project were achieved and even surpassed. The concept of the concurrent handling of multiple soundscapes extends the premise initially settled by eco-composition, enabling ubiquitous exploration of multiple streams of sound material, which is here understood as a valuable contribution to future researches and artistic endeavors for contemporary computer music composers.

References

- Barbosa, A. (2010) "Performance Musical em Rede". In: Damián Keller e Rogério Budasz (ed.). Criação Musical e Tecnologias: Teoria e Prática Interdisciplinar. Goiânia: ANPPOM, p.97-126. http://anppom.com.br/editora/Pesquisa_em_Musica-02.pdf , Junho.
- Barreiro, D. L. e Keller, D. (2010) "Composição com modelos sonoros: fundamentos e aplicações eletroacústicas". In: Damián Keller e Rogério Budasz (ed.). Criação Musical e Tecnologias: Teoria e Prática Interdisciplinar. Goiânia: ANPPOM, p.97 -126. http://anppom.com.br/editora/Pesquisa_em_Musica-02.pdf , Junho.
- Chion, M. (1982) "La musique électroacoustique". Paris: PUF.
- Keller, D. (2004) "Paititi: a multimodal journey to El Dorado". Tese (Doutorado em Artes Musicais) Stanford University. Stanford, CA. <http://portal.acm.org/citation.cfm?id=1048456> , Junho.
- Fornari, J. (2010) "Percepção, Cognição e Afeto Musical". Capítulo de livro da ANPPOM. Título do Livro: "Criação Musical e Tecnologias: Teoria e Prática Interdisciplinar". ISBN:978-85-63046-01-7.
- Manzolini, J. (1996) "Auto-organização um Paradigma Composicional". In Auto-organização: Estudos Interdisciplinares, Campinas, CLE/Unicamp, ed. Debrun, M. Gonzales, M.E.Q. Pessoa Jr. O. p.417-435,
- Naveira, R. B. (1997) "Caos e complexidade nas organizações". Dissertação de Mestrado. São Paulo: FGV.
- Radanovisck, E. A. A. (2011) "MixDroid: compondo através de dispositivos móveis". Rio Grande do Sul: UFRGS. (Trabalho de Conclusão de Curso). <http://www.lume.ufrgs.br/bitstream/handle/10183/31037/000782068.pdf?sequence=1>, Junho.
- Schafer, R. M. (2001) A afinação do mundo. São Paulo: Ed. UNESP.
- Traub, P. "Sounding the Net: Recent Sonic Works for the Internet and Computer Networks". Contemporary Music Review. Vol. 24, No. 6, December 2005, pp. 459 – 481
- Smalley, D. (1990) "Spectro-morphology and Structuring Processes In The Language of Electroacoustic Music", ed. Emmerson, pg. 61-93.
- Denora, T. (2003) "After Adorno: Rethinking Music Sociology". Cambridge University Press. UK.
- Weiser, M. (1991) "The Computer for the Twenty-First Century". Scientific American, v. 265, n. 3, p. 94-101.
- Fornari J. (2012) "Designing Bodiless Musical Instruments". Anais da AES Brasil 2012. 8 - 10 maio. Expo Center Norte. São Paulo. Brasil.

Websites list

<http://www.youtube.com/user/nicsunicamp>

<http://puredata.info/>

<http://www.skype.com/pt-br/>

<http://www.virtualaudiocable.net/>

Avaliando Interfaces Gestuais para Prática de Instrumentos Virtuais de Percussão

Jaziel Vitalino Souza e Silva¹, Filipe Calegario¹, Giordano Cabral², Geber Ramalho¹

¹ Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50.732-970 – Recife – PE – Brasil

{jvss,fcac,glr}@cin.ufpe.br

² Departamento de Informática e Estatística (DEINFO) – Universidade Federal Rural de Pernambuco (UFRPE)

Rua Dom Manoel de Medeiros, 52.171-900 – Recife – PE – Brasil

{giordanorec}@gmail.com

Abstract. *In this paper, we analyze and evaluate the interaction of a set of gestural interfaces in order to identify technical limitations and to correlate them with the low adoption of these technologies in practice. Through literature researches, we made a mapping of gestures used in percussion instruments, and conducted an experiment where professional drummers evaluated the interaction offered by two interfaces: the Microsoft Kinect Sensor and the Nintendo Wii remote.*

Resumo. *Neste trabalho analisamos e avaliamos a interação de algumas interfaces gestuais atuais, no intuito de identificar limitações técnicas e de que forma estas explicam a baixa adoção de tais tecnologias na prática. Através de pesquisas bibliográficas, fizemos um mapeamento dos gestos mais utilizados nos instrumentos de percussão e realizamos um experimento onde profissionais percussionistas avaliaram a interação oferecida por duas interfaces: o Sensor Microsoft Kinect e o Nintendo Wii remote.*

1. Introdução

No século XX, fatores como o avanço do conhecimento sobre aspectos físicos e cognitivos do som, o advento dos instrumentos elétricos e a aplicação das tecnologias eletrônica e digital na geração sonora artificial mudaram a relação entre música e tecnologias eletrônicas e digitais (Iazzetta, 1997). Pesquisas acerca do desenvolvimento de novas interfaces digitais têm crescido principalmente nas últimas décadas e isto pode ser comprovado com o crescimento de conferências como o NIME (New Interfaces for Musical Expression). Instrumentos virtuais, construídos a partir de acelerômetros e câmeras com sensores de profundidade, permitem a interação do usuário através de gestos e a produção do som através de algoritmos de síntese sonora. Boulezz (1977) menciona que estes dispositivos não foram criados essencialmente para aplicações musicais, mas a sua utilização leva a mudanças que podem resultar em novas soluções.

Apesar de tanta tecnologia empregada na criação de novos instrumentos, por que não é frequente vermos músicos profissionais utilizando essas tecnologias em performances reais? Por que a maioria dos músicos não usam as novas interfaces disponíveis? Esta é a questão que esta pesquisa se dispõe a investigar.

2. Interfaces Digitais

Os instrumentos musicais digitais (*digital musical instruments*, DMIs) são sistemas digitais compostos por um módulo de controle de entrada, que são interfaces que capturam a ação do músico, e um módulo de produção sonora, que efetivamente realiza a síntese sonora. Fazendo a ligação entre a entrada e a saída dos DMIs, temos as estratégias de mapeamento.

Dentre os DMIs mais conhecidos, estão os que simulam instrumentos acústicos existentes, como por exemplo: um saxofone MIDI; uma bateria eletrônica, geralmente formada por um conjunto de pads, montados sobre um rack, na mesma disposição de uma bateria convencional.

Entretanto, há também DMIs com dispositivos de entrada alternativos, que não se baseiam em instrumentos existentes (MIRANDA e WANDERLEY, 2006, p. 19). Tais DMIs não possuem padrão de tamanho físico ou forma, podendo ser desenvolvido para funcionar em iPads ou Smartphones, softwares e sintetizadores que simulam teclados, violões, guitarras, contrabaixos e outros instrumentos virtuais. Por exemplo, o Reactable¹ é uma interface em que o usuário interage através de objetos do mundo físico (blocos). Tais objetos possuem diferentes funções na sua execução musical, como exibido na Figura 1.



Figura 1 – Interface do Reactable

3. O Gesto Musical

Pode-se assumir que o gesto é um elemento essencial na execução musical, já que os músicos tocam instrumentos por meio de gestos. Por exemplo, são os gestos dos dedos do violonista que excitam as cordas do violão, provocando ondas sonoras que são amplificadas pelo corpo (caixa acústica) do instrumento, para eventualmente serem depois captadas por microfones ou outros sensores e amplificadas eletronicamente.

A tecnologia de construção de instrumentos sempre levou em consideração: a ergonomia, levando em conta a facilidade ou dificuldade com que os músicos interagem com os instrumentos; e a geração sonora, preocupando-se com a forma com que o som é gerado, ou seja, com a qualidade sonora do instrumento.

A preocupação com gestos acontece não só na criação de instrumentos acústicos, mas também na chamada luteria digital (JORDÀ, 2005). Enquanto os instrumentos acústicos tentam evoluir em qualidade física e timbrística, os instrumentos digitais tentam ganhar espaço, por vezes procurando criar instrumentos musicais com modelos alternativos que busquem novas formas de expressão musical.

¹ Disponível em: <<http://mtg.upf.edu/project/reactable>>. Acesso em: 06 Ago. 2013.

4. Interfaces Gestuais em Computação

Baseando-se em conceitos da área de Interação Humano-Computador (IHC), Wanderley (2006) preconiza que sistemas de geração sonora controlados por gestos podem ser utilizados em vários contextos da interação musical. Desenvolvedores de software tem dado cada vez mais importância a estudos relacionados a gestos e movimentos do corpo (Corrêa et al, 2005). A motivação para isto é poder “herdar” um grande repertório de gestos que os seres humanos desenvolveram ao longo dos anos fazendo com que a interação deles com o computador seja mais natural.

Temos visto se intensificarem pesquisas sobre interfaces gestuais, com aplicações em diversas áreas. Dentre estas interfaces gestuais, algumas merecem maior destaque por sua popularidade:

- **Tela Multi-toque** - é uma tecnologia de interação que reconhece múltiplos contatos simultâneos na tela e possibilita a manipulação de sistemas através do toque de algum objeto ou dos dedos. Exemplos: Tablets, Smartphones e Mesas Multi-toque.
- **Novos Controladores de Jogos** – O Nintendo Wii trouxe o conceito de Nunchuck, um novo tipo de controlador de jogo (como foram os joysticks e joypads). O controlador é segurado com as mãos e se utilizam os dedos para pressionar os botões de comando. Exemplos: Wii Remote e PlayStation Move.
- **Câmera de Detecção de profundidade** – Dispositivo que cria novas maneiras de executar tarefas dispensando a necessidade de contato físico com o instrumento de trabalho. Exemplo: Microsoft Sensor Kinect.

O número de possibilidades é imenso e os aplicativos musicais que utilizam tais interfaces surgem em grande quantidade. Mas uma dúvida perdura: por que não é com frequência que vemos músicos profissionais utilizando essas tecnologias em performances musicais? A maioria dos músicos não usam essas novas interfaces. Isto tem a ver com a falta de qualidade, preconceito, aspectos culturais?

O objetivo desta pesquisa é analisar os aspectos técnicos, identificando as limitações dos sistemas existentes na tentativa de levantar possíveis hipóteses sobre a não utilização destas interfaces mais frequentemente pelos músicos profissionais.

Na próxima seção faremos um mapeamento de gestos instrumentais, procurando relaciona-los com as possibilidades de gestos reconhecíveis por algumas interfaces.

5. Mapeamento de Gestos

Visto que os sensores das interfaces digitais ainda não estão suficientemente refinados e não conseguem captar movimentos curtos com precisão (por exemplo, os dos dedos), neste trabalho foram escolhidos para a análise, alguns instrumentos musicais convencionais que utilizassem gestos mais largos, particularmente os de percussão.

5.1 Gestos Instrumentais

No intuito de relacionar gestos humanos, possivelmente reconhecíveis pelas interfaces e os gestos usados na prática de cada instrumento, foi feita uma tabela descrevendo os gestos de maior relevância, usados na prática de alguns dos instrumentos de percussão (Tabela 1).

Tabela 1 – Tabela descritiva de instrumentos e gestos

Instrumento Musical	Movimentos com o polegar	Movimentos com os demais dedos	Movimentos com o punho	Flexão e extensão do antebraço	Movimentos de rotação com o antebraço	Movimentos com ombro e braço
Agogô 			X	X		
Alfaia 			X	X	X	X
Berimbau 	X	X	X	X		
Bongô 	X	X	X	X		
Caixa 		X	X	X	X	X
Cazari 			X	X	X	
Cajon 		X	X	X	X	X
Castanholas 		X	X			
Conga 		X	X	X		
Cuica 		X	X	X		
Derbake 		X	X	X		
Djembê 		X	X	X		
Ganzá 			X	X		
Maracá 			X	X		
Moringa 		X	X	X		
Pandeiro 	X	X	X	X	X	
Pratos 			X	X		X
Reco-reco 			X	X		
Repique 	X	X	X	X	X	
Sinos 			X	X		
Surdo 			X	X		X
Tabla 	X	X	X	X	X	
Tamborim 			X	X	X	
Timpano 			X	X		
Triângulo 	X	X	X			
Xilofone 			X	X		
Zabumba 		X	X	X		X

5.2 Análise de Captação de Gestos

Abaixo visualizamos uma tabela com a descrição de gestos primitivos humanos e duas interfaces em análise. Foram, assim, mapeados os gestos possíveis de reconhecimento pelas interfaces escolhidas nesse trabalho.

Tabela 2 – Tabela de interfaces

Gestos	Sensor Kinect	Wii remote
Movimentos com o polegar		x
Movimentos com os demais dedos		x
Movimentos com o punho	x	x
Flexão e extensão do antebraço	x	x
Movimentos de rotação com o antebraço		x
Movimentos com ombro e antebraço	x	

Percebemos que há uma demanda de gestos que não são captados pelas interfaces gestuais. Esses gestos são usados na prática de diversos instrumentos de percussão e sem eles os instrumentos perdem características técnicas importantes. Sendo assim, com os resultados descritos de quais gestos mapeáveis são tecnicamente úteis, captáveis pelas interfaces, iremos descrever um experimento onde foi avaliado esse subconjunto de gestos possíveis, nas duas interfaces gestuais.

6. Método

Pretendendo avaliar a interação das interfaces do ponto de vista do usuário, propomos, com base no método de Barbosa (et al, 2011), realizarmos um experimento, onde vários participantes com perfis musicais distintos fossem orientados a usar as interfaces pela primeira vez.

Foram convidados músicos percussionistas profissionais que atuam na área e possuíssem diferentes perfis artísticos, visando obter um resultado mais satisfatório. Os perfis escolhidos foram: estudantes universitários da área erudita, estudantes universitários da área popular e percussionistas autodidatas. Passando pela mesma série de tarefas, cada experimento levou cerca de quarenta e cinco minutos até sua conclusão.

6.1 Coleta de Dados

O recolhimento dos dados foi concluído a partir de três fases:

- Exploração livre: O usuário é orientado a usar a interface por algum tempo sem restrições, à sua própria vontade.
- Exploração guiada: O usuário escuta algumas amostras de áudio com intuito de experimentar possibilidades, inspirando-se na amostra.
- Entrevista semi-estruturada: Nesta fase o usuário é incentivado a discutir sobre a sua experiência realizada com as interfaces gestuais nos momentos de exploração livre e exploração guiada.

6.2 Análise de Dados

Na análise de dados, as informações obtidas no experimento foram examinadas em três eixos: controle temporal – o qual diz respeito à latência do som; variação timbrística – diz respeito às mudanças de coloração sonora oferecidas pelos instrumentos; e controle de dinâmica – que diz respeito ao controle de intensidade do som, mais forte e mais fraco. Nesta análise, utilizamos o método conhecido como análise do discurso (AD).

6.2.1. Análise do Discurso

A análise do discurso possui 5 etapas:

1. **Transcrição** – Neste passo todo material falado é transformado em texto.
2. **Livre associação** – etapa em que as impressões são anotadas para posterior análise.

3. **Discriminação** – Etapa em que todo texto é quebrado em vários pedaços e os elementos mais comuns são reconhecidos, fazendo menção das terminologias usadas na entrevista.
4. **Reconstrução do mundo do usuário** – Etapa em que o analisador constrói uma representação de cada experiência, na perspectiva do usuário com base na lista de elementos mais comuns utilizados e suas relações.
5. **Exame do contexto** – Por final a representação é comparada e analisada.

7. Experimento

O experimento iniciou-se com a apresentação da primeira interface, o *Kinect*, figura 3. Nele foram mapeados gestos para dois instrumentos percussivos: um Ganzá (tipo de chocalho); e uma Conga (tipo de tambor). O Percussionista recebeu a instrução de que teria um momento livre (*free exploration*) onde ele poderia testar ritmos e reconhecer que gestos poderiam ser utilizados no instrumento virtual. O momento livre tinha duração média de dois minutos. Em seguida o percussionista testou o instrumento sendo acompanhado por um metrônomo em três andamentos diferentes (bpm - batidas por minutos): 90 bpm, 120 bpm e 160 bpm. Logo após o experimento continuou com acompanhamento de uma harmonia servindo como base em ritmos de Baião em 80, 100 e 120 bpm; Frevo em 100, 140 e 160 bpm; Maracatu em 80, 100 e 120 bpm; Sambinha em 80 bpm e um ritmo latino americano 120 e 140 bpm. Após às performances com cada instrumento virtual, era feita uma sessão de comentários pelo usuário, falando sua opinião sobre a interação do instrumento musical usando a interface. O mesmo ciclo se repetiu com todos os instrumentos experimentados no *Kinect*.

Em seguida foi apresentada a segunda interface, os controles Wii, figura 4. Nele foram mapeados três instrumentos que utilizam baquetas, aproveitando a formato de bastão e os sensores usados na interface. Os instrumentos virtualizados pelo Wiimote foram: uma Caixa e uma Zabumba (Tambores) e um Agogô (Espécie de sino duplo oco). O experimento com o *Wiimote* seguiu o mesmo ciclo de tarefas usadas no *Kinect*.

Na web é possível visualizar a gravação de alguns trechos do experimento, mostrando os usuários usando os instrumentos virtuais nas interfaces gestuais².



Figura 3 – Perspectiva do usuário com o Kinect

² Disponível em: zabumba virtual <<https://www.youtube.com/watch?v=7PfWE-V8lmk>>, conga virtual <https://www.youtube.com/watch?v=pVkJ_5_OyW0>, caixa virtual <<https://www.youtube.com/watch?v=6IZRyFQ7q4M>>, agogô virtual <<https://www.youtube.com/watch?v=tfNfcmkXPA>> - Acesso em: 06 Ago. 2013.



Figura 4 – Usuário usando o Wii Remote

7.1. Análise do Discurso

7.1.1. Transcrição

Como citado anteriormente em uma das etapas do método utilizado, o material de áudio coletado no experimento foi totalmente transcrito e encontra-se disponível na web³.

7.1.2. Discriminação

Percebemos que os participantes não se sentiram a vontade diante das interfaces e mesmo tendo conhecimento em informática e experiência musical no seu instrumento, levaram um bom tempo para se adaptar a ideia de tocar em um modelo virtual de percussão. Ao experimentar pela primeira vez as interfaces, os movimentos comuns, utilizados na prática do instrumento acústico eram lembrados e automaticamente gesticulados, buscando atingir a mesma sonoridade no instrumento virtual. Porém, quando não conseguiam o retorno esperado, era notório o modo como se sentiam perdidos. Ao passar por essa experiência o usuário 2 comentou:

“Me senti como se nunca tivesse tocado na minha vida, parece que nunca toquei e não sei nada sobre música (...) É diferente do contato com o instrumento material, mas a sensação foi essa, assim, estou mais perdido do que tudo aqui num é?”

Isso acontece também, por causa da importância que os músicos dão a presença do corpo do instrumento, porque ao tocar um instrumento convencional, além de ter um retorno auditivo, ele tem retorno tátil, o que provavelmente deixa o usuário sem confiança diante de uma interface que se manipula apenas por gestos no ar. O limite em que o gesto deve parar é o ponto de ataque onde o músico espera atingir. Contudo, para ele, este ponto precisa ser previsível, pois enquanto o músico toca, ele calcula o tempo em que deve soar à próxima nota e o gesto que será utilizado deve fazer um percurso que atinja no tempo previsto, o ponto de ataque. O usuário 4 chama de instrumento imaginário e ainda salienta:

“Porque é aquela questão, aqui tá a cadeira, eu calculo, automaticamente minha vista calcula onde eu vou atacar e no imaginário isso dificulta.”

Algo que é bastante relevante para o músico é a variação de sonoridades. Em todos os experimentos os percussionistas procuraram variações que segundo eles eram importante para execução de algum ritmo, como por exemplo: a caixa no frevo, que necessita de acentuações o tempo inteiro e sem elas, o ritmo fica sem sentido ou a

³ Disponível em: <http://www.4shared.com/rar/0WDAkDV4/Experimento_Percusso_interativ.html?> - Acesso em: 06 Ago. 2013.

zabumba no baião, que precisa da variação timbrística da pele superior para fazer o swing.

A percepção que o músico tem ao tocar o seu instrumento é a mesma que ele usa para tocar em um instrumento digital, por tanto ele procura usar os mesmo gestos para obter as variações sonoras de timbre, altura e intensidade.

Dentre os problemas citados pelos usuários, os mais relevantes foram: a falta de variação de intensidade das notas, dada a imprecisão encontrada em alguns movimentos; variação de timbres no mesmo instrumento e atrasos em movimentos rápidos.

7.4.3. Reconstrução do Mundo do Usuário

Abaixo, encontra-se um resumo da entrevista semi-estruturada, onde alguns usuários descrevem suas experiências com as interfaces em suas perspectivas, salientamos que uma versão completa desta seção está disponível. Essa fase é uma representação do mundo do usuário, resultado da etapa AD e pode nos ajudar a entender limitações na interação entre usuário e interface.

Usuário 1 - Foi um músico de perfil autodidata. Toda sua experiência musical se deu participando de grupos musicais que tocam na noite e percorrem pelo Brasil. Não teve experiência com instrumentos virtuais, mas tem simpatia por novidades, desde que traga contribuições ao seu trabalho. Apesar de achar que não há condições de tocar com os atrasos encontrados, gostou da experiência e viu nas interfaces pontos positivos referentes à performance. O usuário achou a novidade muito interessante, principalmente por causa do problema de logística (transporte) que sofre o percussionista e acredita que no mercado, as interfaces iriam ajudar bastante. Ele percebeu que havia diferença na captação entre gestos mais lentos e mais rápidos na maioria dos instrumentos, menos na zabumba. No início teve dificuldades, mas salientou a necessidade de conhecer primeiro o equipamento, pra assimilar os gestos e no dia a dia domina-lo.

Usuário 2 - Foi um músico de perfil erudito. Aprendeu em escola de música e estuda percussão erudita na universidade. Possui muita experiência na área de percussão e participa de orquestra. Ele foi o usuário que mais se sentiu indiferente ao usar as interfaces, pois não acredita que instrumentos virtuais possam ser comparados e ter as mesmas qualidades encontradas nos instrumentos convencionais. Apresentou bastantes queixas, mas achou interessante para o uso na iniciação musical, como instrumento didático. Comentou que quando acompanhou o frevo no agogô, tentou fazer semicolcheias em 140bpm, mas o *Wii remote* não conseguiu acompanhar a articulação. Segundo ele, no instrumento acústico, quando é necessário fazer ritmos mais rápidos, os gestos são mais curtos e na interface, gestos curtos não são bem captados, sendo necessário usar movimentos mais longos, gerando atrasos no ritmo. Em seu ponto de vista, a vantagem em usa-las é impressionar o público com a tecnologia, não tecnicamente, nem musicalmente, mas visualmente.

Usuário 3 - Foi um músico de perfil popular. Estudou em conservatório de música e estuda percussão erudita e popular. Trabalha profissionalmente como professor de música e participa de grupos como: *bigband*, orquestra e grupos instrumentais. Nunca teve experiência com instrumentos virtuais, mas possui conhecimento em tecnologia computacional e gosta de novidades. Teve dificuldade em aplicar algumas técnicas, porque de acordo com sua descrição, no instrumento acústico ele percebe o corpo e calcula o golpe, já no virtual é diferente. Percebeu atrasos e apontou como um dos motivos, a falta de visualização do corpo como *feedback* visual,

pois algumas vezes levou a mão para baixo e esperava o som do golpe na pele, mas vinha um pouco depois e outras vezes o som vinha quando estava voltando a mão. O usuário falou que as interfaces apresentam dificuldades de precisão quando são usadas subdivisões, por isso que funcionam melhor em andamentos lentos. O instrumento que ele mais sentiu falta de variação de dinâmica foi a caixa, por ser um instrumento que usa muito esse tipo de técnica, para fazer rudimentos e nuances. Usaria em apresentações por ser uma novidade e pelo aspecto futurista, mas o usuário ainda não se sentiu confiante diante das interfaces. Diz-se defensor dos instrumentos acústicos, mas não é contra o digital, desde que seja interessante e vá somar ao seu set.

Usuário 4 - Foi um músico de perfil popular. Estudou no conservatório e atualmente estuda Licenciatura em música na universidade. Participa de grupos populares e apesar de ter conhecimento informatizado e simpatia por tecnologia, não possui experiência em instrumentos virtuais. O usuário reclamou os atrasos e a falta de controle da dinâmica, mas se sentiu bem com a sensação de liberdade e a possibilidade da troca de timbres. Gostou muito da ideia de colocar diversos timbres sem ter a presença física dos instrumentos e vê como bastante importante para bateristas e percussionistas. O usuário sentiu interesse em possuir as interfaces e usaria em apresentações em público, pois achou que a sensação de liberdade que a interface traz ao músico é fantástica.

7.4.4. Examinando o Contexto

Conforme o contexto visualizado, todos os usuários sentiram bastante dificuldade com relação à **latência**, além da **sensibilidade** funcionar melhor com gestos largos, era perceptível a dificuldade relacionada a **controle de fluxo**, que acontece, quando certa quantidade de comandos é executada em série e o fluxo de dados faz atrasar o programa. Para percussionistas essa situação é intolerável, pois segundo eles, a **métrica rítmica** fica comprometida, não havendo condições de compensar o atraso, e tanto a bateria quanto a percussão são os principais responsáveis pelo andamento rítmico.

Nenhum dos usuários conseguiu sentir **variação de dinâmica** e a **variação timbrística** foi muito resumida, fazendo com que houvesse queixas sobre **acentuação** na *performance* rítmica. Percebemos que a **acentuação é responsável pelos rudimentos e nuances** nos instrumentos solistas do naipe de percussão, como a Caixa, o Agogô e a Zabumba. Sem a acentuação, a música não tem o balanço (*swing*⁴).

Mesmo com toda a atratividade existente nas interfaces gestuais, as limitações encontradas na interação, trazem para os músicos **insegurança**, criando uma espécie de **preconceito a dispositivos digitais**. Se no instrumento convencional, onde normalmente eles manipulam bem, já encontram recursos suficientes para a extração do som desejado, porque trocar por outro artefato que por si só já é problemático?

7.5. Análise de Dados

Na interação com a música é necessário que aquilo que se toca em um instrumento tenha uma resposta imediata. Dillon (2003, apud CALEGARIO, 2010) afirma que:

(...) a resposta imediata do sistema ao gesto de entrada aumenta bastante o significado que a música tem para o usuário que está

⁴ Swing – Característica encontrada na música que faz com que ela tenha uma dinâmica envolvente. Geralmente induz ao desejo da dança.

interagindo com o sistema, em outras palavras, com a resposta em “tempo real” existe uma maior atração natural e envolvimento com a música por parte do usuário.

Como proposto anteriormente, um dos resultados da análise do experimento foi realizada a partir de três critérios de avaliação:

7.5.1. Variação Timbrística

Percebemos dificuldades nas captações de nuances das interfaces *Kinect* e *Wii remote* interferindo nas variações timbrísticas vinculadas aos gestos usados no instrumento virtual. Os usuários testaram, mas acharam limitados porque o instrumento convencional demonstrava muito mais possibilidades timbrísticas. Isso acontece também porque o mapeamento de gestos, que é necessário ser feito, vai de gestos musicais usados no instrumento, até as suas localizações espaciais. Quer dizer, além de a interface nos dar condições de tratarmos como uma variável, o gesto da batida de uma baqueta na pele da caixa, deve também nos dar condições de mapearmos a posição em que o gesto acontece, se no centro da pele da caixa ou na borda ou até mesmo no aro da caixa, pois no instrumento convencional essa variação é comum de ser utilizada na prática do instrumento.

Esse tipo de problema que encontramos, diz respeito às limitações de algoritmo e de hardware. Primeiro porque os algoritmos criados não tratam esse problema e segundo porque muitos dos gestos utilizados pelos músicos nos instrumentos fazem movimentos curtos que a própria interface não consegue captar.

7.5.2 Variação de Dinâmica

Ela se refere à variação de intensidade de volumes dentro da música ou de ostinatos⁵ rítmicos, de algum instrumento, arranjo ou fraseado. Também é responsável pelas notas acentuadas e pela não linearidade em uma música. De acordo com o experimento, nenhuma das duas interfaces ofereceu variação de dinâmica. Sendo assim, os algoritmos utilizados não resolveram problemas de mapeamento de gestos fracos e fortes.

Para ter controle de dinâmica a interface tem que possuir uma boa sensibilidade, e um mapeamento de gestos que contenha algoritmos que tratem desse critério. Uma das razões pelo qual gostamos mais de músicas interpretadas por seres humanos do que por programas de computador, está nessas variações de dinâmica, que são naturais na interpretação humana.

7.5.3 Controle temporal

O tempo foi um dos requisitos que mais deixou a desejar. Sua importância para este tipo de instrumento é fundamental, porque a percussão é um dos responsáveis pelo andamento da música. Usando o *Kinect* a latência foi bem maior do que com o *Wii remote*, pois os sensores usados por ele possuem dificuldade em captar movimentos rápidos. Apesar do instrumento de percussão ser um dos instrumentos que possuem gestos mais largos, existem detalhes como, por exemplo, rudimentos que são feitos com

⁵ Ostinato – é um motivo ou frase musical que é persistentemente repetido num mesmo padrão rítmico.
<<http://pt.wikipedia.org/wiki/Ostinato>>

movimentos curtos e são muito importantes para a expressão musical. Quando foi usado o ganzá também ocorriam sons com latência, porém o tipo de latência que apareceu era um problema de *jitter*⁶, conhecido também como problema de fluxo de dados, acontecia quando o usuário utilizava os gestos do ganzá de maneira mais rápida, fazendo com que as informações ocorressem com maior intensidade.

8. Conclusão e Trabalhos Futuros

Esta pesquisa nos deu condições de visualizar pontos relevantes a respeito de o porquê dos músicos usarem tão pouco as novas tecnologias digitais para a execução musical. Além dos problemas técnicos detectados, ainda existem outros que não foram citados por não fazer parte do nosso escopo de pesquisa, como: falta de informação, preconceito com tecnologia digital, falta de recursos financeiros que possibilitem a aquisição de equipamentos, problemas relacionados à cultura e outros. Apesar das interfaces se proporem a uma determinada interação, muitos dos gestos utilizados pelos instrumentos não são captados com facilidade em tempo real. A princípio podemos dizer que há dois pontos-chaves nessa questão. O primeiro é que existem limitações nas interfaces, que comprometem uma série de atividades na interação musical; e o segundo é que os desenvolvedores precisam criar melhores algoritmos.

Um dos problemas que os algoritmos podem tratar é a questão de variação timbrística. Podemos ter excelentes amostras de sons, e inúmeros tipos de instrumentos, mas se o mapeamento não for feito, levando em consideração a localização das sonoridades, nunca vamos ter uma variação timbrística real. Não foi possível trabalhar variação de dinâmicas e a sensibilidade das interfaces ainda não está refinada. A dinâmica e a sensibilidade estão relacionadas, porém a dinâmica deve ser tratada com algoritmos que percebam diferenças de velocidade, pois um movimento mais rápido deve proporcionar uma dinâmica mais forte e vice-versa, em quanto a questão da sensibilidade deve ser tratada no hardware, porque é ele quem capta os gestos, sejam eles longos ou curtos. No momento em que as interfaces forem mais sensíveis a pequenos gestos e os algoritmos melhorados, a dinâmica irá funcionar perfeitamente.

Para o músico, independente dos recursos que o instrumento venha oferecer ou forma física, gesto novos ou antigos, visual, *expertise* necessário, maior liberdade, novas técnicas, múltiplos timbres, etc... a coisa mais importante a ser oferecida é uma interação que promova a sensação de total controle da interface através de gestos naturais.

Nessa área é interessante pesquisarmos:

- *Feedback* tátil na ergonomia
- Idiomatismo dos instrumentos como referência para mapeamento de gestos
- Interface gestual como ferramenta didática
- Mudança de técnicas instrumentais no uso de interfaces gestuais (novas formas de controle)

O futuro da nossa pesquisa é refazer o experimento com novos questionários, agora incorporando novos gestos e comparando com outra interface. Por exemplo: O *PlayStation Move*, *Leap Motion*.

⁶ Jitter é uma variação estatística do atraso na entrega de dados. <<http://pt.wikipedia.org/wiki/Jitter>> Acesso em: 06 Ago. 2013.

9. Referências

ARAÚJO, C. R. *Uma Proposta de Ferramenta de Apoio à Educação Musical Via Web Usando Java e XML*. 2002. 118 p. Dissertação (Mestrado) – Departamento de Comunicações, Faculdade de Engenharia Elétrica e Computação, Campinas, 2002.

BARBOSA, J.; CALEGARIO, F.; MAGALHÃES, F.; CABRAL, G.; TEICHRIB, V.; RAMALHO, G. Towards an evaluation methodology for digital musical instruments considering performer's view: a case study. In PROCEEDINGS OF THE 13TH BRAZILIAN SYMPOSIUM ON COMPUTER MUSIC (SBCM'11), Vitória, ES. Brazil, 2011.

BOULEZ, Pierre (1977) "Technology and the Composer", *Language of Electroacoustic Music*. Ed. Simon Emmerson, London: Macmillan Press, 1986, pg. 5-14.

CANDÉ, Roland de. *História universal da música v.1*. 2 ed. São Paulo: Martins Fontes, 2001. ISBN 8533615000

CALEGARIO, F. Estudo de geração de música baseada em gestos do usuário. Trabalho de Graduação. Centro de Informática – UFPE, 2010.

CORREA, A.G.D., NETO, O.B., e LOPES, R.D. MusandScene: uma Interface Gestual de Expressão e Criação Musical no Auxílio ao Aprendizado de Música em Educação Especial. In: XXV Congresso da Sociedade Brasileira de Computação, 25, 2005, Rio Grande do Sul. *Anais...Canelas e São Leopoldo*, 2005.

JORDÀ, S.. Digital Lutherie. Tese de Doutorado. Universitat Pompeu Fabra, 2005.

IAZZETTA, F. A Música, o Corpo e as Máquinas. *Opus: Revista da Associação Nacional de Pesquisa e Pós-Graduação em Música*, IV(4), 27-44, 1997.

MIRANDA, Eduardo Reck; WANDERLEY, Marcelo. *New Digital Instruments: control and interaction beyond keyboard*. Middletown: A-R Editions, 2006.

WANDERLEY, M. M. (2006). Instrumentos Musicais Digitais: gestos, sensores e interfaces. In: Beatriz Ilari (ed.) *Em Busca da Mente Musical*. Curitiba: Editora da Universidade Federal do Paraná. <http://idmil.org/_media/wiki/instrumentos_digitais_final.pdf>

WANDERLEY, M. M. Projeto e utilização de instrumentos musicais digitais. In: KELLER, D. e BUDASZ, R. *Criação musical e tecnologias: teoria e prática interdisciplinar*. Goiânia: ANPPOM, 2010. P. 70-96.

GTSpat: um sistema interativo de síntese e espacialização sonora

Gabriel Rimoldi^{1,2}, Adolfo Maia Jr.²

¹Instituto de Artes - Departamento de Música | ²Núcleo Interdisciplinar de Comunicação Sonora (NICS) – Universidade Estadual de Campinas (UNICAMP)
Campinas – SP - Brasil

{gabriel.rimoldi,adolfo}@nics.unicamp.br

***Abstract.** This article reports the development of a real-time system of sound synthesis and spatialization of granular flows, named GTSpat. Firstly, we present the general architecture of the algorithm, and then, point out details of the interactive strategies applied to the system. Finally, we present some adaptive schemes by which the system can be modified by interaction with the user.*

***Resumo.** Este artigo aborda o desenvolvimento de um sistema de síntese e espacialização sonora de fluxos granulares em tempo real, ao qual denominamos GTSpat. Apresentamos inicialmente a arquitetura geral do algoritmo e, em seguida, detalhamos as estratégias interativas aplicadas ao sistema. Por fim, apresentamos alguns esquemas adaptativos pelos quais o sistema pode ser modificado pela interação com o usuário.*

1. Introdução

O grande desenvolvimento tecnologias voltadas para a difusão multicanal e sua crescente acessibilidade tem, nas últimas décadas, renovado o interesse de compositores e pesquisadores sobre métodos formais para o controle da espacialização sonora [Otondo, 2008]. Neste trabalho, reportamos um método desenvolvido para o controle da distribuição espacial dos sons através das chamadas *Medidas de Similaridade* relacionadas ao timbre, no qual associamos um conjunto de Descritores de Áudio [Puckette e Apel 1998; Foote, 1997] obtidos a partir de uma determinada amostra sonora à sua respectiva localização espacial na difusão sonora com suporte eletrônico digital. Mais especificamente, apresentamos uma solução aplicada à distribuição espacial, em tempo real, de fluxos sonoros gerados por Síntese Granular.

O controle espacial aplicado a sons granulares é um procedimento bastante recorrente na literatura. Desde o surgimento da técnica granular, a atribuição de uma localização aural específica a cada partícula ou grão sonoro em sistema multicanal tem se mostrado como um procedimento bastante idiomático e enriquecedor das paisagens sonoras geradas [Roads, 2001]. Dada a enorme quantidade de grãos usualmente necessários para a formação de um determinado segmento sonoro, é comum à síntese granular o emprego de controle macroestrutural, que determine, dentre outros parâmetros, a localização específica para cada grão no espaço de projeção sonora.

Podemos observar o uso frequente de métodos matemáticos na modelagem da distribuição espacial de síntese granular como, por exemplo, através de distribuições aleatórias, ou estocásticas, dos grãos. Estes métodos, no entanto, apresentam, na maioria das vezes, pouca ou nenhuma relação com os outros aspectos sonoros e/ou musicais em contexto. Trabalhos recentes têm adotado outras estratégias de auralização dos grãos sonoros. Kim-Boyle (2006), Wilson (2008) e Barreiro (2011) utilizaram o algoritmo *boids* para o controle de espacialização do fluxo granular em sistemas multicanal. Souza e Maia (2009) desenvolveram um modelo baseado em Algoritmos Genéticos (G.A.) para o controle de diversos parâmetros da síntese granular, dentre eles, a posição panorâmica de cada grão no sistema estereofônico. Em trabalhos atuais sobre Síntese Concatenativa, observamos o emprego de métodos que associam a posição aural de cada segmento sonoro aos Descritores de Áudio extraídos do mesmo. No sistema *CataRT*, do IRCAM, por exemplo, os *corpus sonoros* são distribuídos em um espaço bi ou tridimensional¹ em função da classificação dos descritores escolhidos pelo usuário [Einbond & Schwarz, 2010].

Em direção próxima aos trabalhos supracitados, nossa pesquisa tratou do desenvolvimento de ferramentas composicionais que tratassem a espacialização sonora como um processo conjugado, ou par a par, com outros aspectos da criação musical. O sistema *GTSpaT* foi então concebido como um modelo de síntese e espacialização do fluxo granular, no qual se estabelece uma função entre qualidades timbrísticas e a posição espacial de cada grão sonoro na difusão em suporte eletroacústico. O sistema foi implementado em *Pure Data* (PD) e utiliza a biblioteca *iem_ambi* desenvolvida por Thomas Musil e colaboradores (2005) para a espacialização sonora com *Ambisonics* [Gerzon, 1973] e algoritmos adaptados da biblioteca *PDescriptors*, desenvolvida por Monteiro (2010), para a extração de características do sinal de áudio.

2. O sistema GTSpaT

A arquitetura do sistema foi concebida em três módulos distintos, sendo o primeiro um módulo de análise e classificação dos segmentos sonoros, o segundo de síntese e espacialização e, por último, o módulo interativo (cf. Figura 1). O usuário seleciona inicialmente uma amostra sonora e define os parâmetros de análise da mesma. No módulo de análise, a amostra de áudio é então seccionada em pequenos quadros (ou grãos), que são analisados segundo um conjunto de descritores acústicos. Os dados obtidos dos descritores para cada grão são alocados na forma de um vetor. A posição espacial de cada grão é então uma função deste vetor denominado *vetor de características*. No módulo de interação o usuário define quais grãos serão enviados ao módulo de síntese e espacialização. Os grãos sonoros selecionados, juntamente com sua posição espacial, são endereçados ao módulo de síntese e espacialização, responsável pela codificação do sinal monofônico em relação à sua posição espacial e, em seguida, pela decodificação dos sinais resultantes para a quantidade de saídas de áudio do sistema.

¹ Neste, a terceira dimensão é representada graficamente através de uma escala de cores.

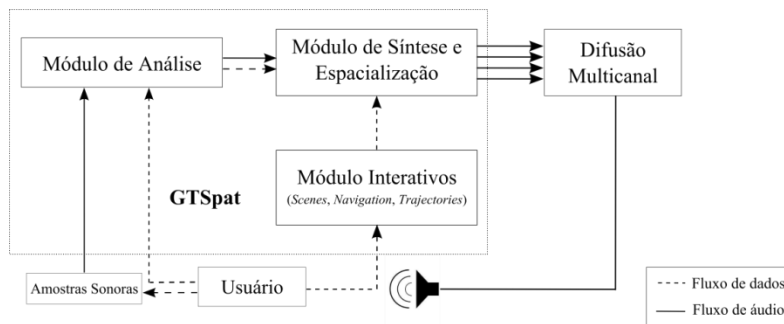


Figura 1: Arquitetura geral do Sistema GTSpac

Da amostra digitalizada selecionada pelo usuário são tomadas sucessivas seqüências de pontos, de tamanho invariável e com uma taxa de sobreposição (interseção das sequências) entre as mesmas. Para cada quadro ou seqüência de pontos, ponderados por uma envoltória de amplitude, atribui-se então um *vetor de características*, resultante de um conjunto de medidas extraídas do sinal, as quais serão denominadas por descritores acústicos. Estes, como o próprio nome indica, representam uma maneira econômica de substituir o objeto por algo, geralmente, muito mais simples, e que é suscetível de uma análise que reflete potenciais propriedades do objeto. Sua utilização torna-se bastante útil, principalmente para sistemas sonoros digitais em tempo real, nos quais processamentos de um sinal podem alcançar tamanhos proibitivos. Optamos então pelo uso de descritores que, embora tenham caráter reducionista em relação ao objeto sonoro analisado, podem ser muito úteis como uma taxonomia para uso imediato em espacialização, por exemplo. No nosso caso, tais descritores permitem o delineamento de comportamentos presentes no sinal através de ponderações estatísticas elementares, tanto no domínio do tempo quanto das frequências, e que possuem correlações diretas com a percepção sonora.

Em nosso modelo, o vetor de características é composto por duas categorias de descritores acústicos: *temporais* e *espectrais*. A primeira consiste de medidas computadas a partir do sinal representado no domínio do tempo, enquanto que a segunda é composta de medidas associadas a informações computadas da representação do sinal no domínio das frequências, através da Transformada Discreta de Fourier. Os descritores utilizados no sistema GTSpac foram: *energia média*, *taxa de cruzamento por zero*, *centroide espectral*, *desvio médio espectral*, *obliquidade espectral*, *curtose espectral*.

A definição do tamanho do quadro de análise e a taxa de sobreposição são previamente definidas pelo usuário. A quantidade de pontos analisados a cada quadro pode variar de 441 a 8820² e a taxa de sobreposição entre quadros pode variar entre 0 a 75 por cento do tamanho da janela temporal de análise. O usuário pode também definir qual tipo de envoltória de amplitude aplicar aos quadros de análise, sendo elas:

² Para os tamanhos de janela não correspondentes a valores de potências de dois, são adicionados zeros até que se complete a quantidade de pontos equivalentes à potência de dois superior mais próxima. Este procedimento é conhecido na literatura como *Zero Padding*.

senoidal, gaussiana, exponencial inversa (*rexpodec*), exponencial (*expodec*), retangular e ADSR³.

Definidos tais parâmetros, cada quadro da amostra sonora é analisado extraindo-se então os dados de interesse, ou seja, os valores numéricos dos descritores do sistema o quais são armazenados em um vetor numérico. A localização espacial de cada grão é dependente das entradas de seu respectivo vetor, as quais são convenientemente normalizadas e ponderadas por pesos definidos pelo usuário.

Matematicamente, em nosso modelo, denotamos por n a quantidade de descritores do sistema e m a quantidade de grãos sonoros obtidos a partir de uma amostra de áudio. Seja d_{kj} o valor do descritor acústico k atribuído ao grão sonoro j , com $k = 1, 2, 3, \dots, n$ e $j = 1, 2, 3, \dots, m$ e ainda p_k, q_k e r_k pesos atribuídos ao descritor k , com $0 \leq p_k, q_k, r_k \leq 1$ e normalizados como $\sum_{k=1}^n p_k = \sum_{k=1}^n q_k = \sum_{k=1}^n r_k = 1$. Definimos a posição espacial (x, y, z) ⁴ para cada grão j como:

$$x_j = \sum_{k=1}^n p_k d_{kj} \quad y_j = \sum_{k=1}^n q_k d_{kj} \quad z_j = \sum_{k=1}^n r_k d_{kj}$$

O módulo de síntese e espacialização é composto por três processos distintos: *granulação, codificação e decodificação espacial*. O primeiro deles é responsável pela seleção do fragmento do áudio amostrado (o grão) e a multiplicação do sinal pela envoltória de amplitude pré-estipulada. Na codificação, o sinal monofônico é processado em relação à distância (r) e direção espacial com ângulos de coordenadas polares (θ, ϑ) . Para a codificação de distância, utilizamos filtros do tipo IIR (*Infinite Impulse Response*) para simular a atenuação da energia do espectro de frequências em decorrência da distância de cada fonte em relação ao ouvinte [Daniel, 2003].

Para a codificação direcional, utilizamos um sistema Ambisonics de 2ª ordem. O codificador é responsável pela renderização do sinal monofônico enviado pelo granulador a uma quantidade de canais que simulam o posicionamento espacial virtual do mesmo e que, por sua vez, são enviados ao decodificador. Na codificação, o sinal de cada um desses canais é multiplicado por uma matriz de ganho que determina o quanto cada um destes será enviado para cada canal da difusão sonora. A matriz de ganho varia em relação à configuração de alto-falantes estipulada ao sistema. A Figura 2 ilustra o processo de síntese e espacialização do sistema GTSpac.

3. Estratégias interativas

Através dos módulos de interação, o usuário pode controlar o sequenciamento do fluxo granular que será enviado ao módulo de síntese e espacialização. A fim de explorar

³ Ressaltamos que o uso de janelas com as envoltórias de amplitude acima não são exclusivas. Muitas outras podem ser definidas. Claramente o conteúdo espectral do sinal sonoro é bastante afetado pelos diferentes tipos de amplitude. Por exemplo, ADSR é nada mais que uma aproximação linear por partes da maioria dos sinais sonoros que, de alguma forma, possuem ataque (*attack*), decaimento (*decay*), sustentação (*sustain*) e relaxamento (*release*). Sua simplicidade e plasticidade é a razão de que é muito utilizada em sintetizadores.

⁴ O modelo admite distribuições espaciais bi- ou tridimensionais. Para o primeiro caso, adota-se $z_k = 0$, ou seja, a elevação da fonte sonora na espacialização é considerada constante.

distintas possibilidades de controle da espacialização, construímos diferentes propostas interativas das quais resultam os três módulos de interação: *GTSpat Scenes*, *GTSpat Navigation* e *GTSpat Trajectories*. Cada um destes módulos explora determinadas propriedades de síntese e espacialização sonora. Nos dois primeiros, procuramos explorar a localização espacial dos grãos como elemento de interação. O primeiro baseia-se na metáfora de múltiplas fontes sonoras fixas no espaço, pelo qual o usuário detém o controle paramétrico de fluxos granulares distribuídos em distintas regiões da espacialização. O segundo, por outro lado, baseia-se na metáfora de uma única fonte sonora que se desloca no espaço através do controle gestual do usuário. No terceiro módulo, procuramos salientar as propriedades cinéticas na interação entre o sistema e o usuário, utilizando a *diretividade* e taxa de deslocamento de movimento como parâmetros de controle da síntese e espacialização. Nos tópicos a seguir descreveremos com maior detalhe cada um desses módulos interativos.

3.1. GTSpat Scenes

Neste módulo, o usuário detém o controle de diversos fluxos granulares espalhados através do espaço de difusão sonora. Tendo distribuído os grãos no espaço bi ou tridimensional de controle da espacialização através de seu vetor de características, o usuário pode então determinar sub-regiões deste espaço as quais definirão cada um dos objetos da cena espacial. Estas sub-regiões são representadas na interface por retângulos (cf. Figura 3), que podem ser manipulados pelo usuário (posição e área) a fim de se definir qual grupo de grãos formará o referido *objeto na cena espacial*. O sistema comporta até quatro objetos sonoros simultâneos, cada qual composto por até dez fluxos granulares síncronos (ou quase síncronos). A quantidade de fluxos e intervalo de tempo entre os grãos em cada um dos objetos são definidos pelo usuário, podendo este último variar entre 1 e 10 vezes a duração do grão nesta versão do GTSpat. Cada objeto possui ainda um controle individual de amplitude, que permite que o usuário determine diferentes níveis de intensidade para cada objeto na cena espacial. Selecionada a área de um determinado objeto, o granulador de cada fluxo deste objeto pode sortear de maneira aleatória qualquer grão que esteja contido na área associada a este objeto.

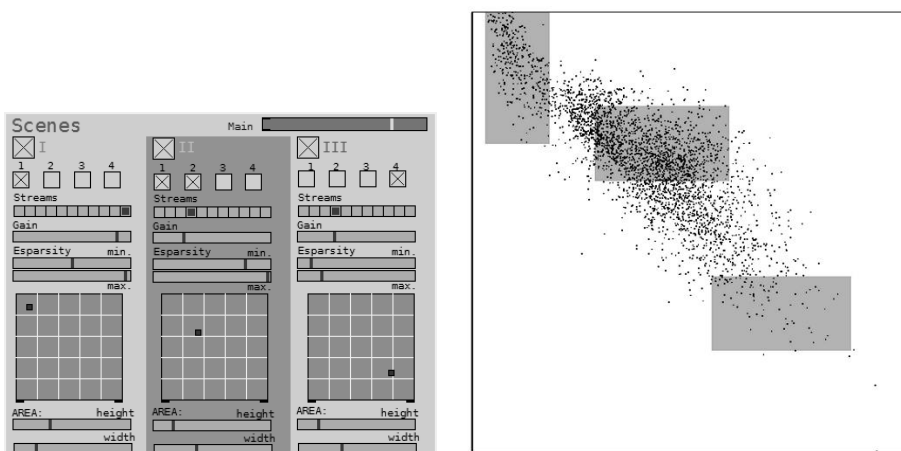


Figura 3: Interface de controle e visualização do GTSpat Scenes

3.2. GTSpat Navigation

Neste segundo módulo interativo, o sequenciamento temporal dos grãos sonoros como também a sua espacialização são controlados pelo usuário através de um “navegador bidimensional”. Este consiste de um quadrado virtual com tamanho definido pelo usuário e posição indicada através de uma interface de controle. A área demarcada pelo navegador no espaço de controle da espacialização, ao qual denominaremos Espaço de Navegação, define a região alvo do sistema. A partir da distribuição dos grãos no Espaço de Navegação, pela ponderação de seus vetores de características, o sistema envia ao módulo de síntese e espacialização os grãos circunscritos na área demarcada pelo navegador.

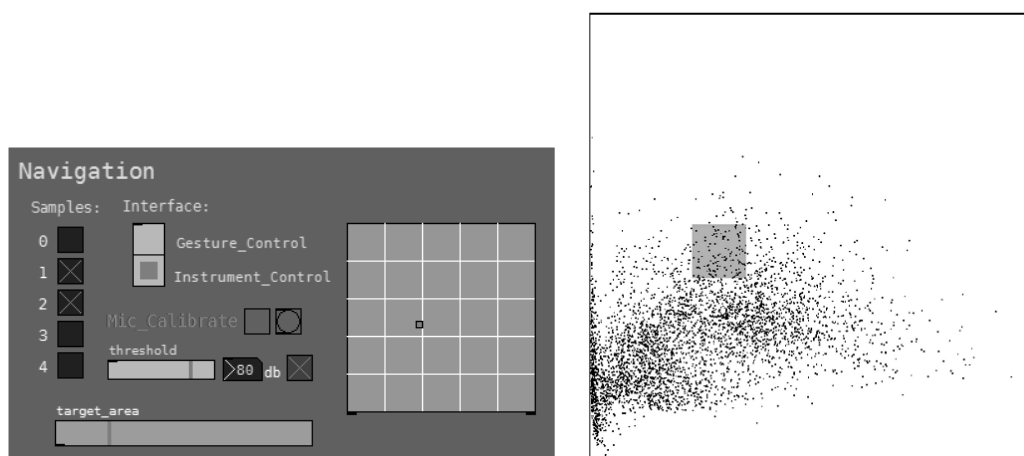


Figura 4: Interface de controle e visualização do GTSpat Navigation

Abordamos dois diferentes tipos de interface de controle. O primeiro consiste em interfaces gestuais capazes de reportar ao sistema informações de posicionamento espacial (*mouse* ou controladores gestuais de jogos, por exemplo). O segundo tipo consiste num microfone, pelo qual o sistema recebe informações de áudio que serão processadas e analisadas pelo próprio sistema e, em seguida, transformadas em informação numérica referente à posição do navegador.

O processamento aplicado à entrada de áudio da interface de controle é similar ao aplicado às amostras sonoras no Módulo de Análise. O áudio captado pelo microfone é analisado em tempo real⁵ pelo mesmo conjunto de descritores do Módulo de análise. O vetor de características extraído em tempo real é ponderado pelos mesmos pesos atribuídos pelo usuário para cada eixo da espacialização e em seguida atribuído à posição bidimensional do navegador no Espaço de Navegação. Os dados absolutos obtidos da análise dos descritores são normalizados através de um procedimento de calibração do sistema.

⁵ A latência do processamento em tempo real é determinada pelo tamanho da janela de análise estipulada pelo usuário para cada amostra analisada previamente, ou seja, pode variar entre 10 a 200 milissegundos.

3.3. GTSpat Trajectories

Com a criação deste módulo, procuramos estabelecer outras possibilidades de interação do sistema, utilizando medidas de segunda ordem das posições registradas pela navegação do usuário. As coordenadas cartesianas do Navegador, ao serem manipuladas pelo usuário no Espaço de Navegação, são registradas no sistema como informações de trajetória, em magnitude (tamanho do deslocamento) e ângulo. A partir desta trajetória, à qual denominaremos por “alvo”, o sistema compara as possíveis trajetórias dentre o conjunto de posições cartesianas atribuídas a cada grão no Espaço de Navegação. O sistema elege então a trajetória, dentre todas possibilidades de deslocamento condicionado pela distribuição dos grãos, que mais se “assemelhe da trajetória alvo” segundo uma conveniente medida, ou função, de comparação. A cada “passo” da trajetória no Espaço de Navegação o sistema compara as possibilidades de deslocamento dos grãos distribuídos no Espaço de Navegação e elege aquele que resultará numa trajetória “próxima” daquela estipulada pelo usuário como “alvo”.

Para a definição das trajetórias resultantes do sequenciamento dos grãos, utilizamos um processo markoviano de decisão, pelo qual o sistema elege o grão subsequente do fluxo em função da posição espacial do grão atual. Dada a coordenada inicial (x_0, y_0) como sendo a posição espacial de um grão emitido pelo sistema. A partir de uma trajetória “alvo” \vec{T} , definida pelo usuário e com magnitude $\|\vec{T}\|$, e ângulo de deslocamento σ_T , o sistema calcula as magnitudes e ângulos de todas as possibilidades de trajetórias em relação ao seu ponto inicial (x_0, y_0) . Dentre as possíveis trajetórias, condicionadas pela quantidade e distribuição dos grãos sonoros no Espaço de Navegação, o sistema escolhe a que mais se assemelhe à trajetória “alvo”. No decorrer de sucessivos “passos” do trajeto, o vetor “alvo” pode manter-se fixo ou também alterar-se a cada novo estado.

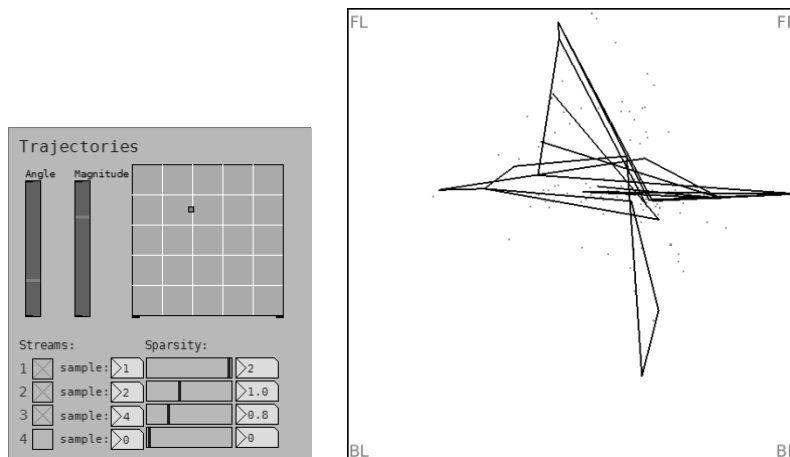


Figura 5: Interface de controle e visualização do GTSpat Trajectories

Em nosso modelo, o sistema sorteia aleatoriamente um grão sonoro inicial de posição definida por (x_0, y_0) para cada fluxo granular (*streams*). Definimos $\vec{U}_j = (x_j - x_0, y_j - y_0)$ como o vetor formado entre a posição (x_0, y_0) do grão inicial emitido e a

posição (x_j, y_j) de cada grão j ⁶, com magnitude $\|\vec{U}_j\|$ e ângulo σ_j . Definimos ainda a trajetória “alvo” do sistema pelo vetor $\vec{T}_n = (x_n - x_{n-1}, y_n - y_{n-1})$ formado pelas diferenças sucessivas⁷ de posição (x_t, y_t) do navegador, com magnitude $\|\vec{T}\|$ e ângulo σ_T . A comparação entre a trajetória alvo e as possíveis trajetórias do sistema é dada pela função:

$$l_{jt} = h \left(1 - \frac{\|\vec{U}_j\| - \|\vec{T}\|}{\sqrt{2}} \right) + \tilde{h} \left(\frac{\sigma_j - \sigma_t}{\pi} \right),$$

onde h e \tilde{h} são pesos que podem ser definidos pelo próprio usuário, a fim de se atribuir maior ênfase ao tamanho ou à direção do trajeto “alvo”.

Nesta fórmula, o sistema calcula dentre as trajetórias possíveis a partir do ponto (x_0, y_0) , com índices $\|\vec{U}_j\|$ e σ_j , a que mais se aproxima da trajetória “alvo” de magnitude $\|\vec{T}\|$ e ângulo σ_T indicada pelo usuário. O sistema seleciona o grão j que obteve maior índice l_{jt} em relação à trajetória alvo e o envia ao granulador. Uma nova medida é então calculada, sucessivamente, a partir da posição espacial do grão selecionado, comparando as possíveis trajetórias em relação à trajetória “alvo”.

4. Estratégias de adaptação

Associados aos módulos de interação do sistema GTSpaT, construímos dois submódulos adaptativos. O propósito da criação destes submódulos foi de estabelecer novas instâncias de interação entre o sistema computacional e o usuário. Para isto, procuramos estabelecer estratégias nas quais o sistema se adapta (ou se auto-regula) em busca de um estado ótimo (distribuição dos grãos) que mais se aproxime das trajetórias do navegador no Espaço de Navegação. Em ambos os submódulos o sistema procura redistribuir os grãos no Espaço de Navegação a partir desses dados cinemáticos extraídos da navegação do usuário.

4.1. Controle adaptativo do vetor de características

No primeiro submódulo, os pesos dos descritores associados a cada um dos eixos da espacialização são redistribuídos de maneira automática, por uma função distância entre a distribuição do conjunto de dados de cada descritor e a distribuição dos pontos das trajetórias do navegador. Adotamos três aspectos como critérios para análise das trajetórias percorridas pelo navegador, sendo eles o ponto médio dentre as posições percorridas em cada eixo do Espaço de Navegação, o grau de dispersão destes pontos em relação ao ponto médio para cada eixo e, por último, o grau de assimetria da distribuição destes pontos em relação ao ponto médio. As mesmas medidas extraídas da navegação do usuário são também realizadas para o conjunto de dados de cada descritor.

⁶ Com $j = \{1, 2, 3 \dots m\}$, referente à quantidade de grãos extraídos da amostra sonora.

⁷ Definimos a taxa de amostragem pela duração temporal equivalente ao *hop-size* aplicado à amostra sonora no Módulo de Análise, conforme descrito na arquitetura geral do sistema.

5. Considerações Finais

Apresentamos neste artigo uma metodologia para controle de espacialização de sons granulares em tempo real. Descrevemos a estrutura algorítmica do sistema GTSpac e, mais detalhadamente, abordamos estratégias de interação e adaptação abordadas. De modo geral, a implementação computacional do modelo comportou-se de maneira bastante estável e apresentou resultados sonoros satisfatórios na distribuição dos grãos pela difusão sonora. Acreditamos que as estratégias interativas por nós adotadas demonstraram-se também bastante eficazes e sintetizam um rico conjunto de potencialidades aplicadas à criação musical. Cada módulo interativo desenvolvido explora diferentes características da espacialização, que são manipuladas através de informações recebidas do usuário. Nos dois primeiros, *Scenes* e *Navigation*, delineamos processos reativos de processamento, enquanto que no terceiro módulo, *Trajectories*, utilizamos mecanismos computacionais de tomada de decisão em função de dados de segunda ordem das informações recebidas. Através dos módulos adaptativos, estabelecemos ainda estratégias de organização interativa do sistema, que pode alterar-se mediante dados de segunda ordem das informações recebidas do usuário.

6. Projeções Futuras

Em sequência ao trabalho aqui apresentado, pretendemos aprimorar o algoritmo de espacialização utilizado, sobretudo no que tange à codificação de informação de distância e na simulação de ambientes acústicos. Pelo custo computacional que isto implicava, optamos nesta primeira versão do sistema restringir a espacialização apenas à simulação de posicionamento das fontes sonoras. A codificação de informações relacionadas às primeiras reflexões e reverberação de fontes sonoras virtuais é elemento bastante rico à espacialização e muito contribui para o grau de realismo da cena acústica criada. Posteriormente, pretendemos utilizar codificadores/decodificadores Ambisonics de ordem superiores, bem como aplicar técnicas híbridas de espacialização, que garantam maior definição da auralização de fontes sonoras virtuais. Com a implementação de sistemas de espacialização mais sofisticados, poderemos ainda explorar interativamente diversos outros atributos espaciais, dentre os quais, destacamos a manipulação em tempo real das propriedades acústicas da sala e mudanças de perspectivas da auralização pela utilização de *sweet spot* dinâmico.

Consideramos ainda que a metodologia de espacialização empregada pode ser ampliada a outros contextos que não propriamente o da síntese granular ou da síntese concatenativa. Procedimentos similares podem, por exemplo, ser utilizados na distribuição espacial de sons de longa duração, que se deslocariam no espaço da difusão em função das medidas extraídas em nível micro-temporal. Em continuidade à nossa pesquisa, pretendemos estender o modelo elaborado pela utilização de diferentes interfaces que, sobretudo, estabeleça um diálogo mais direto entre o espaço físico da difusão e o espaço virtual das fontes sonoras.

7. Agradecimentos

Agradecemos às agências de fomento FAPESP e CAPES pelo financiamento desta pesquisa.

8. Referências

- Barreiro, D. L. (2011) “Manipulação de amostras sonoras em contexto interativo”. *Anais do Encontro Internacional de Música e Arte Sonora (EIMAS)*, Juiz de Fora, Brasil.
- Daniel, J. (2003). “Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format”. *Proceedings of AES 23rd International Conference*, 23-25.
- Einbond, A; Schwarz, D. (2010) “Spatializing Timbre with Corpus-Based Concatenative Synthesis”. International Computer Music Conference (ICMC) June 1–5, 2010, New York. *Proceedings of the International Computer Music Conference*, NY.
- Foot, J. (1997) “Content-based Retrieval of Music and Audio. Multimedia Storage and Archiving Systems II”. *Proceedings of SPIE*.
- Gerzon, M. A. (1973) “Periphony: With-Height Sound Reproduction”. In: *Journal of Audio Eng. Soc.*, v. 21, p. 2-10, 1973.
- Kim-Boyle, D. (2006) “Spectral and Granular Spatialization with Boids”. *Proceedings of the 2006 International Computer Music Conference*. New Orleans, p.139-142.
- Monteiro, A. (2010) *PDdescriptors*, versão 1.2. Campinas, Brasil.
- Otondo, F. (2008). “Contemporary trends in the use of space in electroacoustic music”. *Organised Sound*, 13 (1), 77-81.
- Musil, T.; Noisternig, M.; Hoeldrich, R. (2005) "A Library for Realtime 3D Binaural Sound Reproduction in Pure Data (PD)", *Proc. Int. Conf. on Digital Audio Effects (DAFX-05)*, Madrid, Spain, September 20-22.
- Puckette, M., Apel, T. (1998) “Real-time audio analysis tools for Pd and MSP”, *Proceedings of International Computer Music Conference*. San Francisco: International Computer Music Association, pp. 109-112.
- Roads, C. (2001) *Microsound*. Cambridge: MIT Press, 2001.
- Schumacher, M.; Bresson, J. (2010). "Spatial Sound Synthesis in Computer -Aided Composition." *Organised Sound*, 15(3), p.271-289.
- Sousa, F. F.; Maia, A. (2009) “Evolutive Processes for Granular Synthesis in Extended Gabor Spaces”, 7º Congresso de Engenharia de Áudio da AES Brasil, 2009, São Paulo. *Anais do 7o Congresso de Engenharia de Áudio da AES Brasil*. São Paulo
- Wilson, S. (2008) “Spatial Swarm Granulation”, *Proceedings of the 2008 International Computer Music Conference*. Belfast: SARC, ICMA

Design¹ oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade

Damián Keller¹, Floriano Pinheiro da Silva¹, Edemilson Ferreira da Silva¹, Victor Lazzarini², Marcelo Soares Pimenta³

¹Núcleo Amazônico de Pesquisa Musical (NAP), Universidade Federal do Acre, Rio Branco, AC, Brasil

²Department of Music, National University of Ireland, Maynooth

³LCM, Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Grupo de Música Ubíqua

dkeller@ccrma.stanford.edu, victor.lazzarini@nuim.ie,
mpimenta@inf.ufrgs.br

Resumo. *Relatamos os resultados de um estudo experimental em música ubíqua envolvendo 12 sujeitos. O estudo focalizou o impacto dos fatores ambientais em diversos tipos de atividades musicais criativas. Duas condições foram consideradas: atividade em ambiente doméstico e atividade em ambiente comercial. Formulamos duas hipóteses de trabalho: 1. os ambientes com sons previsíveis reduzem o incentivo à criatividade enquanto os ambientes com sons pouco previsíveis aumentam o desempenho criativo; 2. a similaridade entre o tipo de sons ambientais e o tipo de sons utilizados na atividade criativa aumenta o desempenho criativo, já o contraste e a imprevisibilidade dos sons do ambiente reduz o desempenho. Os sujeitos realizaram 108 ensaios de interação utilizando a ferramenta musical ubíqua mixDroid 1.0 e avaliaram 2 quesitos relativos ao produto - relevância e originalidade - e 4 itens relativos à atividade - esforço cognitivo, engajamento, diversão e colaboração. Os resultados mostram que ambientes sonoros imprevisíveis fomentam a originalidade mas têm um impacto negativo no desempenho criativo - reduzindo o engajamento, a diversão e as possibilidades de colaboração, e aumentando a demanda cognitiva da atividade. Discutimos as implicações para o design de sistemas musicais ubíquos.*

1. Suporte a atividades musicais criativas em contexto ubíquo

As atividades musicais ubíquas [Keller et al. 2011a] demandam novos enfoques de design de sistemas de suporte tecnológico [Flores et al. 2010]. Duas estratégias de desenvolvimento vêm sendo priorizadas. Uma focaliza a implementação de infraestrutura para atividades musicais [Miletto et al. 2011]. A outra tenta aproveitar os recursos materiais já existentes, integrando esses recursos ao fluxo de atividades em contextos normalmente desvinculados da atividade musical [Keller et al. 2013; Pinheiro da Silva et al. 2012]. Dentro da segunda perspectiva, Lazzarini et al. (2012) propõem o

¹ Design de software é o termo adotado pela área de Interação Humano-Computador para as atividades relacionadas ao desenvolvimento de projetos de aplicativos computacionais. A palavra design também é utilizada por outras disciplinas como design gráfico, design arquitetônico, design de interiores, e se estende à profissão designer. Portanto não utilizamos itálico para esses termos que já foram incorporados à língua portuguesa. Reservamos os termos desenho, projeto e implementação para instâncias específicas da atividade de design.

desenvolvimento de ecossistemas musicais ubíquos - ferramentas modulares que podem ser incorporadas para atender necessidades específicas do processo de design de software. Essa proposta pode ser caracterizada como 'design oportunista', e se alinha com uma longa tradição de métodos de design que atualmente vem ganhando força dentro do campo da computação ubíqua [Ball & Ormerod 1995; Cross 2001; Visser 1994]. Um aspecto do design oportunista que pode ter impacto no suporte à criatividade em contextos cotidianos é o aproveitamento dos recursos existentes no local onde é realizada a atividade criativa. Esse é o foco da pesquisa relatada neste artigo.

2. Marcação temporal e criatividade

Na perspectiva da pesquisa em interação humano-computador, metáforas são analogias ou similaridades para se representar ou modelar atividades e/ou projetos do mundo real dentro do ambiente computacional [Pimenta et al. 2012]. Podemos citar a conhecida metáfora da 'área de trabalho' que leva para o mundo virtual a ideia da mesa de trabalho, onde podem existir diversos documentos ou objetos dispersos. Fora do campo da computação, um exemplo de metáfora são os balões de conversas existentes em revistas em quadrinhos que podem representar tanto falas quanto pensamentos dos protagonistas.

A metáfora visual de representação em trilhas utilizada pela maioria dos editores para dispositivos estacionários não é facilmente transposta aos sistemas portáteis. Essa metáfora de interação demanda o uso intensivo de recursos computacionais para dar suporte aos elementos gráficos da interface. Ela não fornece controles intuitivos para atividades síncronas: o tempo fica desvinculado das ações do usuário. E ela não é facilmente escalável. As mixagens com pistas múltiplas demandam a redução do tamanho dos elementos da interface dificultando a visualização e o controle.

Como alternativa, Keller et al. (2010) propõem a metáfora de marcação temporal. Nesta metáfora de interação, a reprodução do som fornece as pistas perceptuais que o usuário utiliza para colocar as amostras sonoras no eixo temporal. Como o compositor se baseia nos sons sendo executados, não existe necessidade de utilizar o referencial visual. Portanto, a interface pode se limitar a um número reduzido de botões para acessar as amostras sonoras.

A marcação temporal vem mostrando bons resultados em diversas atividades criativas realizadas tanto por sujeitos leigos quanto por músicos. Pinheiro da Silva et al. (2013) relatam um experimento com seis sujeitos - leigos e músicos - abrangendo 47 mixagens com amostras de sons urbanos e de sons de animais em duas condições experimentais: dentro do estúdio e nos locais de coleta dos sons. Os autores obtiveram dados sobre o suporte às atividades criativas, incluindo mixagens livres (criação), mixagens seguindo um modelo (imitação), e uso livre da ferramenta sem o objetivo de produzir um resultado sonoro (exploração). Os resultados apontaram para efeitos relacionados ao local onde é realizada a atividade nos fatores explorabilidade e produtividade. O aumento na avaliação da produtividade estaria relacionado tanto ao local (para atividades fora do estúdio) quanto ao tipo de atividade (com leve superioridade para criação). Maior explorabilidade foi consistentemente vinculada a atividades fora do âmbito do estúdio, com preferência leve para as atividades de imitação, independentemente do tipo de amostra utilizada. Essas tendências, em conjunto, mostraram que houve um maior engajamento criativo dos sujeitos nas atividades em ambientes externos.

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

Keller e coautores (2013) realizaram um estudo com dez indivíduos, envolvendo 40 ensaios de interação utilizando a mesma ferramenta que a adotada no estudo de Pinheiro da Silva et al. (2013). Os sujeitos fizeram mixagens com amostras sonoras vocais percussivas e avaliaram a experiência criativa e o suporte da metáfora de interação para os diversos fatores de criatividade. O estudo incluiu três condições: lugar (ambientes domésticos e comerciais); tipo de atividade (imitação de mixagens e criações originais) e postura corporal (com os sujeitos em pé ou sentados). Os músicos que participaram na experiência avaliaram todos os fatores com escores próximos ao máximo da escala. As avaliações dos sujeitos leigos foram inferiores e tiveram maior variabilidade, no entanto, também foram positivas. Em relação ao gênero, o grupo das mulheres teve desempenho muito similar ao grupo dos músicos. Porém os autores frisam que as diferenças observadas entre homens e mulheres necessitam de estudos mais amplos. Os resultados mostraram um aumento da pontuação nos fatores explorabilidade e colaboração para as atividades realizadas em ambientes domésticos. Keller et al. (2013) explicam a leve superioridade nos fatores explorabilidade e colaboração pela menor demanda cognitiva para a realização de atividades nesse tipo de ambiente. O estudo indicou a existência de uma relação entre o perfil dos sujeitos, o local onde foram realizadas as atividades e os resultados das avaliações. Como a ancoragem envolve o aproveitamento das pistas cognitivas fornecidas pelo local da atividade, a relação sujeito-local-atividade é consistente com a utilização do mecanismo de ancoragem para viabilizar o suporte a atividades criativas [Keller et al. 2010; Lima et al. 2012]. Tendo em vista que esse mecanismo é aplicável a diversos tipos de metáforas de interação, os resultados são generalizáveis a diversas plataformas, a atividades diversas e a múltiplos contextos de uso.

Porém, esses estudos não analisaram em profundidade a relação entre as pistas sonoras do local e o material sonoro utilizado na atividade criativa. Esse tipo de análise só seria possível realizando experimentos com material sonoro com o qual os sujeitos já estivessem amplamente familiarizados. Estudos recentes em psicologia ambiental indicam que ambientes desordenados aumentam o desempenho criativo dos sujeitos [Vohs et al. 2013]. No entanto, a variável manipulada nesses experimentos é a disposição dos objetos materiais no ambiente. No caso das fontes sonoras cotidianas, ordem pode ser considerada como equivalente a previsibilidade e a familiaridade. Se a proposta de Vohs e coautores for válida para as variáveis sonoras vinculadas ao ambiente, a hipótese a ser testada é se ambientes com sons previsíveis reduzem o incentivo à criatividade, e se ambientes com sons pouco previsíveis aumentam o desempenho nesse quesito. Essa hipótese contrasta com a ideia de uma relação positiva entre os recursos fornecidos pelo ambiente e o desempenho criativo. Pinheiro da Silva et al. (2012) sugerem um mecanismo de atenção que aproveita os elementos existentes no ambiente (pistas cognitivas) como forma de sustentação do processo criativo durante a mixagem: a atenção criativa. Segundo essa proposta, ambientes em que as pistas sonoras são consistentes e em compasso com a atividade criativa do sujeito, forneceriam melhor suporte para a explorabilidade (equivalente ao fator facilidade ou ao menor esforço cognitivo) e para a colaboração. Portanto, a hipótese alternativa é que a similaridade entre o tipo de sons ambientais e o tipo de sons utilizados na atividade criativa aumentaria o desempenho criativo. Já o contraste e, conseqüentemente, a imprevisibilidade dos sons do ambiente reduziria o desempenho. Essa é a principal pergunta a ser respondida pelo presente estudo.

3. Preparação das sessões experimentais

3.1. Ferramentas e dispositivos

Como prova de conceito da marcação temporal foi desenvolvido o protótipo mixDroid [Radanovitsck et al. 2011] no sistema operacional aberto Android para dispositivos portáteis. O protótipo mixDroid 1.0 permite combinar sons em tempo real através de um teclado virtual com nove botões acionados pelo toque na tela sensível. As amostras sonoras são selecionadas através do utilitário AndExplorer. As mixagens são mantidas na memória do dispositivo para posterior arquivamento em formato PCM, estéreo, com taxa de amostragem de 44.1 kHz e resolução de 16 bits. Para realizar o presente estudo, instalamos mixDroid em três dispositivos portáteis rodando o sistema operacional Android versões 2.2 e 2.3.6: um telefone celular Samsung Galaxy 15500B, um tablet Samsung Galaxy Pocket, e um tablet Koby.

3.2. Características das amostras sonoras

Utilizamos 36 amostras sonoras de ambientes domésticos, divididas em 4 grupos de 9 amostras cada uma. A duração dos sons variou entre 8 e 41 segundos. As amostras foram gravadas com um microfone Sony estéreo modelo ECM 907, utilizando um gravador digital portátil Fostex. A coleta foi realizada em dois ambientes diferentes de uma residência (ver figuras 1 e 2 e descrição das condições experimentais), na cozinha e no banheiro, durante atividades cotidianas de sujeitos que depois participaram nas sessões experimentais. As amostras foram captadas e editadas em formato PCM, estéreo, 44.1 kHz, 16 bits. A edição e segmentação foi feita no editor Audacity.

3.3. Perfil dos sujeitos

Devido ao perfil cotidiano das atividades experimentais, foi necessário que houvesse um acordo prévio entre os pesquisadores e os sujeitos em relação à disponibilidade de horários. Durante a preparação dos experimentos os participantes foram contactados para agendamento das sessões. Cada participante disponibilizou aproximadamente duas horas, e esse foi definido como o tempo máximo para cada sessão.

Participaram 12 sujeitos, 6 homens e 6 mulheres. Todos os sujeitos tiveram alguma experiência prévia com tecnologia. É interessante observar que todos os participantes possuem telefone celular, porém somente 9 participantes declararam ter experiência com uso de computadores. Seis sujeitos já utilizaram ferramentas multimídia (YouTube, MediaPlayer) e 3 sujeitos conhecem e usam tecnologias desenvolvidas para fins musicais, incluindo Audacity, Musescore e editores comerciais. Um sujeito teve experiência prévia com a ferramenta mixDroid.

Em relação ao conhecimento musical, cinco sujeitos disseram ter alguma experiência em música e quatro sujeitos participaram em cursos formais de música por dois anos ou mais. Portanto, para fins de análise consideramos que 8 sujeitos são leigos e 4 são músicos ou estudantes de música. Esse grupo de participantes que realizam atividades musicais regularmente está formado por 3 homens e uma mulher. Isso significa que o grupo de homens tem 50% músicos e 50% leigos. Já o grupo das mulheres tem 83% de participantes sem conhecimento musical. Esse aspecto do perfil dos sujeitos é importante ao considerar o impacto do sexo e do treino musical nos resultados obtidos.

4. Procedimentos

O experimento incluiu cinco condições: o tipo de ambiente (doméstico ou comercial); a postura corporal do sujeito durante as atividades (sentado ou em pé); o tipo de atividade criativa (exploração, criação ou imitação); o número de participantes (solo ou duo); e o tipo de amostra sonora (sons de cozinha ou sons de banheiro). Neste artigo relatamos e discutimos os resultados relacionados ao tipo de ambiente e ao gênero dos sujeitos.

Foram realizadas 108 atividades criativas (iterações) abrangendo diversas combinações de condições experimentais. 54 iterações foram feitas por dois sujeitos e 54 foram individuais, 43 correspondendo a homens e 65 a mulheres. Metade das iterações utilizaram sons sampleados no banheiro e a outra metade sons da cozinha. 61 mixagens foram feitas no ambiente doméstico e as restantes no ambiente comercial (47). Durante 27 iterações realizadas no ambiente comercial, os sujeitos ficaram em pé, e nas outras 20 os sujeitos permaneceram sentados.

Tabela 1. Quadro de iterações por condição experimental.

variável	condição	número de iterações
número de sujeitos	solo	54
	duo	54
sexo dos sujeitos	homens	43
	mulheres	65
tipo de ambiente	lar	61
	loja	47
tipo de amostra sonora	sons banheiro	54
	sons cozinha	54
postura corporal	em pé	27
	sentado	20

4.1. Descrição das condições experimentais

4.1.1. Tipo de ambiente: doméstico e comercial.

A primeira sessão do experimento foi realizada no ambiente doméstico. Dentro do lar, o sujeito escolheu o local priorizando o ambiente confortável para as atividades criativas. No entanto, não foi realizada nenhuma modificação intencional nas atividades cotidianas – mantendo barulhos intrusivos e outros distratores – de forma de não afetar a validade ecológica do experimento. A casa utilizada na condição de ambiente doméstico está localizada no bairro Conjunto Esperança próximo a Avenida Ceará (figura 2). A casa tem um espaço construído de 7 metros de largura por 11 de comprimento dividido em três quartos, uma sala e uma cozinha. Parte das iterações foram realizadas nos locais de captação das amostras sonoras - cozinha e banheiro - e a outra parte foi feita na sala.

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

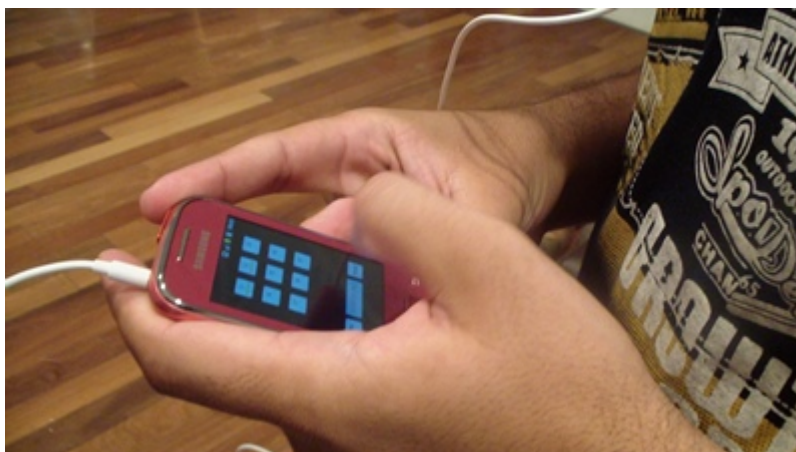


Fig. 1. Sujeito fazendo mixagem em ambiente doméstico (casa)

O local da segunda sessão foi o Shopping Via Verde, em Rio Branco, Acre: um espaço comercial altamente reverberante com circulação de grande número de pessoas. Para indicar esse ambiente utilizamos alternadamente a denominação shopping, loja e centro comercial (figura 3). O ambiente usado para as sessões fica localizado no centro do shopping, numa área de passagem com grande fluxo de pessoas. Nesse espaço há lugares para as pessoas se acomodarem para descansar. Durante o experimento identificamos algumas classes sonoras que também estavam presentes no ambiente doméstico, por exemplo: motor de geladeira, motor de liquidificador, sons de gavetas abrindo e fechando, sons de frigideiras na cozinha do restaurante, e sons de vassoura.

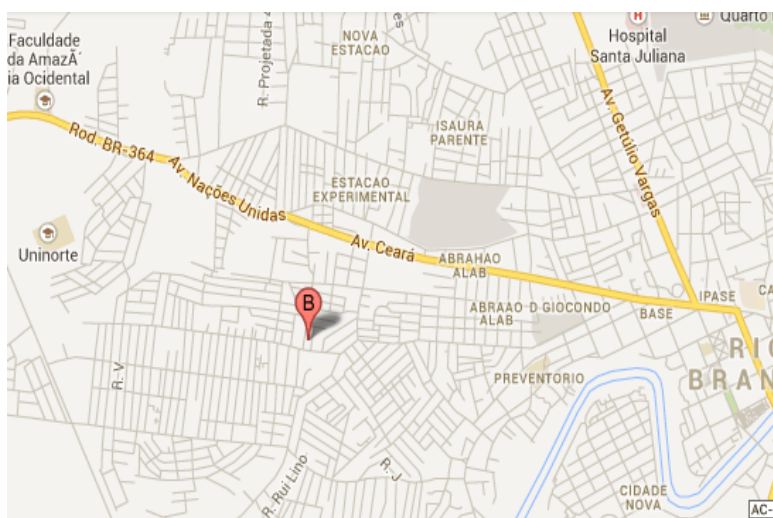


Fig. 2. Localização do ambiente doméstico (casa), indicada com a letra B.

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

4.1.2. Postural corporal: sentado, em pé.

As posturas corporais dos sujeitos durante a realização das atividades foram duas: sentado e em pé. Tanto na posição sentado quanto na posição em pé, o sujeito segurava o dispositivo com as duas mãos.

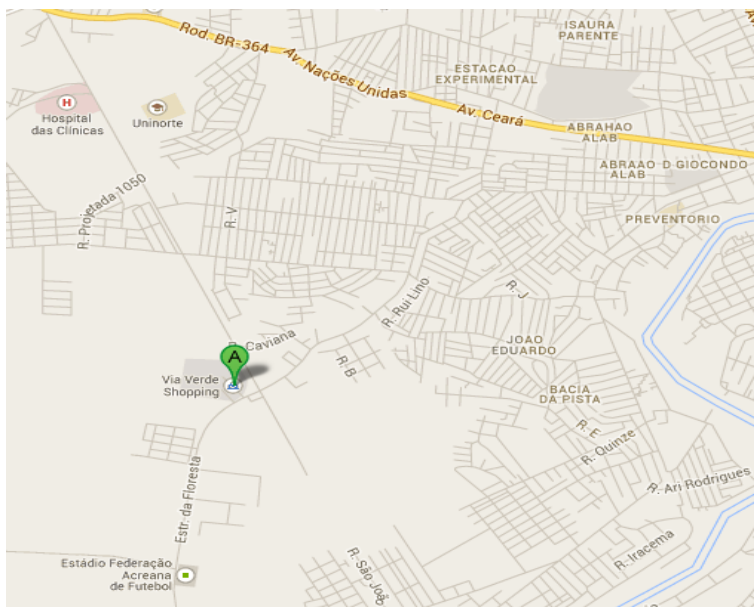


Fig. 3. Localização do ambiente comercial (loja), indicada com a letra A.

4.1.3. Atividade criativa: exploração, criação e imitação.

As sessões incluíram três tipos de atividades. Após uma demonstração sucinta de uso por parte do experimento, inicialmente o participante era convidado a utilizar a ferramenta sem limitação em relação ao tempo de uso e sem indicar qual era o objetivo da atividade (condição exploração). Seguidamente, os sujeitos realizavam uma mixagem dentro do tempo limite de 30 segundos, utilizando até nove amostras sonoras (condição criação). A terceira atividade consistia em imitar um modelo de mixagem já existente (imitação).

4.1.4. Número de participantes na atividade criativa: solo e duo.

Na atividade de imitação na condição solo, o sujeito repetia um modelo de mixagem já existente. Já na condição duo, ele tinha que reproduzir a mixagem realizada por outro participante. Tanto a criação quanto a exploração eram feitas simultaneamente pelos dois sujeitos. Em alguns casos não foi possível realizar a condição sentado em duplas por falta de espaço físico (no banheiro do espaço doméstico).

4.2. Ferramentas de aferição

Foi utilizada a ferramenta CSI-NAP versão 0,03 (Keller et al. 2011b) para aferir o nível de suporte a criatividade. O CSI-NAP consiste em um formulário eletrônico com os itens apresentados na tabela 4, incluindo um campo para observações por parte dos sujeitos. Os fatores de avaliação visam determinar o tipo de suporte necessário para realizar atividades criativas. Nesta versão, a escala de aferição é de -2 a +2 (tabela 2), com os correspondentes descritores semânticos (tabela 3). Para fins de aplicação, as perguntas foram impressas em folhas de papel e o questionário foi apresentado a cada sujeito imediatamente após a conclusão da atividade.

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

Tabela 2. CSI-NAP v.03: avaliação de fatores de suporte a criatividade.

construto (fator)	avaliação (no formulário)	escala
relevância	o resultado foi bom	-2 a + 2
originalidade	o resultado foi original	
esforço cognitivo (inverso de facilidade)	a atividade foi fácil	
engajamento	fiquei atento na atividade	
diversão	a atividade foi divertida	
colaboração	foi fácil colaborar	

Tabela 3. CSI-NAP v.03: Escala Likert e descritores semânticos.

valor numérico	equivalente semântico
-2	discordo totalmente
-1	discordo parcialmente
0	não sei
1	concordo parcialmente
2	concordo totalmente

5. Resultados

Os dados coletados mostraram efeitos em diversas combinações de variáveis. Porém, somente discutiremos as condições que tiveram maior impacto. Utilizamos seis fatores de avaliação da criatividade: relevância e originalidade (do produto), facilidade, engajamento, diversão e colaboração (relativos à atividade). Facilidade é o inverso de esforço cognitivo, portanto valores altos correspondem a pouco esforço e valores baixos indicam que a atividade demandou muito esforço dos sujeitos. Na apresentação dos resultados usamos a abreviação Média \pm Desvio Padrão, correspondente ao total das iterações realizadas para cada condição.

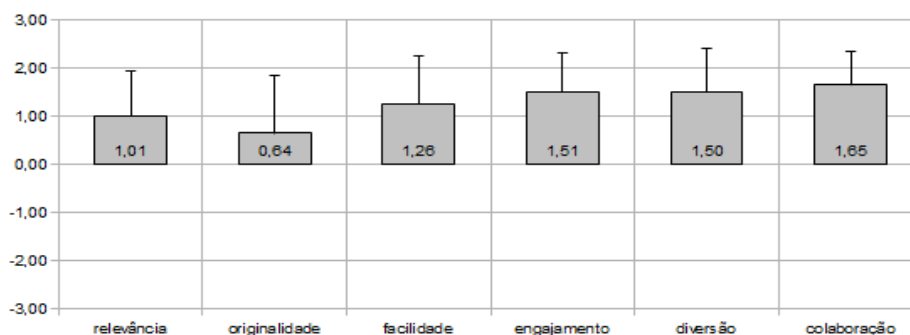


Fig. 5. Resultados do experimento com sons domésticos (108 iterações). A barra indica a média das aferições e a linha indica o desvio padrão.

Os resultados gerais mostram aferições positivas em todos os fatores, tanto para o produto quanto para a atividade. Em particular, engajamento, diversão e colaboração tiveram escores próximos do máximo ($1,51 \pm 0,81$; $1,50 \pm 0,91$; $1,62 \pm 0,69$). A atividade foi avaliada como sendo fácil ($1,26 \pm 0,99$), o que indica um bom suporte para as atividades criativas. Já a relevância do produto obteve escores levemente positivos ($1,01 \pm 0,91$). Na combinação de todas as condições, não é possível determinar se os

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

produtos foram originais ou não ($0,64 \pm 1,21$). Portanto, é necessária a análise das condições específicas para tirar conclusões sobre os produtos criativos.

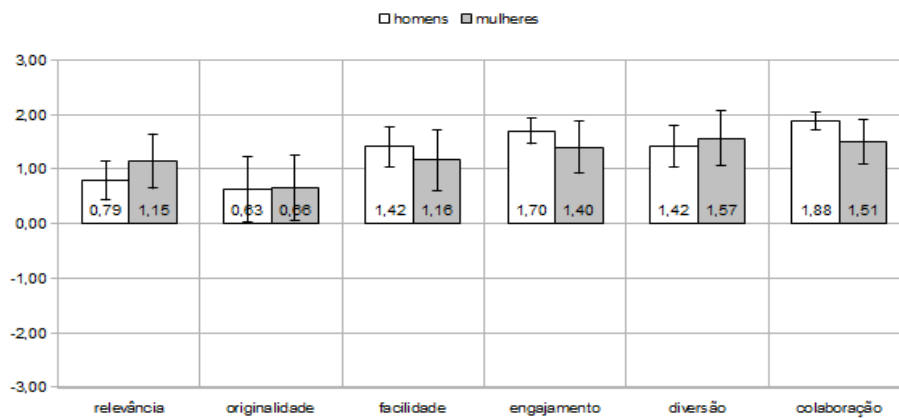


Fig. 6. Resultados: homens e mulheres. A barra indica a média das aferições e a linha indica o desvio padrão.

Como os sujeitos que participaram no experimento foram metade mulheres e metade homens, podemos fazer algumas observações iniciais em relação ao impacto do gênero na avaliação do suporte à criatividade (mantendo em mente o perfil apresentado no item 3.3). As diferenças maiores corresponderam aos fatores engajamento, colaboração e relevância. Os escores dos homens foram mais altos do que os das mulheres nos itens colaboração (37 centésimos) e engajamento (30 centésimos). Em contrapartida, as mulheres acharam os produtos mais relevantes do que os homens (36 centésimos). As diferenças nos fatores facilidade e diversão foram menores, e tanto homens quanto mulheres não deram definições sobre a originalidade dos produtos.

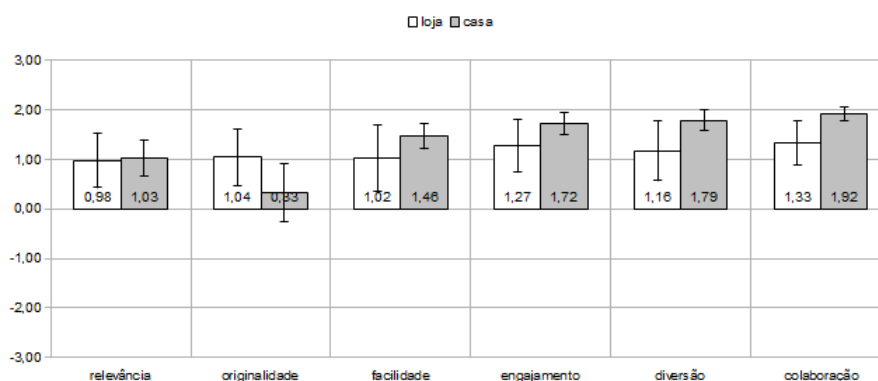


Fig. 7. Resultados: ambiente comercial (loja) e ambiente doméstico (casa). A barra indica a média das aferições e a linha indica o desvio padrão.

A comparação entre a condição ambiente comercial e ambiente doméstico mostra um perfil muito diferente. Todos os fatores relativos às atividades feitas no lar tiveram escores maiores. Facilidade, engajamento, diversão e colaboração foram melhor avaliadas no ambiente doméstico, com pouca divergência entre os escores dos sujeitos. O desvio padrão foi respectivamente: 0,25; 0,23; 0,21; 0,14. E as diferenças entre as médias da condição loja e da condição lar foram: 44; 35; 63 e 59 centésimos. Destacamos que esta foi a única condição na qual os sujeitos deram respostas positivas

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

para o quesito originalidade. Para os produtos realizados no ambiente comercial, a média e o desvio padrão dos escores de originalidade foram $1,04 \pm 0,57$, enquanto a média e o desvio padrão dos escores para os produtos criativos feitos em casa foi somente $0,33 \pm 0,58$. O item relevância não mostrou diferenças entre as duas condições.

6. Discussão geral dos resultados

Iniciamos o trabalho tentando responder as seguintes perguntas. Primeiro, o ambiente tem impacto na aferição do suporte à criatividade? Segundo, se tem, esse impacto é positivo ou negativo? Terceiro, quais são as condições e os aspectos que precisam ser levados em conta na hora de projetar o suporte para os sistemas musicais ubíquos?

A partir dessas perguntas definimos duas hipóteses experimentais. A primeira é decorrente de trabalhos recentes em psicologia ambiental: ambientes desordenados fomentam a criatividade [Vohs et al. 2013] portanto pistas sonoras imprevisíveis deveriam ter um impacto positivo nas atividades musicais ubíquas. A hipótese alternativa é decorrente dos trabalhos experimentais em música ubíqua [Pinheiro et al. 2013], com ênfase na proposta de uso do mecanismo de ancoragem para sustentar a atividade criativa em contexto ubíquo [Keller et al. 2010]: as pistas sonoras consistentes com a atividade criativa fomentariam um melhor desempenho criativo na atividade.

Para coletar dados exploratórios iniciais sobre a relação entre ambiente e desempenho criativo, definimos cinco variáveis com duas condições cada uma: número de sujeitos, sexo dos sujeitos, tipo de ambiente, tipo de amostra sonora, e postura corporal. Dessas cinco variáveis, escolhemos duas para discutir neste artigo: o sexo dos sujeitos e o tipo de ambiente. Esta última é relevante para responder qual das duas hipóteses levantadas tem melhores chances de ser adotada em projetos futuros de design de software para música ubíqua. Quando observamos exclusivamente os aspectos relevantes ao suporte da atividade, os resultados apoiam a segunda hipótese. Porém, ao analisar o impacto nas avaliações do produto criativo, e em particular a originalidade, os resultados são favoráveis à primeira proposta. Em síntese, *ambientes sonoros imprevisíveis fomentam a originalidade mas têm um impacto negativo no desempenho criativo - reduzindo o engajamento, a diversão e as possibilidades de colaboração, e aumentando a demanda cognitiva da atividade.*

Os resultados relativos ao gênero dos sujeitos foram intrigantes. Segundo as perspectivas econômicas da criatividade [Rubenson & Runco 1992; Sternberg & Lubart 1991] um maior investimento de recursos aumenta as probabilidades de resultados criativos. Portanto, um maior engajamento na atividade deveria estar correlacionado a uma avaliação positiva da relevância do produto. Esse não foi o caso. Por um lado, os escores dos homens indicam que as atividades incentivaram a atenção e a colaboração, e que foram mais fáceis de executar do que o que indicam os escores das mulheres. Por outro lado, as mulheres acharam que os produtos foram mais relevantes. Pinheiro da Silva et al. (2012) sugeriram que o perfil dos sujeitos poderia estar influenciando o tipo de avaliações dos produtos criativos. Essa proposta é consistente com os resultados de Barbosa et al. (2010), quanto maior o nível de escolaridade melhor a avaliação do produto. Porém, são necessários estudos mais amplos para determinar se o nível de treinamento musical tem impacto nas aferições da relevância do produto criativo. No estudo relatado em [Keller et al. 2013], os escores das mulheres não musicistas ficaram no mesmo patamar que os escores dos músicos, descartando a possibilidade de um

Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.

alinhamento entre conhecimento musical, gênero dos participantes e escores dos produtos criativos. Um aspecto que não consideramos e que pode estar influenciando os resultados é o sexo do pesquisador que coordena a sessão experimental. Os resultados obtidos colocam em destaque uma possível relação entre quatro variáveis: o gênero do experimentador, o sexo dos sujeitos, o engajamento na atividade e a relevância do produto.

6.1. Implicações para o design de sistemas musicais ubíquos

O presente estudo se insere no contexto da pesquisa recente em música ubíqua preenchendo uma lacuna importante: quais recursos materiais podem ser utilizados no desenvolvimento de estratégias de design oportunista? Os resultados indicam o seguinte. A utilização de pistas cognitivas do ambiente tem impacto em dois aspectos do suporte tecnológico à criatividade: 1. ambientes domésticos ajudam no desempenho criativo, reduzindo o esforço cognitivo e fomentando o engajamento, a diversão e a colaboração; 2. segundo os sujeitos criadores, ambientes públicos reforçam o fator originalidade dos produtos criativos porém não têm impacto na relevância dos produtos.

As implicações são múltiplas. Metáforas de interação que aproveitam recursos ambientais precisam levar em conta a relação entre ambiente e atividade criativa. Ambientes desordenados ou imprevisíveis podem ser adotados para fomentar a originalidade. No entanto, ambientes familiares ou previsíveis melhoram o desempenho geral e podem ajudar a ampliar o acesso de leigos a atividades criativas. Estratégias de design oportunista permitem ampliar o aproveitamento de recursos locais para o suporte a criatividade incorporando diversos tipos de ambiente. Ao vincular pistas cognitivas complexas a atividades que demandam resultados originais e pistas cognitivas familiares a atividades que demandam alto esforço cognitivo, o design de sistemas musicais ubíquos pode oferecer respostas às demandas de leigos e músicos interessados em aumentar seu potencial criativo.

7. Referências

- Ball, L. J. & Ormerod, T. C. (1995). Structured and opportunistic processing in design: a critical discussion. *International Journal of Human-Computer Studies* **43** (1), 131 - 151. (Doi: <http://dx.doi.org/10.1006/ijhc.1995.1038>.)
- Cross, N. (2001). Design cognition: results from protocol and other empirical studies of design activity. In C. Eastman, W. Newstetter & M. McCracken (eds.), *Design Knowing and Learning: Cognition in Design Education* (pp. 79–103). Oxford, UK: Elsevier. (ISBN: 0-08-043868-7.)
- Keller, D., Barreiro, D. L., Queiroz, M. & Pimenta, M. S. (2010). Anchoring in ubiquitous musical activities. In *Proceedings of the International Computer Music Conference* (pp. 319-326). Ann Arbor, MI: MPublishing, University of Michigan Library. <http://hdl.handle.net/2027/spo.bbp2372.2010.064>
- Keller, D., Ferreira da Silva, E., Pinheiro da Silva, F., Lima, M. H., Pimenta, M. S. & Lazzarini, V. (2013). Criatividade Musical Cotidiana: Um Estudo Exploratório Com Sons Vocais Percussivos. In *Anais do Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música - ANPPOM*. Natal, RN: ANPPOM. <http://anppom.com.br/congressos/index.php/ANPPOM2013/Escritos2013/paper/view/2098/420>

- Keller, D., Pinheiro da Silva, F., Ferreira da Silva, E., Lazzarini, V. & Pimenta, M. S. (2013). Opportunistic design of ubiquitous music systems: the impact of anchoring on creativity (Design oportunista de sistemas musicais ubíquos: O impacto do fator de ancoragem no suporte à criatividade). In E. Ferneda, G. Cabral & D. Keller (eds.), *Proceedings of the XIV Brazilian Symposium on Computer Music (SBCM 2013)*. Brasília, DF: SBC.
- Keller, D., Flores, L. V., Pimenta, M. S., Capasso, A. & Tinajero, P. (2011a). Convergent Trends Toward Ubiquitous Music. *Journal of New Music Research* **40** (3), 265-276. (Doi: 10.1080/09298215.2011.594514.)
- Keller, D., Pinheiro da Silva, F., Giorni, B., Pimenta, M. S. & Queiroz, M. (2011). Marcação espacial: estudo exploratório. In *Proceedings of the 13th Brazilian Symposium on Computer Music*. Vitória, ES: SBC.
- Lazzarini, V., Yi, S., Timoney, J., Keller, D. & Pimenta, M. S. (2012). The Mobile Csound Platform. In *Proceedings of the International Computer Music Conference* (pp. 163-167). Ann Arbor, MI: MPublishing, University of Michigan Library.
- Lima, M. H., Keller, D., Pimenta, M. S., Lazzarini, V. & Miletto, E. M. (2012). Creativity-centred design for ubiquitous musical activities: Two case studies. *Journal of Music, Technology and Education* **5** (2), 195-222. (Doi: 10.1386/jmte.5.2.195_1.)
- Miletto, E. M., Pimenta, M. S., Bouchet, F., Sansonnet, J.-P. & Keller, D. (2011). Principles for music creation by novices in networked music environments. *Journal of New Music Research* **40** (3), 205-216. (Doi: 10.1080/09298215.2011.603832.)
- Pimenta, M. S., Miletto, E. M., Keller, D. & Flores, L. V. (2012). Technological support for online communities focusing on music creation: Adopting collaboration, flexibility and multiculturalism from Brazilian creativity styles. In N. A. Azab (ed.), *Cases on Web 2.0 in Developing Countries: Studies on Implementation, Application and Use*. Vancouver, BC: IGI Global Press. (ISBN: 1466625155.)
- Pinheiro da Silva, F., Keller, D., Ferreira da Silva, E., Pimenta, M. S. & Lazzarini, V. (2013). Criatividade Musical Cotidiana: Estudo Exploratório de Atividades Musicais Ubíquas. *Música Hodie* **13**, 64-79.
- Pinheiro da Silva, F., Pimenta, M. S., Lazzarini, V. & Keller, D. (2012). A marcação temporal no seu nicho: Engajamento, explorabilidade e atenção criativa. In *Proceedings of the III Ubiquitous Music Workshop (III UbiMus)*. São Paulo, SP: Ubiquitous Music Group. <http://compmus.ime.usp.br/ubimus/pt-br/node/33>.
- Radanovitsck, E. A. A., Keller, D., Flores, L. V., Pimenta, M. S. & Queiroz, M. (2011). mixDroid: Marcação temporal para atividades criativas. In *Proceedings of the XIII Symposium on Computer Music (SBCM)*. Vitória, ES: SBC.
- Rubenson, D. L. & Runco, M. A. (1992). The psychoeconomic approach to creativity. *New Ideas in Psychology* **10** (2), 131 - 147.
- Sternberg, R. & Lubart, T. (1991). An Investment Theory of Creativity and Its Development. *Human Development* **34** (1), 1-31.
- Visser, W. (1994). Organisation of design activities: opportunistic, with hierarchical episodes. *Interacting with Computers* **6** (3), 239 - 274. (Doi: [http://dx.doi.org/10.1016/0953-5438\(94\)90015-9](http://dx.doi.org/10.1016/0953-5438(94)90015-9).)
- Vohs, K. D., Redden, J. P. & Rahinel, R. (2013). Physical Order Produces Healthy Choices, Generosity, and Conventionality, Whereas Disorder Produces Creativity. *Psychological Science*, 1-8. (Doi: 10.1177/0956797613480186.)
- Zawacki, L. & Johann, M. (2012). A Prospective Analysis of Analog Audio Recording with Web Servers. In *Proceedings of the III Ubiquitous Music Workshop (III UbiMus)*. Retrieved from <http://compmus.ime.usp.br/ubimus/pt-br/node/23>.

A Preliminary Evaluation of the Leap Motion Sensor as Controller of New Digital Musical Instruments

Eduardo S. Silva¹, Jader Anderson O. de Abreu¹, Janiel Henrique P. de Almeida¹,
Veronica Teichrieb¹, Geber L. Ramalho¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE) Caixa Postal
50.740-560- Recife – PE – Brasil.

{ess, jaoa, jhpa, vt, glr}@cin.ufpe.br

***Abstract.** The introduction of new gesture interfaces has been expanding the possibilities of creating new Digital Musical Instruments (DMIs). Leap Motion Controller was recently launched promising fine-grained hand sensor capabilities. This paper proposes a preliminary study and evaluation of this new sensor for building new DMIs. Here, we list a series of gestures, recognized by the device, which could be theoretically used for playing a large number of musical instruments. Then, we present an analysis of precision and latency of these gestures as well as a first case study integrating Leap Motion with a virtual music keyboard.*

1. Introduction

Human-Computer Interaction (HCI) is an important field of study that has gained increasing attention with the emergence of new interaction devices (such as de Nintendo Wii Remote¹ and the Microsoft Kinect²). The use of gestural interfaces are arousing interest, since interaction is essential in several domains of application such as art, medical assistance, simulation, etc. [Freitas et. al. 2012; Singer, Larke and Bianciardi 2003; Wong et. al. 2008].

In this context, the computer music field has been witnessing a growing number of new Digital Musical Instruments (DMIs), that have benefit from new interface technologies [Miller and Hammond, 2010; Montag, Sullivan, Dickey and Leider, 2011; Jordà, Kaltenbrunner, Geiger and Bencina 2005].

Several of these DMIs can be found all over internet sites like Youtube³ and Vimeo³. DMIs such as the Kin Hackt⁴ developed by Adriano Clemente and the work of Chris Vik in collaboration with the dancer Paul Walker⁵, integrate the Kinect sensor

¹ Wii Remote is a motion controller from the Nintendo© Wii Console developed by Nintendo© Company, Limited. Nintendo© Wii website: <http://www.nintendo.com/wii>.

² Microsoft© Kinect is a depth sensor developed by Microsoft© Corporation initially for the Microsoft© Xbox 360 videogame console to be a motion controller, then updated to function with Microsoft© Windows computers. Microsoft© Kinect websites: <http://www.xbox.com/en-US/kinect> and <http://www.microsoft.com/en-us/kinectforwindows/>.

³ Youtube and Vimeo are video-sharing websites that allow users to upload, view and share videos.

⁴ Video available at <http://www.youtube.com/watch?v=YW86yyz0gj0>.

⁵ Video available at <http://www.youtube.com/watch?v=qXnLxi2nZrY>.

capabilities with the software Live⁶ by Ableton to enable musical composition. Projects like the V motion project⁷, built a DMI that uses two Kinect sensors to enhance precision and reduce latency at the same time as giving visual feedback for artistic purposes.

New instruments using other videogame motion controllers such as the Wii remote (Silva, 2012; Miller and Hammond, 2010), projects using multi-touch tablets [Calegario, 2013], multi-touch tables [Jordà, Kaltenbrunner, Geiger and Bencina, 2005] or building their own innovative DMIs such as Mike Waisvisz's "Hands"⁸, developed in the 1980's, have also been explored.

Moreover, several researchers have addressed the evaluation of new technologies and DMIs. Studies evaluated precision in musical properties such as tempo, latency and precision with instruments built with the Wii remote controller alone [Kiefer, Collins, and Fitzpatrick, 2008], with the Wii sensor bar [Peng and Gerhard, 2009] and with several multi-touch platforms available at the time [Montag, Sullivan, Dickey and Leider, 2011]. Others approached piano movement analysis with inertial sensors attached to the user [Hadjakos and Mühlhäuser, 2010].

Since the area is relatively new, there is a lack of consolidated method to evaluate DMIs. Therefore, researches have used different methods to approach this problem, varying from using methodologies based on HCI concepts and theories [Wanderley, and Orio 2002] to a more direct analysis focusing on comparison of acoustic and digital instruments performances [Collicutt, Casciato and Wanderley, 2009].

In 2012, the Leap Motion controller was introduced [Leap Motion inc., 2012]. This gesture controller provides an approximately 150° field of view and uses a depth sensor to track hand features up to 1/100th of a millimeter. This fine-grained control may represent an opportunity to create a new generation of DMI. However, to confirm Leap Motion potential, an evaluation should be performed concerning latency and precision, some of the common bottlenecks in the use of gestural interfaces in DMIs [Silva 2012; Costa Júnior et. al. 2011].

In this paper, we perform a preliminary study and evaluation of the Leap Motion sensor as a tool for building DMIs. We start by listing the conventional music gestures that can be recognized by the device. Then, we analyze precision and latency of these gestures. As part of the evaluation method, we also propose a first case study integrating Leap Motion with a virtual music keyboard, which was called "Crystal Piano".

2. Leap Motion Controller

A DMI is an interactive artifact used for musical purposes that separates the input (gesture control) from the output (sound synthesis) using mapping strategies to associate them [Malloch, Birnbaum, Sinyor, and Wanderley, 2006; Calegario, 2013].

In the last 10 years, various projects illustrate the impact of new interfaces in building DMIs [Silva 2012; Jordá 2005; Wong et. al. 2008]. Leap Motion is one of these recent technology advances that may have an impact on the creation of DMIs.

⁶ Live is a digital audio workstation created by Ableton. <https://www.ableton.com/en/live/>.

⁷ The V motion project video is available at <http://vimeo.com/45417241>.

⁸ "Hands" is a Digital Music Instrument developed at STEIM (steim.org). Description available at <http://www.crackle.org/TheHands.htm>.

The Leap Motion works with two infrared (IR) cameras and three IR LEDs as a depth sensor in a limited field of view (FOV) (Figure 1) of 8 cubic feet (approximately 61 cubic centimeters). Using the stereoscopy from both cameras, the device can minimize errors from tools, fingers and hand features and is built on a unique mathematical model to maximize speed and precision. As the device detects these features, it provides updates in frames of data. Each frame has a list of tracking data such as hands, fingers, tools, recognized gestures and factors describing the overall motion of the scene.

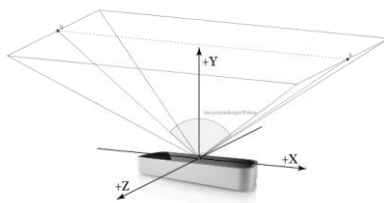


Figure 1. Leap Motion's Field Of View.

This allows Leap Motion to recognize hand features such as hand palm orientation, fingers' length, width and orientation, hand opening and other non-hand features including tools and computer screen location. Furthermore, it incorporates in its Software Development Kit (SDK) the recognition of a few gestures including “circle gesture”, swipe and tap that are shown in its interface (Figure 2). With respect to controllers like Nintendo Wii Remote and Microsoft Kinect, more focused on body and body members, Leap Motion provides a fine-grained hand control, which is clearly promising for building new DMIs.

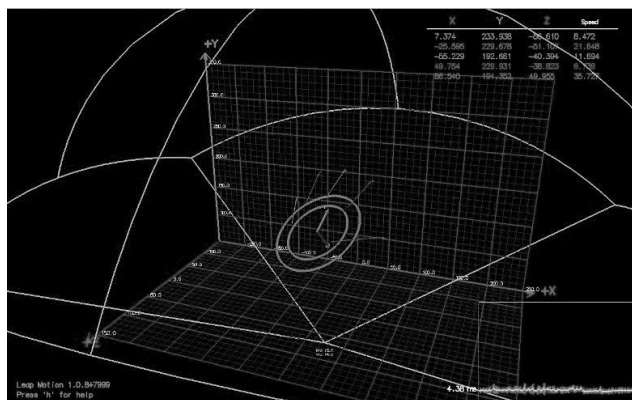


Figure 2. Leap Motion SDK's FOV and tracking features.

Although it has a limited space to work with, the hand tracking is very precise. The Leap Motion SDK comes with three different presets of tracking: Robust mode, Low resource mode and Auto orientation device mode. Furthermore, it has also three different performance sets: High precision, High Speed and Balanced tracking.

In the High Precision mode, the dataflow provides circa 50 data frames per second (data fps), representing about 20 milliseconds (ms) of delay in the computer music field. With the Balanced tracking its data fps increases by a factor of two, reducing the delay to 10 ms, combined with a still good precision. By choosing the High Speed mode, it loses

a perceptible amount of precision with tracking but reduces the delay to 5 ms (approximately 200 fps).

3. Methodology

Considering that the musical field is highly demanding of time precision, low latency is a parameter to be considered when building a new instrument. In digital music instruments, this problem is even more critical, since the gesture control has to be processed by a computer that will perform a series of calculations, then will synthesize a sound, trigger an event or continually manipulate sound properties or processing.

Taking into account that triggering an event requires a lower latency than a continuous sound manipulation [Calegario, 2013], we decided to evaluate the more critical one. Therefore, measuring the latency in which the sensor can capture gestures and trigger an event is an important parameter to evaluate a new technology for building DMIs.

With that in mind, we decided to perform a series of gestures and measure the latency thereof. A simple software was developed so when a finger surpasses a certain threshold a sound was played. A glass table placed at the threshold's position interrupted the movements. To capture the sound of the finger hitting the table, a microphone was placed on it. The delay measure is the time difference between the sound of the finger hitting the table and the sound played by the software.

Furthermore, a simple DMI was built allowing us to evaluate latency and precision of multiple events being triggered sequentially or simultaneously.

After an analysis with images, videos, articles, and the device's documentation, the study of possible applications was made. After choosing to simulate a simple piano, different interfaces and tools were analyzed to identify which of those available in the literature would be appropriate to integrate with the Leap Motion Controller.

Among several interfaces, the Synthesia⁹ game [Synthesia© LLC., 2012], was identified as the most promising one. Using gamification aspects for teaching users to play piano, it functions, naturally, with any MIDI keyboard or device, however it also can function with the conventional QWERTY computer keyboard. The software is able to stop, resume, slow down and speed up the song's rhythm and can play any MIDI song.

However, this former open-source interface has turned private and its code could not be used.

Another attempt to use this interface was trying to simulate the QWERTY keyboard with the Leap Motion. But it brought up two problems. The first one was that, the time of a musical note had to vary with the pressed key, and the synchrony was not ideal. The other problem was due to an Operation System (OS) aspect. When dealing with multi-processing, the application that is being used acquires a focus and has priority of execution, therefore the input from the QWERTY keyboard and the device, goes straight to the Synthesia, without being processed by the application that emulates the Leap Motion, i.e., the system's architecture was not helpful.

⁹ The Synthesia© software is a music educational tool developed by Synthesia©, LLC. Description at: <https://www.synthesiagame.com>.

We, then, built our own interface that simulates one octave of a piano and evaluated the DMI.

4. Precision and Recognition Studies

4.1. Gesture Mapping

After understanding the articulations and movements of the upper body members [Gray 1918], it is possible to understand which of those are responsible for accomplishing the musical gestures. It is necessary to emphasize that most of musical gestures are made by using the upper body extremities (shoulders, arms, elbows, forearms, wrists and hands). Previous studies classified those movements by type of articulation and highlighted its Degrees Of Freedom (DOFs) [Silva 2012].

Particularly, the piano is an instrument that requires short and precise hand gestures to be played. With that in mind, the classification of the interphalangeal articulations was taken as an additional study of the digits:

- Proximal interphalangeal articulation: Articulation between the first and second phalanges (1 DOF).
- Distal interphalangeal articulation: Articulation between the second and the third phalanges (1 DOF).

The possible movements for these articulations are only flexion and extension (therefore 1 DOF) and are more extensive in the proximal articulations. It is important to highlight that the thumb has only one interphalangeal articulation, while the remainder of the fingers (on the hand) have two.

Therefore, combining the movements listed in previous studies and the detail of digits articulations, it is able to reproduce the gestures for playing the piano. Follows a list thereof, according to Table 1:

Table 1. Upper body movements to play a piano.

Movement	Recognition
Shoulder Movement	X
Arm Movement	X
Elbow Movement	X
Forearm Movement	X
Wrist Movement	X
Metacarpo-Interphalangeal Movement	X
First Interphalangeal Movement	X
Second Interphalangeal Movement	X

4.2. Experiments and Results

After the study of the possible gestures, it was time to analyze which of these the tracking sensor in study can recognize. The device was tested using the version v0.8.1.6221 of the SDK. The test included a series of hand manipulations, articulations movements and gestures similar to those required to play the selected instrument.

The device can capture hand features as palm position, orientation and aperture, fingers length, width and orientation. Although it cannot capture the position of the junction points and segments from the whole arm, it can accurately detect the motion passed into the hands, according to Table 2.

Table 2. Piano movements captured by the Leap Motion Controller.

Movement	Recognition
Shoulder Movement	X
Arm Movement	X
Elbow Movement	X
Forearm Movement	X
Wrist Movement	X
Metacarpo-Interphalangeal Movement	X
First Interphalangeal Movement	X
Second Interphalangeal Movement	-

After the study of which gestures the sensor can capture, the simple process to measure the latency thereof was made.

When calculated, the delay for the recognition was approximately of 71 milliseconds (ms) in the High Speed mode. According to [Jordà, 2005], an acceptable latency regarding triggered events has to be lower than 20 ms, so the latency found is unacceptable for triggering events.

5. Latency Study

5.1. Case Study

The developed interface (Figure 3) simulates one octave from a piano and allows the user to play it using the Leap Motion controller. This interface shows the user where his fingers are in space.

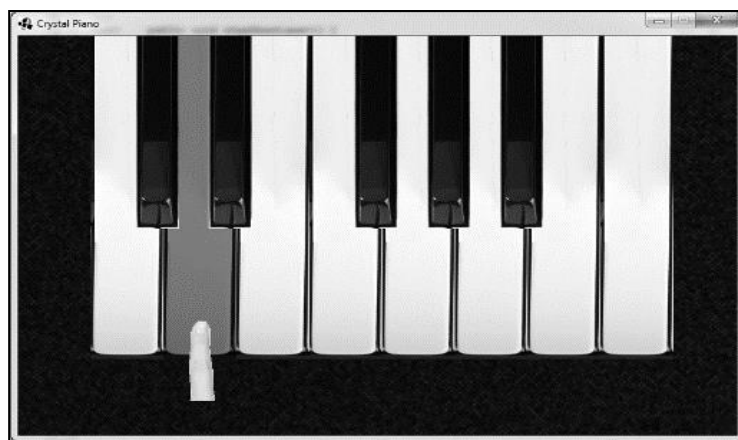


Figure 3. Crystal piano interface.

5.2. Usability Aspects

In matter to provide tactile feedback to the user, it was created a set-up with a glass “table” above the device (Figure 4). To simulate the keys from the keyboard, the glass was marked in the same way a piano should look.

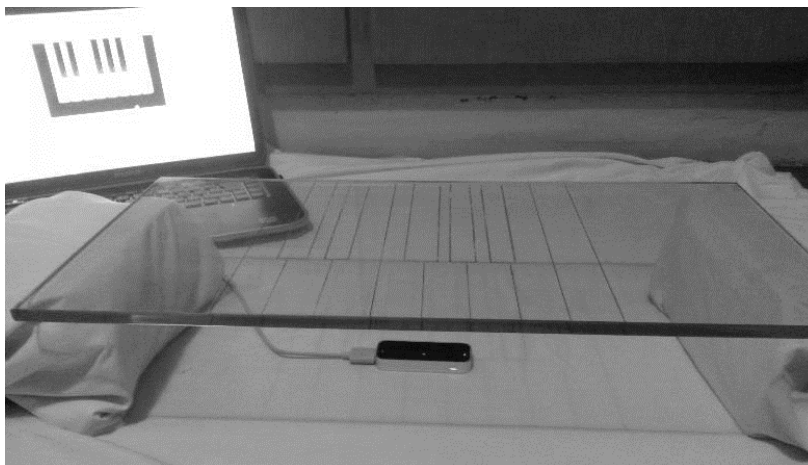


Figure 4. Piano DMI setup with Leap Motion controller under the glass table.

To provide better usability to the system, hand and body postures from online piano educational classes were analyzed. Usability is defined as the ability of an interactive system offers its user in a given context of operation, to carry out tasks in an efficient, effective and enjoyable [ISO 9241, 1993].

Thus, we analyze how the position to play a piano should be. This will help you understand which aspects should be kept to the user to make a good use of Crystal Piano.

The posture is influenced by factors such as:

- The biotype of each user and his body size;
- The position to play standing or sitting, since this may influence respiration;
- The chosen way of supporting the instrument, since this will determine the position of arms and lips;
- The curvature of the nozzle, since it affects the angle at which the blade will be in relation to the lips.

These factors need to be observed to obtain an efficient and comfortable position to play the piano.

It is noteworthy that, in addition to good posture, the position of the hands and their indications are references in the act of playing. To do so, the fingers are numbered for easy positioning in writing scores, where the numbering is the same for the right and left hands.

5.3. Implementation Details

The software used to perform the case study was developed in Java to simulate a music keyboard. It uses gesture input associated with a MIDI protocol to simulate the sound.

The software was created using the following libraries:

- LeapMotion: Official library of the Leap Motion sensor that allows the program to capture the sensor's data information and requires a pre-configured structure to identify the gestures and tools recognized by the device.
- LWJGL: Java library that gives access to high performance cross-platform libraries, such as OpenGL, OpenCL and OpenAL.
- SLICK: Java library compatible with LWJGL to facilitate the development of 2D games.

5.4. Experiments and Results

As a first case study, the user experienced playing different songs. Some aspects were analyzed:

1. Latency of the output.
2. False positive input (the note is played, but the user did not played it).
3. False negative input (the note is not played, but the user did played it).
4. Presence of tactile feedback set-up.
5. Software Interface.
6. Posture.

Follows, the user's perspective (Figure 5):

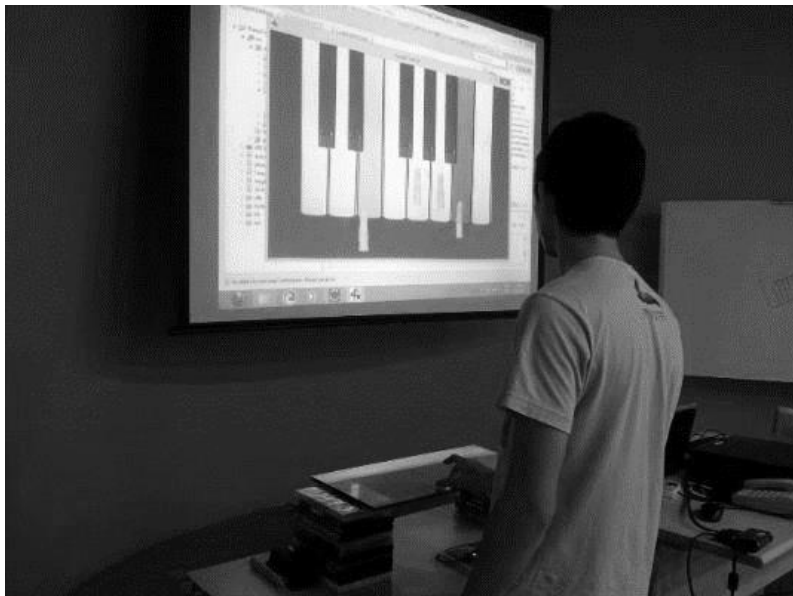


Figure 5. User's perspective.

Regarding physical limitations of the current set-up, the interface only shows one octave (from Do C4 to Do C5).

By using a projector system, the user was able to follow the hand tracking. Virtual fingers provided visual feedback to the user.

The tactile feedback was provided by mounting a glass table above the device as shown in (Figure 6). Markers were used to provide visual feedback on the glass. The table's height was adjusted by an improvised stacking of boxes.



Figure 6. User's perspective of the setup.

The user played three different songs and the following aspects were noted:

- All fingers appear (on the interface) as the pointing finger, losing fidelity;
- The interface shows where the fingers are;
- It appears to be intuitive, in the way that when the user places the hand above the glass, the interface shows the fingers' positions. However, an user that does not have the musical skill to play the piano, cannot perform a complex song (as with a real piano);
- Only displays one octave;
- The latency measure with the render interface was made using the same method as with the gestures recognition and presented a large latency;
- When a song was played fast, the user had problems related to false positive input, false negative input and latency;
- The presence of the glass table was great when it comes to having something to touch and to be able to know where is the sound trigger;
- The user could move his hand freely, but when occlusion and tracking failure factors were considered, the gesture freedom had some constraints;
- When two fingers were too close, the tracking was lost and then recovered, causing false positive input.

The major problem was related to the tracking. When two fingers were too close to each other, the Leap Motion cannot resolve them, losing tracking. In piano gestures, when the user is playing a song, his fingers can be close at some point. If the finger is pressing one note, the system, by losing tracking and recovering it, plays another note or a false positive input. However, if the system does not capture the pressed finger, no note is played and the user experiences a false negative input.

Occlusion plays a great part in the errors as well. The player has to adjust the hand position in a way to the system to identify all the fingers.

When playing a song in a fast way, the user experienced some tilt from the interface. Combining the fast tracking with occlusion and tracking loss, the system,

sometimes, tilts causing frustration in the user. However, when the user plays a song in a slow way, the system flows well.

The presence of a tactile feedback was very enjoyable to the user, which experienced playing songs with and without the glass table. Playing without the table can cause loss of the key's position, forcing the user to suppose it. With the glass, the trigger position is physically displayed and the user gains a more precise control. In addition, the markings on the table provides visual feedback of which piano key is pressed. The downside is that there is no actual key (it is a touch key) nor a force feedback.

Another good point of the tactile feedback set-up is that the user "rests" his hand on the table, since there is a contact to overcome gravity force applied in the hands (normal force). When playing without the set-up, the user experienced some discomfort from hanging his arms and hands in the air for some time.

7. Conclusions

With this study, the analysis of the device shows that it is a powerful tool for recognizing hand gestures with great precision. The SDK showed a dataflow that produces a delay from 5 ms up to 20 ms, with a middle level of 10 ms with great precision. The latency of recognition measured with the prototype developed in Java code presented a still unacceptable delay to build DMIs with precise hand tracking. However, a further study with this device has to be made to verify the performance with other code languages (such as C++) and algorithms.

The case study system proved to be a good simulation only when played in a slow rhythm. In normal-fast paced songs, the system presented an unacceptable latency. The occlusion and loss of tracking problems with the current SDK are relevant, but the Leap Motion proved to be a powerful tool to simulate musical instruments, by having a great gesture repertoire. Meanwhile, the presence of a tactile feedback was of very importance to the user's experience.

The piano, by requesting a particular hand precision, may not be the ideal instrument to be simulated with only one depth sensor.

Concerning future works, the analysis of other music gestures using the device itself or its tracking principle has to be made. Furthermore, the analysis of other music instruments and their gestures is of utmost importance to investigate more the device's capabilities.

It is also important to test other code languages and algorithms to verify the device's performance. Furthermore, it is extremely important to study the integration of this device with other technologies, a work that is already in progress.

A thorough analysis of the prototype's software and improvement of thereof has to be made in order to correct the software's problems and to simulate more than one octave. Another possibility is the development of a driver that, once installed, captures the sensor's data and turns them into a MIDI input signal or that simulates a QWERTY keyboard smoothly. That way, it will be possible to integrate the Leap Motion controller with other tools like the Synthesia© Game or even other applications like a text editor.

8. References

- Calegario, F. C. A., (2013). Sketchument: ambiente de experimentação para criação de instrumentos musicais digitais. Dissertação de mestrado. Universidade Federal de Pernambuco, Recife.
- Collicutt, M., Casciato C. and Wanderley, M. M. (2009). From Real to Virtual: A Comparison of Input Devices for Percussion Tasks, Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 1–6.
- Costa Júnior, J. B., Calegario, F. C. A., Magalhães, F., Cabral, G., Teichrieb, V. and Ramalho, G. L. (2011). Towards an evaluation methodology for digital music instruments considering performer s view: a case study. In: Simpósio Brasileiro de Computação Musical, 2011, Vitória - ES. Anais do 13º Simpósio Brasileiro de Computação Musical.
- Freitas, D. Q., Gama, A. E. F., Figueiredo, L., Chaves, T. M., Oliveira, D. M., Teichrieb, V. and Araujo, C. (2012) Development and Evaluation of a Kinect Based Motor Rehabilitation Game. Simpósio Brasileiro de Jogos e Entretenimento Digital, 2012, Brasília. SBC – Proceedings of SBGames 2012.
- Gray, H. (1918). Anatomy of the Human Body. Disponível em <<http://www.bartleby.com/107/>>. Acessado dia 30 de agosto, 2013.
- Hadjakos, A. and Mühlhäuser, M. (2010). Analysis of Piano Playing Movements Spanning Multiple Touches. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from http://www.nime.org/proceedings/2010/nime2010_335.pdf.
- ISO 9241 Part 1(1993). Ergonomic requirements for office work with visual display terminals, Part 1 General Introduction; International Standard ISO 9241-1.
- Jordà, S. (2005). Digital Lutherie. Universitat Pompeu Fabra. Disponível em <<http://dialnet.unirioja.es/servlet/tesis?codigo=19509>>. Acessado dia 30 de agosto, 2013.
- Jordà, S., Kaltenbrunner, M., Geiger, G. and Bencina R. (2005). The reactable. Proceedings of the international computer music conference (ICMC 2005), Barcelona, Spain, pages 579-582. Citeseer.
- Kiefer, C., Collins, N. and Fitzpatrick, G. (2008). HCI methodology for evaluating musical controllers: A case study. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from <http://nime2008.casapaganini.org/documents/Proceedings/Papers/193.pdf>.
- Leap Motion Incorporation. Retrieved from: <https://leapmotion.com/>. 29 de October 2012.
- Malloch, J., Birnbaum, D., Sinyor, E. and Wanderley, M. M. (2006). Towards a New Conceptual Framework for Digital Musical Instruments. Proceedings of the 9th International Conference on Digital Audio Effects (pp. 49–52). Retrieved from http://www.dafx.ca/proceedings/papers/p_049.pdf.
- Miller, J. and Hammond, T. (2010). Wiolin: a virtual instrument using the Wii remote. Proceedings of the 2010 Conference on New Interfaces for Musical Expression (NIME 2010), Sydney, Australia.

- Montag, M., Sullivan, S., Dickey, S. and Leider, C. (2011). A Low-Cost, Low-Latency Multi-Touch Table with Haptic Feedback for Musical Applications. Proceedings of the International Conference on New Interfaces for Musical Expression, (June), 8–13.
- Peng, L. and Gerhard, D. (2009). A Wii-based gestural interface for computer conducting systems. Proceedings of the International Conference on New Interfaces for Musical Expression. Retrieved from www.nime.org/proceedings/2009/nime2009_155.pdf.
- Silva, J. V. S., (2012). Avaliando Interfaces Gestuais Para Prática de Instrumentos Virtuais de Percussão. Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife.
- Singer, E., Larke, K. and Bianciardi, D. (2003). LEMUR GuitarBot: MIDI Robotic String Instrument. Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03), Montreal, Canada.
- Synthesia Game LLC. Retrieved from: <https://www.synthesiagame.com>. 29 October 2012.
- Wanderley, M. M., and Orio, N. (2002). Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, 26(3), 62–76.
- Wong, E., Yuen, W. and Choy, C. (2008). Designing wii controller: a powerful musical instrument in an interactive music performance system. In Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia, pages 82-87. ACM.

A Hybrid Approach to Recommend Partners in Collaborative Musical Environments

Guilherme G. Testa¹, Marcelo S. Pimenta¹, Damián Keller², Victor Lazzarini³

¹Instituto de Informática – Federal University of Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brazil

²NAP – Núcleo Amazônico de Pesquisa Musical - Federal University of Acre (UFAC) – Rio Branco – AC – Brazil

³National University of Ireland, Maynooth, Ireland

{ggtesta,mpimenta}@inf.ufrgs.br, dkeller@ccrma.stanford.edu, victor.lazzarini@nuim.ie

***Abstract.** Collaboration is an important element in the human creative construction process, and takes special significance in music composition. Systems that support collaborative activities with creative focus have been growing largely lately. However, there is a lack of tools that assists users of these systems to find partners to collaborate with in their activities. In this context, this paper presents a hybrid approach – based on both content-based and collaborative filtering – to recommend partners in creative collaborative environments focused on their produced work. Its main characteristics and functionalities are demonstrated by its design on CODES, a platform for music prototyping for lay users.*

1. Introduction

Collaboration is a key concept in the human creative process. Michael Farrel, in his extensive study about collaborative circles, mentions a passage of Henry James where he declares that “every man produces more when it is in company of people working in the same field of work” [Farrel 2003]. Farrel complements saying that collaborative circles affect significantly the creative work. Artists, writers, composers and politicians reported that cooperation is an indispensable part for the developing of their works. Several important studies of sociology, anthropology and psychology have demonstrated the importance of collective creativity to the intellectual and technological advancement of human society [Palispis 1996][Winthrop 1991][Tuomela 2000].

In general, collaborative processes to build products that are not generated by methodological steps require special support for collaboration. Creative activities tend to feed from mutual inspiration between its collaborators: ideas flow around the discussion of elements that composes the creative work in the pursuit of a satisfactory outcome for all (which is not always known *a priori*). This means that all parts must be tuned to the final results and having similar expectations. But at the same time, it is desirable that every person could offer new and innovative ideas to contribute in a unique manner to the final result.

In this sense, conventional approaches to stimulate cooperation in the design of technical products, where the objectives and responsibilities are usually well defined, do

not allow a systematic and opportunistic trading of ideas [Pimenta et al. 2011]. Therefore, they are not suitable for the dynamic and creative nature typically associated with collaboration among musicians. In this sense, a new concept of awareness is needed to define this new form of collaboration [Miletto et al. 2006]. Similarly, we believe that the recommendation of partners in non-technical byproducts should incorporate some main differences to optimize the suggestion of compatible partners.

In this context, we present an approach for recommending people (partners) for collaborative musical environments. This proposal will be demonstrated from its implementation and usage in CODES, a collaborative music prototyping environment focused on people with no prior musical experience [Miletto 2010]. Our recommendation approach is hybrid, combining both content-based and collaborative filtering. Thus, characteristics of users are recovered to create musical profiles, so they can be compared and analyzed for partner recommendation.

This article is organized as follows: section 2 presents an overview of CODES. In section 3, we discuss some related works and studies. Section 4 explores the details of the recommendation system being proposed. Section 5 describes the user experience of the partner recommendation in CODES. Finally, section 6 presents some conclusions and future work.

2. Exploring CODES

CODES is a Web-based environment designed to support Cooperative Music Prototyping (CMP), built with special focus on music novices. The main focus of CODES is to approximate the lay user to music creation, without a need of theoretical music concepts. CODES offers a high level music representation and user interface features to foster easy direct manipulation of icons representing sound patterns (predefined MP3 samples with 4 seconds of duration).

Built with appropriate support features, CODES users can create, edit, share and publish simple musical pieces - or Musical Prototypes (MPs) collaboratively to a restricted group or to the Web. These shared MPs can be repeatedly listened to, tested, and modified by the partners who cooperate on prototype refinement. Users can start a new MP just by choosing the name and dragging the musical instruments to the editing area. Edition in CODES includes intuitive actions to build a MP. It is possible to “drag-and-drop” sound patterns from the sound library, “move”, “organize”, “delete”, “expand” the duration, and “collapse” to listen to the final result. When sharing a musical prototype, the “owner” user can invite other users to listen and collaborate with his prototype, or may send explicit invitations via e-mail to non-members asking them for cooperation. When someone accepts such an invitation, the user becomes a prototype partner and can edit the MP like the owner does.

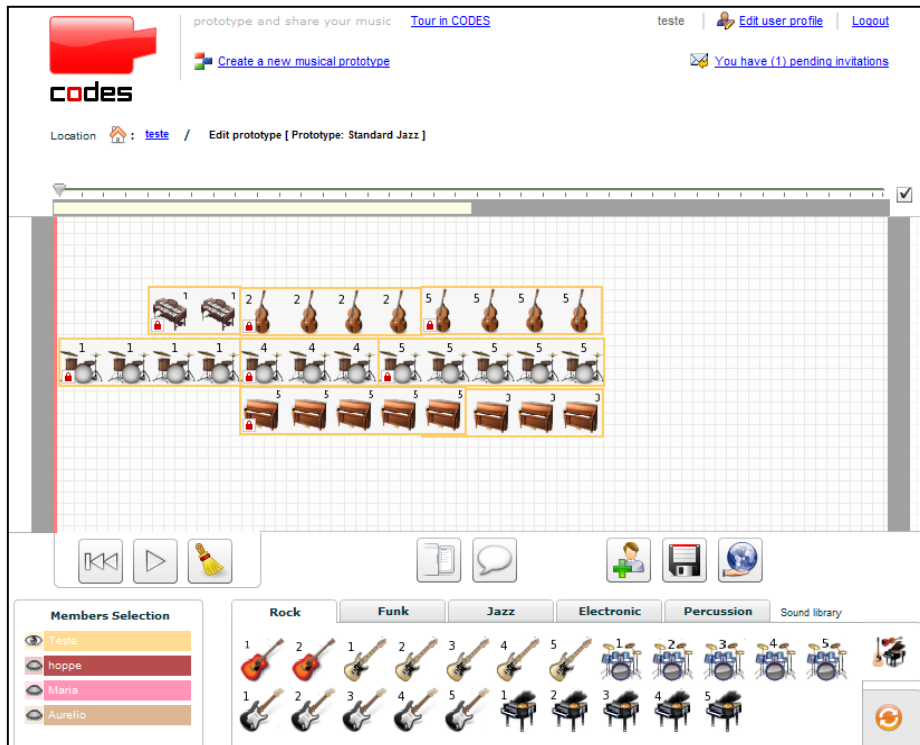


Figure 1. Editing a MP in CODES

The prototypical nature of CODES is designed and built to provide a novice-oriented perspective. All the interactions with the editing area are made as intuitive as possible, without any music theory terms. All prototype partners can discuss and change ideas about each step of the prototype refinement, in order to understand each other's decisions. It is possible to link arguments to every decision made. When someone considers that the resulting sounds are good, a "publication request" can be triggered and the group may discuss and deliberate about the publication of this musical prototype in the CODES home page. This activity is called musical prototype publishing. As an alternative to publishing their music, users can export their musical prototype, and share it at will.

2.1. Awareness Mechanisms in CODES

Through CODES, ordinary users may have the opportunity to be the actors of their own musical experiences. Also, the system offers tools to allow a full cooperation among partners. This implies a focus not only on community management but also on experimenting and participating in specific design practices using a suitable interaction vocabulary. This process suggests the existence of noteworthy distinct kinds of cooperation activities. Awareness and conflict resolution are already considered critical issues in general Computer Supported Cooperative Work (CSCW). However, mechanisms existing in other systems need some adaptation to take into account the idiosyncrasies of the CMP context. The ultimate goal is to provide actual cooperation, social knowledge construction, argumentation and negotiation among the actors of the MP design activities. This type of cooperation is supported by a set of mechanisms borrowed from the Software Engineering and Human-Computer Interaction (HCI) areas and specially adapted for CODES [Miletto 2010].

2.2. Design and Development of CODES

The design and development of CODES adopted a user-centered and incremental approach, taking into consideration social aspects such as the characteristics of the users, contexts, purpose, minimal technology requirements, and the nature of its possible influence on the novice user.

CODES is based on the classical client-server architecture for Web applications. In the current version of CODES, special attention was given to aspects related to interaction flexibility and usability since one of the main goals is to implement an adequate support for manipulation of complex musical information, cooperative activities and group awareness, to provide an effective interaction of the users with each other and with the environment itself. Thus, in the client-side, CODES uses scripts embedded within standard HTML (Figure 1).

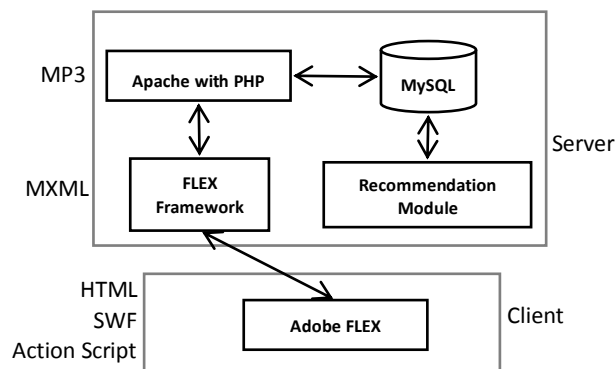


Figure 2. CODES current architecture

On the server side, CODES implements the Model-View-Controller (MVC) architecture. The “Model” part (Apache with PHP) connects the Web server with MySQL database, and represents all the information (the data) of the application; and the “Controller” part manages the communication of data and the business rules used to manipulate the data to and from the model. For this, CODES makes use of Adobe MXML (an XML-based language used to lay out user-interface components for Adobe Flex applications). The Adobe Flex as script language was chosen to allow actions like drag-and-drop, use of sliders, scalable windows, and other facilities to manipulate the sound samples provided for this technology, while the HTML 5 standard is still being on development [W3C 2013]. The sound files used in CODES are small MP3 files which can be quickly downloaded by the client-side ensuring a standard audio quality.

The recommendation module connects to the MySQL database to extract information about the users and to write the data structures of the recommendations. This communication does not interfere in the response of the server to the client.

3. Related Work

There are several studies that approximate partially to the work being proposed in this paper. CODES and its recommendation system grabs several researches areas as

Recommendation Systems, Sound and Music Computing (SMC), Music Recommendation, Computer Supported Cooperative Work (CSCW) and so on.

To introduce the analysis of these areas of study, we will present, in the next subsections, important works classifying them into three main areas: people recommendation, music recommendation and collaborative environments of music composition.

3.1. People Recommendation

Although recommendation of general items is a problem already well researched both in academia as in business and commercial environments, it is not possible to assert the same about people recommendation. In CODES, our focus is to recommend other users (partners) to collaborate with in the process of music prototyping. Recommending individuals is a difficult task due to the subtle inherent and instinctive aspects of the human relations, especially in creative activities. Studies in this area are growing, mainly focusing on social networks.

In music, a person has greater empathy with people with a similar musical taste. This assertive actually can be propagate to several other areas of interest. When more things in common we have with another individual, more we tend to sympathize and relate better with him [Guy et al. 2010][Guy et al 2009]. The key point here is similarity. Briefly, the problem of recommending people is to construct a profile that reflects faithfully the main characteristics of a user, and then, compare those profiles to look for similarities.

Therefore, to make a good recommendation it is necessary an extra effort than just the application of the usual recommendation algorithms. These differences of recommending people instead of products became evident with the growth of virtual communities, and the direct application of the collaborative filtering technique in such systems demonstrated to be ineffective [Cai et al. 2010]. Individuals differ in ways that the standard algorithms are not adequate to present a good recommendation [Lopes et al. 2010]. It is necessary to adopt further measures to encompass all these details.

In this sense, CODES differs from the recommendation engines usually designed to recommend people because our focus is partnership. People recommendation is mainly driven by virtual communities centered on personal relationships (new friendships, for example), or for technical activities. In a creative environment like CODES, characteristics as empathy are not essentials for a good partnership. Otherwise, is very important that the collaborators have the same goals and are also able to work together efficiently.

3.2. Music Recommendation

Several interesting academic studies about music recommendation can be found in the literature, using different focuses. In generic terms, music recommendation deal with the problem of recommending a list of songs that pleases a certain user. To achieve that, it is necessary to understand the elements that construct the musical taste of a user, including as many details as possible, to feed a personal musical profile of the user. In this sense, several information can be used to enrich this profile, like the explicit searches of the user, his reviews or feedback about songs listened, favorites songs and

artists, his personal profile and others [Chen and Chen 2001][Shao et al. 2009][Bogdanov et al. 2011].

With all this information, another question reveals to be crucial: the automatic analysis of the music themselves. To categorize songs, artists and genres efficiently, systems extract semantic information from audio files and use it to group similar artists and songs. Algorithms that provide sound analysis require a considerable amount of processing, but they are becoming viable with the advancements of hardware. Websites that implements these techniques are growing and becoming popular on the Internet and people are changing for listening downloaded music to streaming media as Grooveshark and Pandora

However, almost all of those websites are focused on simple playback of tracks and are not related to music creation or collaboration itself. Also, the focus of recommendation in CODES is not audio analysis, once the information about the genre and style are already explicit in the system.

3.3. Collaborative environments for music composition

Collaborative environments for music composition are also growing with the increase of interactive community websites. Some interesting services are already available on the Internet today, and they have a significant number of users, which demonstrates the growing potential of this area.

Some of them are Indaba, Kompoz and SoundCloud. All these provide a platform for publication, storage and support collaboration in different ways. In Indaba, the user can also make use of an interesting tool to edit and save directly voice or instrument sounds using a microphone connected to the computer. Kompoz is focused mainly in collaboration itself, and it is possible to upload and publish recordings of musical material so that other users can use freely. Sound Cloud, differently, serves as a repository of sounds and songs.

CODES, unlike, focuses in users who have no previous knowledge of music theory and, in this sense, it differs from the examples cited. Additionally, none of these websites provides recommendations to bring users with similar musical tastes together for musical composition.

4. Recommendation in CODES

There are two approaches widely used in recommender systems in general: content-based and collaborative filtering. The first uses representations of data that were accessed in the past to create user profiles. Based on these profiles, the system then recommends new items for a specific user that are possibly relevant, according to his activity history [Balabanović and Shoham 1997]. In collaborative filtering, similarities between the actions of different users are analyzed to predict a behavior. Users with similar profiles are grouped together to help the recommendation. Collaborative filtering is based on the premise that people who historically have similar activities, are more likely to behave similarly also in the future [Sarwar et al. 2001].

Looking to take advantage of the main benefits of both techniques of recommendation, some hybrid algorithms were created mixing characteristics of these two methods. In a hybrid approach, we use information about both the history of user

activities as its similarity with other users to search for compatible profiles. The recommendation can also count with other relevant information to refine the similarity calculation [Kim 2008][Cai et al. 2009].

In CODES, as the user interact with the system, the database is being populated with relevant information that is used by the recommendation engine to produce a musical profile. CODES has a sound library with several predefined samples of instruments of different genres. A combination of these sounds patterns produces a MP. The number of occurrences of different instruments and different genres are collected of the range of MPs created by the user to approximate the preferred genre for composition of a user. Then, with the musical profile created, similarity calculation takes place to achieve recommendation.

The musical profile of the user is represented by a matrix that is filled while the user composes MPs. These matrixes are them combined and compared with other parameters to mold the final profile. The initial genre matrix is populated as follows.

$$ncomp(p)_u = \frac{\sum_{i=0}^m nmp_{i,u} \times 0,9^i}{\sum_{i=0}^m 0,9^i} \quad (1)$$

Where in (1), $ncomp(p)_u$ is the combined grade of the prototypes of the user p calculated separately for each genre u . nmp represents the value of the genre u for a specific music prototype, while i corresponds to the order of the MPs of the user, being 0 the most recent to the oldest m . Therefore, each MP will have a 10% decrease in the relative score of its immediate precedent. The sum of all the music prototypes of the user as their respective genres will result in a matrix of genres that takes into account a large number of compositions to quantify the musical preference of a user.

In order to avoid costly calculations that can compromise the scalability of the system, it is preferable to assign a fixed value to calculate the genre matrix for each user [Agrawal and Srikant 1994]. In this case, considering that the compositions themselves are the predominant factor to draw the music profile, this maximum value should not be too low. For this reason, it is initially assumed $m = 20$. The genre matrix is updated by the profile manager to each new MP edited or created by the user through the collaborative prototyping module.

4.1. Evaluation of public items

The recommender system in CODES implements a well-known rating system to evaluate the opinion of the user to any item. This explicit information is an important way to get a user feedback about others' MPs. It helps to reflect interests and rejections that may go unnoticed by analyzing only the contributions of a single user.

CODES has a rating system of public MPs that follows a standard widely used in evaluations of items, with grades assigned from 1 to 5. The impact of an evaluation alters the influence of a single MP positively or negatively in the total of the public MPs calculated. Table 1 lists the relative influence of each grade.

Table 1. Relative influence according different grades.

Grade assigned	Opinion associated	Relative influence
1	Terrible	-0.25
2	Bad	-0.1
3	Neutral	0
4	Good	+0.1
5	Great	+0.25

For each user, it is constructed an evaluation matrix that stores information about the ratings given by the user to public MPs. Initially, the matrix is completed with equally values distributed to each genre. As the user evaluates MPs, the relative influence of each MP (according to Table 1) of the range of all MPs that the user gave feedback is calculated and the table updated. This matrix is constructed as follows:

$$n\text{atr}(p)_{i,u} = n\text{atr}(p)_{i-1,u} \times (1 + nmp(p)_u \times \text{feedback}) \quad (2)$$

In (2), $n\text{atr}(p)_{i,u}$ corresponds to the grade given by the user p relative to the genre u to all public MPs that the user gave feedback until i . Initially, $n\text{atr}(p)_{i,u}$ has a value equally distributed among all genres; and feedback represents the value of the relative influence column on Table 1 according to the grade assigned.

4.2. Clustering and similarity calculation

Clustering is an interesting way to reduce the universe of comparisons needed to build a recommendation. First, the information about the user activity history and the public evaluations of data items collected in the previous steps are combined. Then, the users are grouped according to a defined criterion and, finally, the similarity calculation is conducted, where users are actually chosen for the recommendation.

The genre and evaluation matrixes are unified using a weighted arithmetic mean (3), where each grade assigned to a public MP will have an effect of 5% of the final grade, up to a limit of 30%. I.e. the user's own MPs correspond to at least 70% of the final grade. This difference in weights seeks to give greater emphasis to the genres of MPs made by the user himself.

$$n\text{final}(p)_u = \begin{cases} \frac{(20 - n) \times n\text{comp}(p)_u + n \times n\text{atr}(p)_u}{20} & \text{if } n < 6 \\ 0.7 \times n\text{comp}(p)_u + 0.3 \times n\text{atr}(p)_u & \text{if } n \geq 6 \end{cases} \quad (3)$$

In (3), $n\text{final}(p)_u$ represents the weighted final score for each of the genres u for the user p . $n\text{comp}(p)_u$ refers to the relative grade of the genre u in the genre matrix of the user p . And finally, $n\text{atr}(p)_u$ corresponds to the given grade of the genre u in the range of all MPs that the user gave feedback (as explained in section 4.1).

After set the unified genre matrix, the clustering part takes place. Each user is assigned to a specific group according to the majority of its genres of interest. This group separation is intended to reduce the universe of calculations needed for

comparison, making the system faster and more scalable [Cataltepe and Altinel 2007]. In CODES, clustering is made according to each user's favorite genres, and these genres are those who contribute, jointly, with at least 80% of the total number of MPs in the unified genre matrix of the user.

The next step is the similarity calculation. One of the most used methods to calculate the proximity between users in recommend systems is the Cosine Similarity (or Vector Similarity), which presents an interesting measure for comparison and prediction [Vozalis and Margaritis 2003]. Note that the summations are calculated over a number of items for which both users u and p have expressed their opinions. The correlation then is calculated as follows:

$$dist_{u,p} = \sum_{i=1}^n \frac{r_{i,u} * r_{i,p}}{\sqrt{\sum_{i=1}^n (r_{i,u})^2} \sqrt{\sum_{i=1}^n (r_{i,p})^2}} \quad (4)$$

In (4), $r_{i,u}$ is the rating of the user u for the corresponding item i , in the n -dimensional item space. If the user didn't rate the current item, the value is set with 0.

A similarity or neighborhood matrix, S , can now be generated, including the similarity values between all users of a given group. The entry at the x -th row and the y -th column corresponds to the similarity between two given users, u_x and u_y .

In possession of S , is possible to recommend the n closest users of the current user consulting the neighborhood matrix. The distance of the acceptable users for recommendation is dynamic, being evaluated and refined through the user feedback about the previous recommendations. As said in the introduction, in creative activities, sometimes people are not looking for a similar user to interact but, unlike, for someone that can contribute with different ideas. This creates a recommendation area, which is shaped with an increase or decrease of 10% of the similarity value for each evaluation of a collaborator recommended by the system.

The list of users of the recommendation area for a specific user u is stored in the user profile itself, and is updated regularly. This communication intends to keep the processing and time consumption of the similarity calculations transparent to the user, considering that the total amount of calculations for the whole recommendation procedure can be costly.

5. User Experience

As explained in the previous section, the process of recommendation is transparent to the user. A list of suggestions about potential partners is showed in three different situations: in the main page, in the summary page and in the music prototyping edition page.

In the first situation, the user invites a recommended partner to start a new MP. The user writes an invitation detailing his intentions, so the potential collaborator can analyze and decide to accept it or not. If the user denies a suggestion, it will vanish from the list of potential collaborators. Otherwise, if the user accepts and collaborates to prototype a MP, this user is allocated to a list of accepted recommendations. Both users then can leave a feedback about the mutual experience, and this feedback will be used to refine the next recommendation procedure.

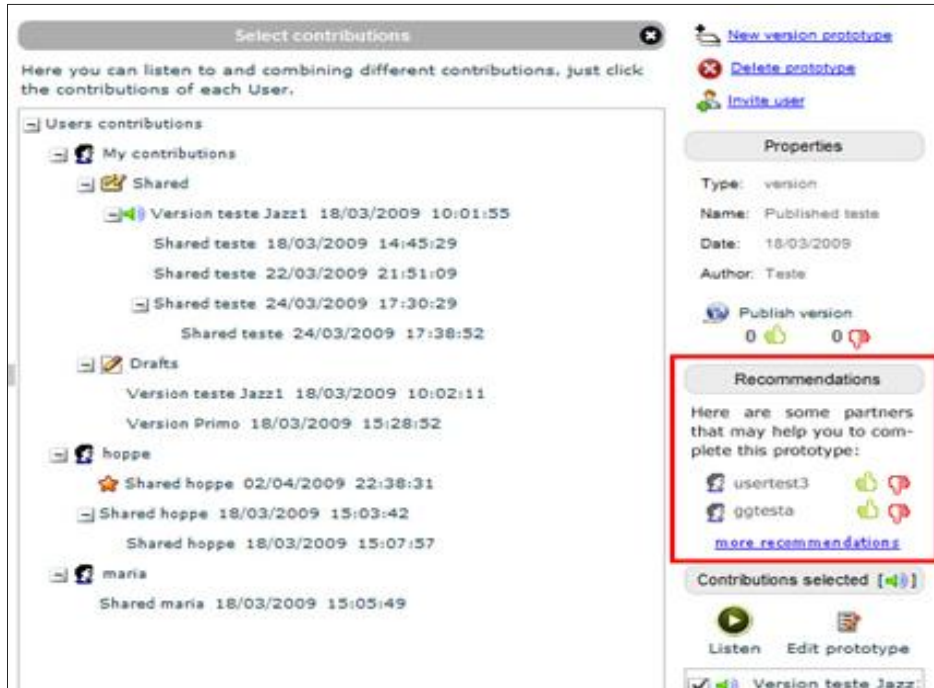


Figure 3. Editing a MP in CODES

In the other two situations, the users will be invited to collaborate in a MP that is already being prototyped (by one or more users). In these cases, the MP is attached to the invitation, so the invited user can listen to the current MP that he is being invited to collaborate with, and can have a better decision.

6. Evaluation

A multi-criteria evaluation of the recommender system proposed in this article is in phase of development, which means that aren't preliminary results of it. The main focus of this evaluation is to get user feedback about the recommendations of partners in CODES, and to adjust eventual needs. The evaluation will be conducted in 5 steps, as follows:

1. Prototyping: a set of 20 different people without prior knowledge about music composition will be invited to create at least 5 music prototypes in CODES, respecting their own musical taste;
2. Evaluation: the set will be indicated to give feedback (rate) of at least 5 others' MPs;
3. Recommendation: with the data collected in stages 1 and 2, the recommendations will be followed by the participants;
4. Cooperation: the users will be guided to create at least 2 MPs with each of the top 2 user recommended;
5. Evaluation: a final questionnaire will be presented to the participants to evaluate the collaborative composition and the recommendation of the users.

This evaluation will be very important to measure the effectiveness of the recommender system presented, as its capacity to give good recommendations. Also it will be an important source to guide future revisions or changes in the recommendation approach.

7. Conclusions

In this paper we presented a hybrid approach to recommend partners in a musical collaborative environment. First, we introduce inherent factors that influence partnership work in collaborative creative activities, such as music. Then we proceeded with a brief presentation about CODES – a platform for music prototyping for lay users –, and an explanation about the recommender engine projected and being implemented in CODES.

Although the preliminary results of the system evaluation with real users have not been collected at this stage, we believe that a hybrid approach concerned and intended to include the subtle aspects of partnership in creative activities is the more suitable approach to make recommendations in collaborative musical environments.

8. References

- Agrawal, R. Srikant, R. (1994) “Fast Algorithms for Mining Association Rules in Large Databases”. VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases. Morgan Kaufmann Publishers Inc. San Francisco, CA.
- Balabanović, M. Shoham, Y. (1997) “Fab: content-based, collaborative recommendation”. Magazine of Communications of the ACM. CACM Homepage archive. Volume 40 Issue 3. ACM New York, NY. 1997.
- Bogdanov, D. Haro, M. Fuhrmann, F. Xambo, A. Gomez, E. Herrera, P. (2011) “A content-based system for music recommendation and visualization of user preferences working on semantic notions”. 9th International Workshop on Content-Based Multimedia Indexing (CBMI). 249-252.
- Cai, J. Francis, J. Gheysens, S. (2009) “Creating a Hybrid Music Recommendation System from Content and Social-Based Algorithms”. GSET '09, Governor's School of Engineering and Technology.
- Cai, X. Bain, M. Krzywicki, A. Wobcke, W. Sok, Y. K. Compton, P. Mahidadia, A. (2010) “Collaborative Filtering for People to People Recommendation in Social Networks”. AI 2010. Lecture Notes in Computer Science. Volume 6464/2011, 476-485.
- Cataltepe, Z. Altinel, B. (2007) “Music recommendation based on adaptive feature and user grouping”. 22nd International Symposium on Computer and Information Sciences, ISCIS,,07. 1-6.
- Chen, H. Chen, A. L. P. (2001) “A Music Recommendation System Based on Music Data Grouping and User Interests”. CIKM '01 Proceedings of the tenth international conference on Information and knowledge management. ACM New York, NY, USA.
- Farrel, M. P. (2003) “Collaborative Circles: Friendship Dynamics and Creative Work”. University Of Chicago Press (Nov. 2003). ISBN 978-0226238678. 328p.
- Guy, I. Jacovi, M. Perer, A. Ronen, I. Uziel, E. (2010) “Same places, same things, same people? Mining user similarity on social media”. Proceedings of the 2010 ACM conference on Computer Supported Cooperative Work. 41-50. ACM New York, NY.

- Guy, I. Ronen, I. Wilcox, E. (2009) "Do you know?: recommending people to invite into your social network". IUI '09: Proceedings of the 14th international conference on Intelligent User Interfaces. ACM New York, NY.
- Kim, K. Lee, D. Yoon, T. Lee, J (2008). "A music recommendation system based on personal preference analysis". First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT,,08. 102-106.
- Lopes, G. R. Moro, M. M. Wives, L. K. Palazzo, J. M. O. (2010) "Collaboration recommendation on academic social networks". ER'10 Proceedings of the 2010 international conference on Advances in conceptual modeling: applications and challenges. 190-199.
- Miletto, E. M. (2010) "CODES: an Interactive Novice-oriented Web-based Environment for Cooperative Musical Prototyping". Ph.D. Thesis. PPGC, Instituto de Informática, UFRGS.
- Miletto, E. M. Flores, L. V. Rutily, J. Pimenta, M. S. (2006) "CODES: supporting awareness in a web-based environment for collective music prototyping". In Proceedings of VII Brazilian symposium on Human factors in computing systems (IHC '06). ACM, New York, NY, USA, 50-55.
- Palispis, E. S. (1996) "Introduction to Sociology and Anthropology". Manila: Rex Printing Company, Inc. ISBN 971234875X. 313p.
- Pimenta, M. S. Miletto, E. M. Flores, L. V. Hoppe, A. (2011) "Cooperative mechanisms for networked music". Future Generation Computer Systems archive. Volume 27, Issue 1, January, 2011. 100-108. Amsterdam, The Netherlands.
- Sarwar, B. Karypis, G. Konstan, J. Riedl, J. (2001). "Item-based collaborative filtering recommendation algorithms". WWW'01 Proceedings of the 10th international conference on World Wide Web. ACM New York, NY, USA.
- Shao, B. Wang, D. Li, T. Ogihara, M. (2009). "Music Recommendation Based on Acoustic Features and User Access Patterns". IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, No. 8.
- Tuomela, R. (2000) "Cooperation: A Philosophical Study". Kluwer Academic Publishers, Netherlands. ISBN 0-7923-6201-2. 431p.
- Vozalis, E. Margaritis, K. G. (2003) "Analysis of Recommender Systems Algorithms". HERCMA'03 Hellenic European Research on Computer Mathematics & its Applications Conference. 732-745, Athens, Greece.
- W3C. (2013) "HTML 5". Available on <http://www.w3.org/html/>. 2013.
- Winthrop, R. H. (1991) "Dictionary of Concepts in Cultural Anthropology". Greenwood Press, New York. ISBN 0-313-24280-1.

Simulation of a “Giannini True Reverber” vacuum-tube guitar amplifier by using wave digital filters

Thomaz Chaves de A. Oliveira^{1*}, Gilmar Barreto¹,
Alexander Mattioli Pasqual²

¹ Departamento de Semicondutores, Fotônica e Instrumentação
Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas, SP

² Departamento de Engenharia Mecânica
Universidade Federal de Minas Gerais
Belo Horizonte, MG

thomazchaves@gmail.com, gbarreto@dmcsi.fee.unicamp.br

Abstract. *The objective of this work is to simulate a tube guitar amplifier, the Giannini True Reverber designed by Carlos Alberto Lopes in the nineteen sixties. The nonlinear “overdrive” characteristics of these devices make them attractive for guitarists since odd harmonics are added into the guitar sound as well as sound compression. The shortcomings of these amplifiers led to the development of DSP simulation techniques over the last few years. Many past DSP simulations of tube amplifiers were implemented using Static Digital Waveshapers for the task of replicating the tube transfer characteristics. Since the physical behavior of such systems is quite complex, physically informed models are necessary for more precision in the simulation, requiring more computer power. A Wave Digital Filter (WDF) simulation of the Giannini True Reverber double 12AX7 preamp is accomplished in this work using Koren’s triode equations and Block Compiler, where each parameter was acquired in the original electronic schematic or by measurement of the real amplifier. The real preamp is compared to the WDF model using the following test signals: single tone, logsweep and transient signal analysis. The results suggest that the current triode models do not cover all circuit topologies, so that further research is required.*

1. Introduction

One of the elements that characterizes the electric guitar sound is the amplifier that the guitar is connected to. The first amplifiers were made up of electron tubes or valves (European name), which were the electronic active devices that dominated the industry up to the 1970’s. Although the solid-state technologies have progressively replaced the vacuum tubes in most applications, these are still widely used in the electric guitar amplification. As a matter of fact, most of the guitar players prefer the valve amplifiers for their “warm” and “soft” sound [Bussey and Haigler, 1981, Oliveira et al., 2012, Barbour, 1998]. However, tube amplifiers — such as those shown in Figure 1 — are heavier, larger, less durable and more expensive than transistor amplifiers. This has motivated the development of devices that aim to emulate the timbral characteristics of vacuum-tube audio amplifiers. Because the interest in these devices is closely related to their nonlinear

*Supported by CAPES.

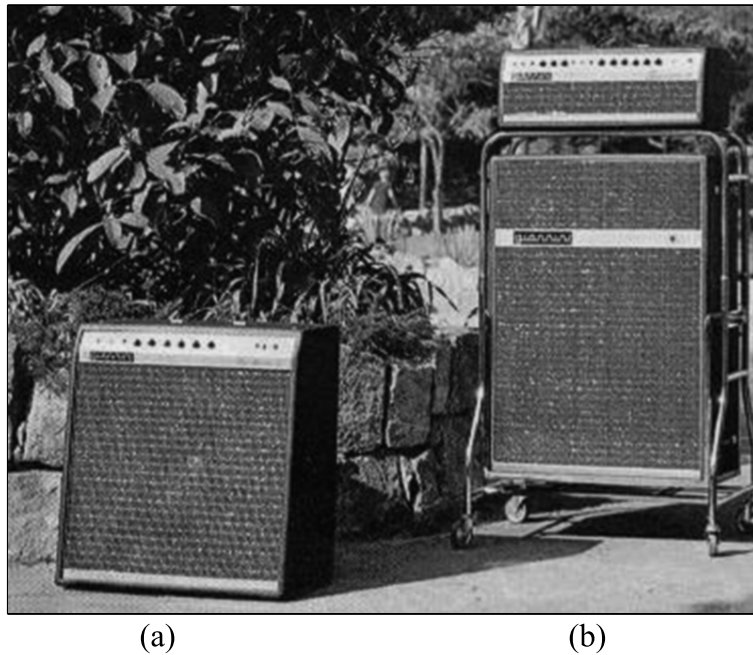


Figure 1: Vacuum Tube Electric Guitar Amplifiers - (a) Giannini True Reverber, (b) Giannini Tremendão, source [Giannini, 1962].

behavior when overdriven [Hamm, 1973, Bussey and Haigler, 1981], there are no simple models that are able to accurately reproduce the tube amplifier sound, as observed by [Pakarinen and Yeh, 2009]. Many different techniques were proposed in recent years attempting to reproduce the sound characteristics of these devices via DSP, as covered in section 3.

This work's objective was to digitally emulate a Brazilian vacuum-tube amplifier designed in the 1960's by Carlos Alberto Lopes, the "True Reverber", manufactured by Tranquilo Giannini S. A.. This simulation used vacuum-tube equations obtained by [Koren, 1996] to resemble the triode tube's non-linear transfer and was implemented using Wave Digital Filters. It was not possible to model the entire amplifier due to its complexity, so only the first two triode 12AX7 stages in cascade configuration were implemented. All the circuit element characteristics such as capacitor values and voltages sources were either obtained from measurements from the real amplifier or by consulting the original circuit schematic. The digital virtual model was validated by a set of test signals, that were input in both the "Virtual Analog" digital amplifier and the real amplifier for the comparison of the outputs. The main objective was to test if such vacuum-tube emulation procedure can produce good sound reproduction of these devices when "overdriven" when using real circuit parameters, since many past work use non-circuit derived parameters, specially lower power supply voltages [Pakarinen and Karjalainen, 2010].

2. Giannini True Reverber

In recent years Brazilian made "vintage" tube amplifiers have been rediscovered by musicians. Since then, many amplifiers from the 1960's and 1970's have been restored by technicians to be used by the electric guitarists of the 21th century. One of the reasons for this is the electron tube comeback of the late 1990's [Barbour, 1998] and the fact that internet has enabled vacuum tube enthusiasts to gain knowledge on such circuits, in order for them to be restored back to proper working conditions.

The amplifiers developed by the Brazilian manufacturers from the 1960's and 1970's are characterized by their high standard of electronic components and tonal char-

acteristics, it is also worth noting that since most of them utilized 6L6GC as output tubes, these amplifiers excel in clean sound as opposed to EL34 output tubes amplifiers such as the ones produce by Marshall and other European brands. Another characteristic of these amplifiers are their high output power, ranging from 18 W to 300 W. In today standards, a amplifier output power rarely exceeds 100 W. As a consequence, amplifiers such as Palmer P200 and Giannini Tremendão are considered “*vintage*” relics, with prices around US\$1500.00, when these are either original or carefully restored to their original standards.

The amplifiers designed by Carlos Alberto Lopes and manufactured by Giannini equipped great names of Brazilian music of the 1960’s and 1970’s, such as the guitarists Pepeu Gomes, Sérgio Hinds from “O Terço”, and Sérgio Dias, lead guitarist of “Mutantes”. In this work, the simulation of a tube amplifier of this time is accomplished. The simulated amplifier is the 1965 Giannini True Reverber designed by Carlos Alberto Lopes in the 1960’s [Sabatinelli and Solon, 2004]. Figure 1 shows this amplifier on the left and the more famous Giannini Tremendão amplifier on the right, named after the famous artist Erasmo Carlos, who was known in the 1960’s as “Tremendão”. The technical specifications of the True Reverber II amplifier from 1965 are listed in Table 1.

Output Power	60W RMS
Preamp Tubes	3 x Miniwatt 12AX7
Poweramp Tubes	2 x RCA 6L6GC
Phase splitter	1x 12AT7
Tone Controls	Grave (bass), Agudo (treble) Volume, Brilho (bright switch)
Tremolo Controls	Velocidade (velocity), Intensidade (intensity)
Reverb Controls	Reverber (Reverb unit volume)
Switches	Liga (Power) and Fase (AC phase switch)
Connections	Entrada(input) Reverb Send, Reverb Return, Foot Switch, alto-falantes (speakers)
Input Impedance	680 k Ω
Speaker impedance	4/8 Ω
AC input	110V/ 220V (1960’s AC standard)
Dimensions	30cm x 72cm x 24,5cm

Table 1: True Reverber II technical specs, source [Giannini, 1962].

Figure 2 presents the Giannini True Reverber II pre-amplifier electronic schematic. In this circuit a double triode 12AX7 vacuum tube is the amplification device. This tube has two independent triodes inside its glass enclosure. This pre-amplifier circuit is a double common cathode circuit where the gain in each single triode stage is close to 70. The electric guitar is plugged into the grid of first triode of the 12AX7 (on the right), its signal receives a gain of 70 and the its output is injected in the grid of the second triode of the 12AX7 (on the left) thorough a coupling capacitor. The overall gain of this circuit is around 74 dB, and varies according to the tube’s manufacture, since vacuum tubes of the same type have significant variations. For more theory on vacuum tube amplification refer to Radiotron Handbook [Langford-Smith, 1953].

The DSP simulation theory for this type of circuit and related work is covered in the next section of this work.

3. Tube Guitar Amplifier Simulation

Since the modeling of vacuum tube amplifiers and analog effects possesses its own challenges, this research field has been called “Virtual Analog” [Pakarinen et al., 2011]. One

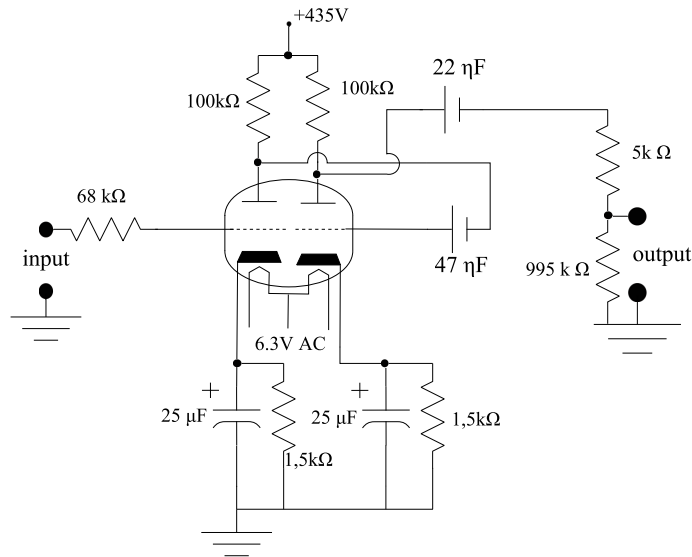


Figure 2: Giannini True Reverber pre-amplifier schematic

of the main advantages of virtual analog models is that many different amplifiers can be simulated in the same DSP (Digital Signal Processing) system by adjusting the simulation parameters [Pakarinen and Yeh, 2009].

As far as amplifier simulation is concerned, the models must be categorized as either linear or nonlinear since analog amplifiers have corresponding linear circuits (RC or LCR filters) and nonlinear circuit elements (transistors, operational amplifiers, diodes and electron tubes). Unlike nonlinear systems, a linear system is characterized by its impulse response and does not increase the bandwidth of the input signal. From this point on, it is necessary to identify a circuit model as being linear or nonlinear. Digital models of linear circuits are mostly accomplished by the use of digital filters that recreate the frequency response of the corresponding analog filtering circuit. For digital models of nonlinear circuits, special functions must be used to generate the nonlinear transfer of these circuits. The nonlinearities will be considered in the next section and are the heart of tube amplifier sound simulations.

3.1. Nonlinear Digital Filters

The most straightforward way to generate nonlinear distortion in digital audio signals is by applying a nonlinear function into each sample of the signal, as illustrated in Figure 5. These functions can be of many types and are known as static waveshapers, where the term “static” is due to the fact that such waveshapers are memoryless systems. Waveshaping functions for audio applications were introduced by [Arfib, 1979] and [Le Brun, 1979] in the late 1970’s. Many simple DSP devices for tube amplifier simulation use waveshaping functions to generate the nonlinearities. Normally the waveshaping functions are run oversampled up to eight times the original sampling rate to avoid aliasing in the output signal [Pakarinen and Yeh, 2009]. An example of these functions was implemented by [Gallo, 2011] proposed in a patent that comprises a complex waveshaper to emulate the effects of vacuum-tube amplifiers:

$$f(x) = \begin{cases} \frac{(k_1 + x)}{(k_2 - x)}, & \text{if } x < a \\ x, & \text{if } a \leq x \leq b \\ \frac{(x - k_3)}{(x + k_4)}, & \text{if } x > b \end{cases} \quad (1)$$

where $k_1 = a^2$, $k_2 = 1 + 2a$, $k_3 = b^2$ and $k_4 = 1 - 2b$. The values of a and b can be freely chosen between -1.0 and $+1.0$ in order to control the characteristics of the nonlinear function. Since these two parameters are independent of each other, the positive and negative values of the input signal can be treated separately, which helps mimic the behavior of real vacuum-tube amplifiers. Small signals in the range $a \leq x \leq b$ remain undistorted.

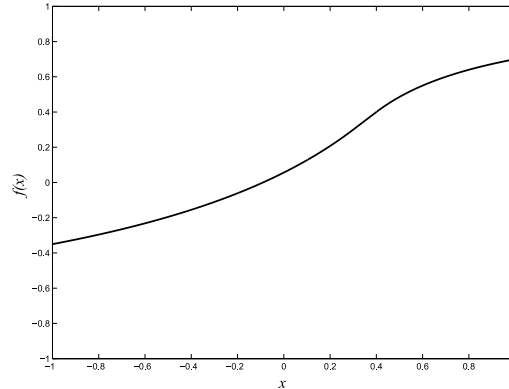


Figure 3: Asymmetric wavelshaper proposed by [Gallo, 2011]

The wavelshaper of Eq. 1 is illustrated in Fig. 3, with parameters $a = 0.3$ and $b = 0.7$. These parameters were chosen to highlight the asymmetric characteristic of this function.

Figure 4 exhibit the distortion introduced by Eq 1 into a 1 kHz test tone. It is worth noting that this asymmetric distortion introduces odd harmonics that have have larger magnitude than the even harmonics. Values a and b can be altered creating many different curves and distortion characteristics for Eq. 1.

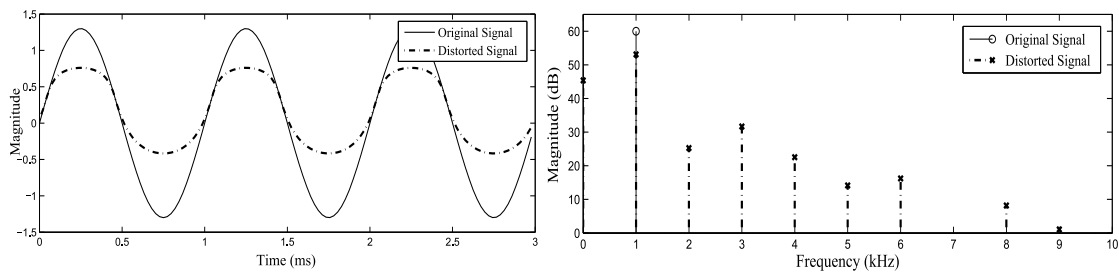


Figure 4: A 1 kHz sine wave filtered by Eq. 1 in the time and frequency domains with parameters $a = 0.3$ and $b = 0.7$.

The patent by [Gallo, 2011] also describes other parameters that can be added to this function to increase the system versatility, which are not presented in this paper for the sake of clarity. Benchmarks works for vacuum-tube amplifier simulations using waveshaping functions are present in the patents by Yamaha [Araya and Suyama, 1996] and Line Six [Doidic et al., 1998].

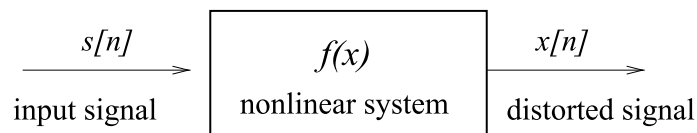


Figure 5: Nonlinear system representation.

4. Physically Informed Amplifier Models

In order to obtain better sound simulation of tube amplifiers alternative techniques have been developed over the last few years based on physical simulation of the amplifying circuits. In order to achieve this, it is necessary to convert a circuit schematic into a system of equations that correspond to the amplifier circuitry. Circuit elements such as capacitors, resistors and inductors are modelled directly using classical circuit equations whereas nonlinear circuit elements such as diodes, transistors and vacuum tubes are modelled with more complex equations to approximate their electric transfer.

4.1. Koren's Vacuum Tube Equations

Vacuum tube phenomenological equations, that is, equations that are not derived from fundamental physics but model the behaviour of a physical phenomena using fitted parameters were developed by [Koren, 1996]. These models were successfully used in most of the physically informed vacuum tube guitar amplifiers digital emulators developed over the last few years, and in SPICE (Simulation Program with Integrated Circuit Emphasis) simulations. Most physically informed vacuum tube amplifier simulations still use Koren's models [Yeh et al., 2010] [Macak and Schimmel, 2011] to model the nonlinear vacuum tube transfer. The triode equations are presented by Eq. 2 and Eq. 3:

$$E_1 = \left(\frac{E_p}{k_P} \right) \log \left(1 + \exp \left(k_p \left(\frac{1}{\mu} + \frac{E_g}{\sqrt{k_{VB} + E_p^2}} \right) \right) \right) \quad (2)$$

$$I_P = \left(\frac{E_1 X}{k_{G1}} \right) (1 + \text{sign}(E_1)) \quad (3)$$

where E_p is the plate voltage, E_g is the grid voltage (where the electric guitar signal is applied), I_p is the plate current, k_{VB} , k_p , $E_1 X$ and k_{G1} are fitting parameters that are altered according to the type of electron tube to be modelled. For that reason, Koren's triode equations are able to model many different triodes. In this model, the plate current is always $I_p > 0$ for all positive plate voltages ($E_p > 0$). Another important aspect of these equations is that grid current is absent, as the grid circuit's impedance is considered to be infinite. Novel equations developed by [Cohen and Helie, 2012] include the grid current, but have not yet shown their potential since they were developed very recently.

5. Wave Digital Filters

Wave Digital Filters (WDF) is a special type of digital filter that have a valid interpretation in the real world. This means that the behaviour of a physical and complex system can be modelled by this approach [Valimaki et al., 2006]. WDF have been used to create digital models of tube amplifiers over the last few years [Karjalainen, 2004, Karjalainen, 2005, Karjalainen, 2008, Pakarinen and Karjalainen, 2010].

The main advantages of this model methodology are: high modularization potential, energy preservation by the use of Kirchoff laws and good numerical properties in its implementations, leading to efficient real time digital models of virtual analog circuits for audio effects.

WDF were originally developed to solve lumped electronic circuits, by creating digital circuit models from the original schematic. Each circuit element is modeled by its circuit equation. The interconnection among elements is accomplished by WDF ports (adapters), in the same way as the original circuit. These adapters can either be serial adapters or parallel adapters.

The main characteristic of the WDF models is that the bidirectional interactions of circuit elements are considered. This is accomplished by the concept of wave scattering propagation. The formalism of WDF theory is based on ‘Voltage’ wave notations, in Equation 4, a corresponds to incoming wave and b corresponds to the reflected wave, V is the Voltage, I stands for current in the same ways as Kirchoff variables and R_p is the port or reference resistance, where all these variables are represented in Figure 6.

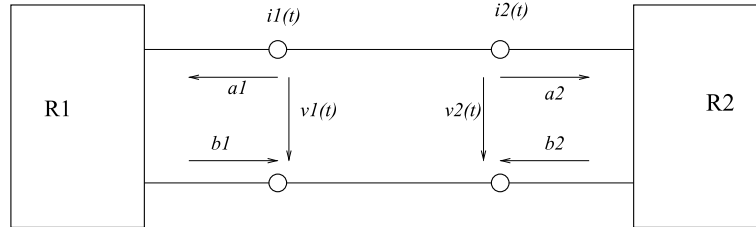


Figure 6: Wave Digital Filters Signal Propagation of Kirchoff Variables.

$$\begin{cases} a = V + R_p I \\ b = V - R_p I \end{cases} \quad (4)$$

The first WDF model of a triode common cathode amplifier was implemented by [Karjalainen, 2005]. The nonlinear tube characteristics is implemented by a nonlinear resistor implemented using Koren’s triode tube equation.

An enhanced version of this triode was accomplished by the same authors [Pakarinen and Karjalainen, 2010], where second order effects were added, such as blocking distortion and the Miller capacitance. The nonlinear processing core also used Koren’s triode models.

A particular digital WDF model for triodes was implemented for the Csound environment by [Fink and Rabenstein, 2011], where the classic comon-cathode stage circuit was implemented. The C-sound opcode was compiled into C code for real-time efficiency. This work also implements the nonlinear behavior of Koren’s triode equations.

A WDF simulation of a linear output transformer model with a triode amplifying power amplifier was reported by [Pakarinen et al., 2009]. This model does not include the nonlinear behaviour of the output transformer but was accomplished with the use of parameters for the speaker and transformer of real devices. In opposite way, a Wave Digital Filter Model of an output chain of power amplifier with a KT88 pentode in triode connection with an audio transformer was accomplished by [Paiva et al., 2011]. In this WDF model, the nonlinearities were introduced by the use of a gyrator and capacitor transformer model. The generated WDF model’s parameters were adjusted from measurements of two real transformers: a Fender NSC041318 and a Hammond T1750V transformers (both in single ended configuration). The model was also validated using these two real transformers. The nonlinearities of real audio frequency transformers are derived from hysteresis and core saturation. More details on WDF theory is covered by [Fettweis, 1986].

6. Nonlinear Audio Metrics

When a digital model is implemented, the comparison of both the analog and the virtual analog models must be made since the Digital Model is always a simplified version of the analog circuit. Special signal processing techniques have been proposed in order to compare output of different systems. One special testing methodology was created

by [Pakarinen, 2011]. This methodology uses single tone, intermodular, logsweep and transient signal analysis of a system's output signal and also the input signal. These techniques were used in a case study by [Oliveira et al., 2012] to characterize nonlinear distortion of an all tube Giannini True Reverber amplifier. Another common validation procedure is to use SPICE simulation of circuits and compare both output signals. Often a musician's trained ear is the best tool for judging either if a digital model performs its task in satisfactory way [Pakarinen and Yeh, 2009].

7. Giannini True Reverber WDF Model

The digital WDF simulation of the True Reverber pre-amplifier was accomplished by the use of two WDF triode models. Each of these models are altered versions of WDF triode models originally developed by [Karjalainen, 2008]. These models use Koren's triode equations in order to produce the nonlinear transfer and high gain of the common-cathode triode amplifiers. The inclusion of real physical parameters derived from the original Giannini True Reverber schematic, obtained from [Giannini, 1980] and from measurements of the real amplifier enabled the creation of a new WDF model. This model has a cascade amplifying circuit that models the circuit of figure 2. The WDF representation of the Giannini True Reverber preamplifier is presented in Figure 7. The model was written in C-Lisp functional programming language and Compiled in the Block Compiler software by [Rabenstein et al., 2007] using the Lisp Works 6.0 environment. The implementation of the double triode stage uses a block approach where the implementation of the first amplifying block uses the input impedance of the second block as an additional parallel output impedance to the plate circuit. The output impedance of the second amplifying block is implemented by the addition of an additional parallel impedance that simulates the phase splitter input impedance on the original circuit. The simulation was executed at 96kHz to avoid aliasing into the audio bandwidth.

7.1. Test Signal Parameters

The test signals for the validation of the virtual True Reverber model were generated using the methodology developed by Pakarinen [Pakarinen, 2011] in order to compare the virtual pre-amplifier and the real pre-amplifier. All test signals were input in the virtual model and in the analog amplifier and were recorded using an Intel I3 2.53 GHz processor PC, with 3 GB of RAM with Windows 7 operating System and a Pro-tools M-Box audio interface. The single tone signal spectral analysis utilized 1 kHz sine wave as input to both real and virtual amplifiers. The logsweep analysis started from 20 Hz to 20 kHz in 2 s for five harmonics. Transient response analysis utilized a 1/10 ratio for 1 kHz signal and transient signal, for 2 s.

8. Results

The inspection of Figures 9, 10 and 11 suggests that the virtual and real True Reverber pre-amplifier have distinct plot patterns for the test results. This suggests that WDF's and Koren's triode models led a simulation that does not match the real amplifier. [Pakarinen, 2011] and [Macak and Schimmel, 2011] also reported distinctive results when comparing real and virtual analog amplifiers.

In most past virtual analog physically informed tube amplifier simulation studies [Karjalainen, 2004, Karjalainen, 2005, Karjalainen, 2008, Pakarinen and Karjalainen, 2010, Yeh, 2012], the authors compared the virtual amplifiers with SPICE simulation results for model validations, as opposed to this work that

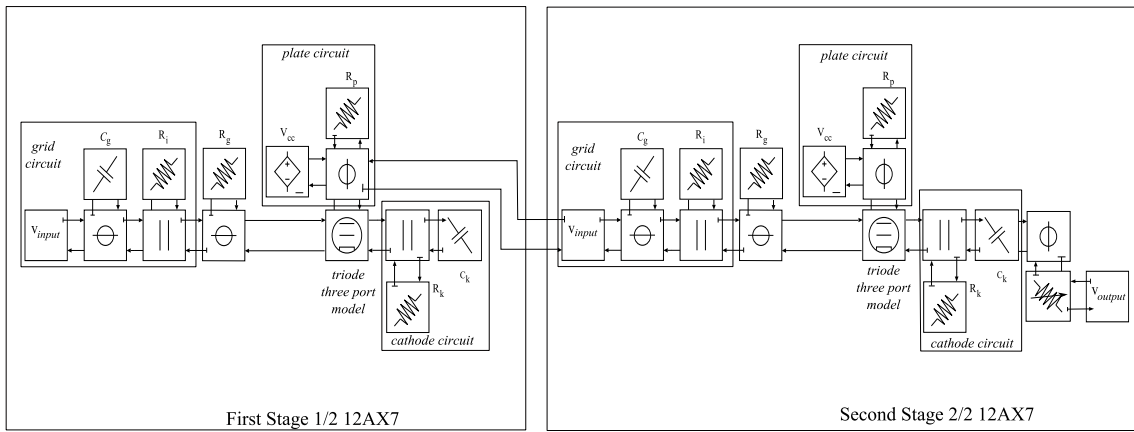


Figure 7: Giannini True Reverber pre-amplifier WDF representation.

compares a digital virtual amplifier to a real amplifier. This suggests that many models were biased since SPICE also uses Koren's triode equations for triode models and it is also a digital simulator. Another important aspect of the past cited works is that the power supply of these virtual circuits utilize lower voltages, such as 250 V. Koren's Equations were designed to work in that range, as opposed to most real amplifiers that operate in the 350 V - 450 V range. That might be the case for which none of these works make mention to any real amplifiers for reference. The simulation of the Giannini True Reverber amplifier was accomplished using parameters from the real schematic and measurements from the real amplifier (435 V power supply voltage), suggesting that in cases of real physical parameters the simulations may be unable to match the transfer of real tube amplifiers using Koren's triode Equations.

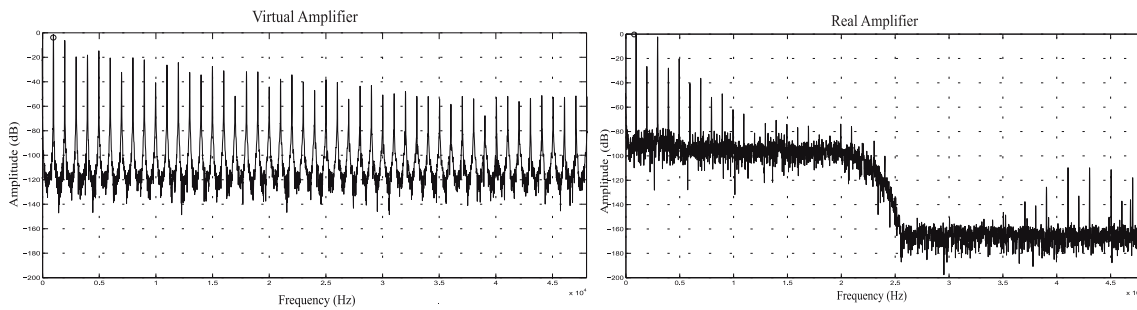


Figure 8: Spectral Analysis of Real and Virtual Amplifiers

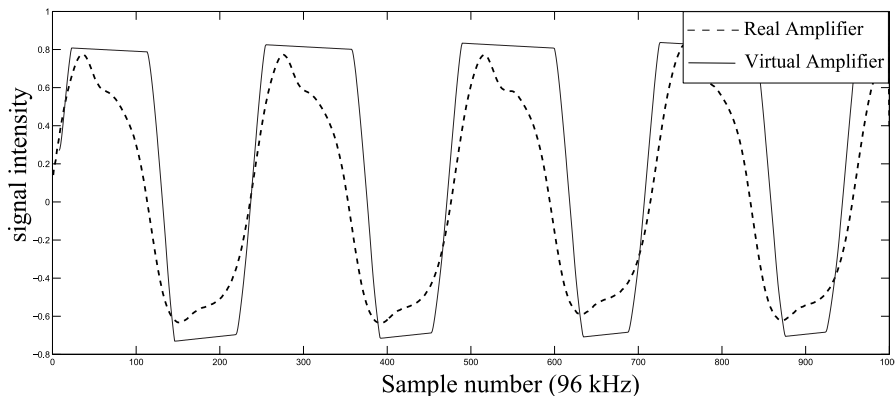


Figure 9: Time domain signals of distorted sine waves

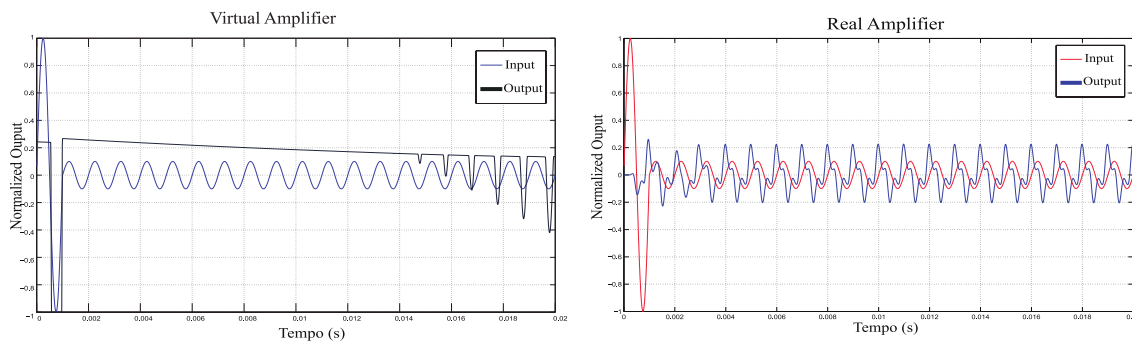


Figure 10: Transient Signal Analysis

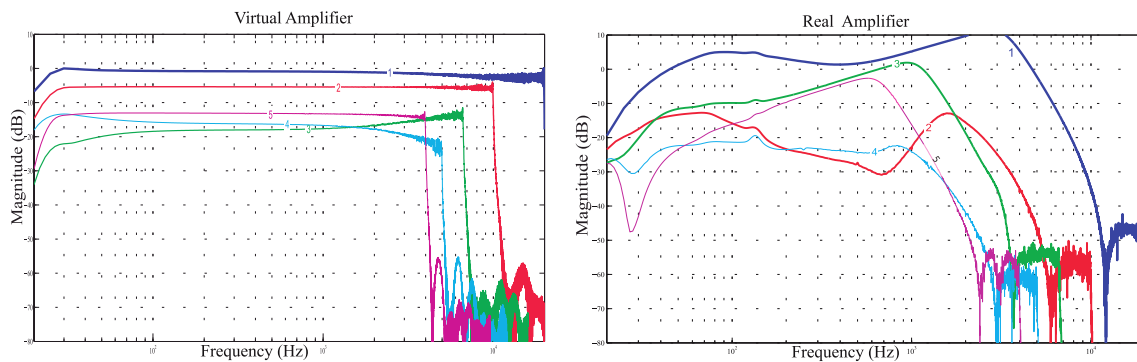


Figure 11: Logsweep Signal Analysis

9. Conclusions

In this work, results of Wave Digital Filters simulation using Koren's triode equations were unable to reproduce the transfer of a real amplifier, the Giannini True Reverber, suggesting that current Vacuum tube amplifier simulation techniques were unable to precisely match the sound characteristics of this amplifier. For real time simulations, there is always a trade-off between accuracy and computation time. Besides, more research should be conducted in order to derive a set of equations that will satisfactorily describe the physical behavior of vacuum tube amplifiers, as proposed by [Cohen and Helie, 2012], but these new triode equations have yet to prove their potential, since they were developed very recently. Finally, it is worth emphasizing that this work's results reinforce the psychoacoustic investigation conducted by [Macak, 2012a]. The quantitative data of this study describes that in 90 % of the blind tests, a trained listener can distinguish if a distorted "Crunch" guitar sound implemented in was either generated by a simulation program or by the actual physical analogue equipment. The same investigation for the Fender style "Clean" by [Macak, 2012b] sound also described that in 83 % of the blind tests the listeners were able to correctly identify if the sound was generated by a simulation program or the actual amplifier. Both this work and [Macak, 2012a] utilized Koren's vacuum tube equations, suggesting that these equations do not lead to digital simulations that can replace the vacuum tube amplifiers for trained listeners. This leads to the conclusion that vacuum tube amplifiers and related audio equipment will not be replaced by simulation programs in the near future. However this does not discard the emulation softwares as accessible tools for the simulation of these devices, as the simulations can be considered as satisfactory for the less trained listeners and musicians. The flexibility of these emulation tools is also an advantage over expensive, large, heavy and hard to handle vacuum tube amplifiers. It is also worth noting that these simulation softwares can be utilized as a sound reference tool for musicians to be acquainted to the classic electric guitar vacuum tube sound.

References

- Araya, T. and Suyama, A. (1996). Sound effector capable of imparting plural sound effects like distortion and other effects. U.S. Patent no. 5570.424.
- Arfib, D. (1979). Digital synthesis of complex spectra by means of multiplication of nonlinear distorted sine waves. *Journal of the Audio Engineering Society*, 27(10):757–768.
- Barbour, E. (1998). The cool sound of tubes - one of the last remaining tube domains is in music applications, but there the devices flourish and even innovate. *IEEE Spectrum*, 35(8):24–35.
- Bussey, W. and Haigler, R. (1981). Tubes versus transistors in electric guitar amplifiers. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 1981*, pages 800–803, Atlanta, Georgia, U.S.A.
- Cohen, I. and Helie, T. (2012). Measures and models of real triodes, for the simulation of guitar amplifiers. In *Proceedings of the Acoustics 2012 Nantes Conference*, Nantes, France.
- Doidic, M., Ryle, M., and Senffner, C. (1998). Tube modeling programmable digital guitar amplification system. U. S. Patent no. 2004.0258.250.
- Fettweis, A. (1986). Wave digital filters: Theory and practice. *Proceedings of the IEEE*, 74(2):270–327.
- Fink, M. and Rabenstein, R. (2011). A csound opcode for a triode stage of a vacuum tube amplifier. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-11)*, pages 1–6, Paris, France, 2011.
- Gallo, M. N. (2011). Method and apparatus for distortion of audio signals and emulators of vacuum tube amplifiers. U.S. Patent no. 2011/0033057.
- Giannini, T. G. S. A. (1960-1980). Esquemas eletrônicos dos amplificadores valvulados da giannini. Electronic Schematic scanned from “Esquematoteca da Rua Aurora”-Santa Ifigênia, São Paulo-SP.
- Giannini, T. G. S. A. (1962). Catalogo da eletrônica giannini. Printed Catalogue of Giannini Products.
- Hamm, R. O. (1973). Tubes versus transistors-is there an audible difference. *Journal Audio Engineering Society*, 21(4):267–273.
- Karjalainen, M. (2004). Discrete-time modeling and synthesis of musical instruments. In *Joint Baltic-Nordic Acoustics Meeting*, Mariehamn, Åland, Finland.
- Karjalainen, M. (2005). Recent advances in multi-paradigm modeling. In *Forum Acusticum*, Budapest, Hungary.
- Karjalainen, M. (2008). Efficient realization of wave digital components. *IEEE Transactions On Audio, Speech, and Language Processing*, 16(5):738–746.
- Koren, N. (1996). Improved vacuum tube models for spice simulations. *Glass Audio*, 8(5):18–27.
- Langford-Smith, F. (1953). *Radiotron Designer’s Handbook*. Wireless Press RCA Victor Division, Radio Corporation of America.
- Le Brun, M. (1979). Digital waveshaping synthesis. *Journal of the Audio Engineering Society*, 27(4):250–266.

- Macak, J. (2012a). Real-time digital simulation of guitar amplifiers as audio effects. Ph. d. thesis, Brno University of Technology - Faculty of Electrical Engineering and Comunication - Department of Comumnications.
- Macak, J. (2012b). Simulation of fender type guitar preamp using approximation and state-space model. In *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-15)*, York , U.K.
- Macak, J. and Schimmel, J. (2011). Real-time guitar preamp simulation using modified blockwise method and approximations. *EURASIP - Journal on Advances in Signal Processing*, 2011(629309):1–11.
- Oliveira, T., Barreto, G., and Pasqual, A. (2012). Analise de distorcao nao linear em amplificadores de audio valvulados. In *XXIV Encontro da SOBRAC*, Belem-PA, Brazil.
- Paiva, R., Pakarinen, J., Valimaki, V., and Tikander, M. (2011). Real-time audio transformer emulation for virtual tube amplifiers. *EURASIP - Journal on Advances in Signal Processing*, 2011(347645):1–15.
- Pakarinen, J. (2011). Distortion analysis toolkit a software tool for easy analysis of nonlinear audio systems. *EURASIP - Journal on Advances in Signal Processing*, 2011(617325):1–15.
- Pakarinen, J., Abel, J., Fontana, F., Lazzarini, V., and Valimaki, V. (2011). Musical applications of real-time signal processing. *EURASIP - Journal on Advances in Signal Processing*, 2011(108):1–2.
- Pakarinen, J. and Karjalainen, M. (2010). Enhanced wave digital triode model for real-time tube amplifier emulation. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(4):738–746.
- Pakarinen, J., Tikander, M., and Karjalainen, M. (2009). Wave digital model of the output chain of a vacuum-tube amplifier. In *Proceedings of the 12th International Conference on Digital Audio Effects (DAFx-09)*, Como , Italy.
- Pakarinen, J. and Yeh, D. T. (2009). A review of digital techniques for modeling vacuum-tube guitar amplifiers. *Computer Music Journal*, 33(2):85–100.
- Rabenstein, R., Petrausch, S., Sarti, A., De Sanctis, G., Erkut, C., and Karjalainen, M. (2007). Block-based physical modelig for digital sound synthesis. *IEEE signal Processing magazine*, 24(2):42–54.
- Sabatinelli, R. and Solon (2004). Da válvula ao simulador. *Áudio Música e Tecnologia*, (150):55–65.
- Valimaki, V., Pakarinen, J., Erkut, C., and Karjalainen, M. (2006). Discrete-time modelling of musical instruments. *Reports on Progress in Physics*, 69(1):1–78.
- Yeh, D. T. (2012). Automated physical modeling of nonlinear audio circuits for real-time audio effects - part ii: Bjt and vacuum tube examples. *IEEE Transactions on Audio, Speech, and language Processing*, 20(4):1207–1216.
- Yeh, D. T., Abel, J. S., and Smith III, J. O. (2010). Automated physical modeling of nonlinear audio circuits for real-time audio effects - part i:theoretical development. *IEEE Transactions on Audio, Speech, and language Processing*, 18(4):728–737.

Interactive Free Improvisation Using Time-domain Extrapolation of Textural Features

Stéphan Schaub^{1*}, Tiago Fernandes Tavares^{2†}, Adriano Claro Monteiro^{1‡}

¹Núcleo Interdisciplinar de Comunicação Sonora – Universidade Estadual de Campinas
Rua da Reitoria, 165 – CEP 13083-872, Campinas, SP.

²Faculdade de Engenharia Elétrica e de Computação – Universidade Estadual de Campinas
Av. Albert Einstein, 400 – CEP 13083-852, Campinas, SP.

schaub@nics.unicamp.br, tavares@dca.fee.unicamp.br, monteiro.adc@gmail.com

Abstract. *In the following paper a simple interactive system for non-idiomatic improvisation is presented. The general approach assumes that a musician improvising in such a setting does not restrict his output to any fixed alphabet, avoids any pre-established grammar and concentrates on articulating the continuation of a musical flow through constant anticipation of its future developments. Using audio classification techniques we map each played phrase to a vector space representing textural features. The system then deduces possible continuations of the ongoing phrase sequence and re-injects past ones in accordance with (manually controlled) instructions as to “contrast” or to “follow” on its predictions. The system was tested with a professional saxophonist and proved a coherent and responsive environment with a wide range of possible extensions.*

Resumo. *No seguinte artigo, apresentamos um sistema interativo para improvisação musical não-idiomática. A abordagem geral assume que um músico improvisando nesse contexto não restringe os elementos de sua linguagem musical a um alfabeto fixo, evita gramáticas pré-estabelecidas, e concentra-se na articulação e continuidade do fluxo musical através da constante antecipação de elementos futuros. Usando técnicas de classificação de áudio, mapeamos cada frase tocada e as transpomos em um espaço vetorial que representa características texturais. O sistema então prediz continuidades possíveis para a sequência em curso e re-injeta segmentos passados segundo instruções (controladas manualmente) para “contrastar” ou “seguir” as predições. O sistema foi testado com um saxofonista profissional e demonstrou ser um ambiente de improvisação coerente e reativo, além de apontar para possibilidades de futuras ampliações.*

1. Introduction

An automatic interactive improviser is considered here to be a musical system capable of accompanying the playing of one or more live improvising musician(s). It is assumed that the output it provides does not reduce to pre-established sets of formulas but is elaborated “on the fly” from information extracted from the session of which it becomes an active participant. Though no *a priori* restriction applies to the form its outputs must take, a

*Supported by FAPESP.

†Supported by CNPq.

‡Supported by CAPES.

general stylistic and contextual coherence is one of the general aims of the development of such systems.

More often than not, automatic interactive improvisers are applied to free, or at least non-idiomatic, improvisational settings. This may be attributed to the facts that their mastery of truly idiomatic styles are still somewhat sketchy and that the exploration of new musical horizons has been, from the start, an important motivation behind their development. Free improvisation, however, does not solely - or even at times primarily - rely on the pitch dimension to develop its musical discourse. Microtonal inflections, “gestures”, alternative modes of playing, etc. might play a central role and bring the “sounding” or “textural” dimensions more or less explicitly into the foreground. Furthermore, as no pre-established grammar or *referent* (in Pressing’s sense [1]) is supposed, the principles from which continuity and general coherence of the musical flow are supposed to follow are difficult to pinpoint.

In the following contribution, we present a simple automatic interactive improviser for non-idiomatic improvisation the development of which was premised on the following two principles. First it had to avoid any reliance on a pre-existing alphabet to accommodate more “textural” types of features. Second, rather than to prolong a given *context*, it would base its responses on an anticipation of the musical session’s current developments into the immediate future.

In the approach proposed here the system captures the playing of an improvising musician, segments it into phrases and stores these alongside a (possibly extendable) set of “textural” features. Basing its anticipation on this latter information using linear prediction, it then re-injects past phrases into the present in accordance with a (manually controlled) instruction as to “contrast” or “follow” on its prediction. The resulting system was then tested for coherence and reactivity in improvisation sessions with a professional saxophonist. Though it is far from bringing the principles on which its development was premised to a close, it proved a flexible environment from which a wide range of extensions and applications can be envisioned.

After a brief overview of some related work (section 2), the proposed system will be presented in details (section 3). This will be followed by the description and analysis of the tests that have been conducted (section 4) followed by some general comments concluding remarks (section 5).

2. Related work

Since Lewis’s first experiments with Voyager in the 1980’s [2] a number of digital systems have been proposed that can be considered as falling into the definition proposed in our introduction. Some, however, consider that “the interactive nature of a system can be obscure, mysterious and opaque” [4] and have preferred to shift efforts away from attempting to produce plausible imitations of interactive behavior towards more general criteria and the designing of solutions that draw from a variety of techniques ([3], [5]).

Amongst the systems that are concerned with such imitation, an important family has taken on from research conducted in musical style modeling and imitation that goes back to the earliest days of computer applications to music and the work of Hiller and Isaacson [6]. The algorithms developed in this context have attained high level of sophistication as attested, for instance, by [7] and [8]. The transposition to interactive improvisation has drawn from a variety of algorithmic resources. Pachet’s Continuator [9], for instance, uses probabilistic decision trees, while OMax by Assayag *et al.* [10] uses factor oracles. All of these “learn” either from a corpus or from the past of the ongoing

session to produce stylistically suitable continuations based on the most recent past of the improvisation (called the *context*).

Other types of approaches must also be mentioned. Some, such as Biles’s GenJam [11], use genetic algorithm. In this particular case, phrases of fixed length are generated in response to the author’s own playing. Neural networks have also been experimented with as exemplified by [12]. All these examples require lengthy, supervised, processes to train their underlying algorithms.

The notion of anticipation (or expectation) has integrated music theory and analysis through the pioneering work of Meyer [13] and known important developments ever since (see, for instance, [14]). It is also part of the psychology of improvisation behavior at least since Pressing introduced it into the field in [15]. More recently, it made its way into style analysis and imitation (see Conklin [7]) and into the development of automatic interactive improvisers, with a notable contribution provided by Cont *et al.* [17]. All the approaches involving the notion of anticipation the processes are based predominantly on structures extracted from pitch configurations.

The system described below can be considered as loosely inspired from the latter approach in that its responses are based on extrapolations of a current *context* into the future, rather than on its continuation *per se*. Though the implementation presented here does rely on pitch, it does not in principle depend on it as decisively. Indeed, it builds, for each incoming phrase, a vector containing features that describe its *overall* textural profile. The system extrapolates a possible continuation using a linear predictor (operating in a continuous space) and re-injects segments from the session’s past based on this prediction. This system is now described in more detail.

3. Proposed system

Our system articulates a total of four steps. The first is responsible for recording the audio stream incoming from (for the time being) one live improviser and for segmenting it into successive phrases s_t , $t = \{1, 2, 3, \dots, T\}$. The second calculates an acoustic feature vector x_t that maps s_t into a N -dimensional vector space \mathbb{R}^N . The third calculates a prediction of what the vector s_{T+1} might be. The fourth, finally, plays, at each trigger, one of the past phrases in accordance with instructions as to “contrast” or “follow” on its prediction. Each step is described in more details below.

3.1. Recording and Segmentation

Segments have been defined as audio slices having amplitude above a certain threshold and boundaries between two silent moments (i.e. with values below the threshold). This decision was based on the preliminary assumption that silences naturally articulate musical ideas and phrases. In order to capture points of segmentation, a simple threshold was set for a low-passed *RMS* function extracted from the incoming signal. The threshold value is determined to each performance by capturing the amplitude of the background noise from the recording environment and adding 6 to 9 dBs to it.

3.2. Mapping

The mapping module assigns a feature vector defining textural characteristics of the segments. To generate it, low-level features – pitch (with the octave divided into 24 micro-tones) and RMS – are extracted from the audio segments. These are calculated by using Sigmund, a standard pitch tracking algorithm in Pure Data [18]. For each segment, the chronologically ordered list of note durations is determined by considering the pitch curve

as a sequence of time intervals between changes of pitch. Finally a higher-level feature vector \mathbf{x}_t is generated that comprises:

1. The mean, standard deviation and slope (angular coefficient of the linear regression) estimated over time from RMS, pitch and duration curves;
2. The Pearson correlation between each two of the three feature curves;
3. The total duration of the audio segment;
4. A 24 chroma pitch-histogram, obtained by the pitch curve, defining the harmonic/intervallic characteristic of the segment.

3.3. Forecasting

The forecast module assumes that the feature vectors \mathbf{x}_t form a time series $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$. Standard forecasting procedures can be performed to estimate a value for the next vector \mathbf{x}_{T+1} . To obtain this we used a linear predictor, based on the equation:

$$\mathbf{x}_M = \sum_{k=1}^K \mathbf{a}_k \mathbf{x}_{M-k}, \quad (1)$$

where the prediction coefficient vectors \mathbf{a}_k are obtained by minimizing the average estimation error over the training data and K corresponds to the predictor order, set *a priori*. Order $K = 5$, performed well while allowing the system to become operational from the onset of the improvisation. The predictor was obtained using standard pseudo-inversion techniques and is re-computed with each new incoming phrase.

In another possible setting, used mainly for testing, the predicted vector \mathbf{x}_{T+1} is systematically taken to be the last completed phrase (\mathbf{x}_T). Transition between the two modes can be affected in real time if desired.

3.4. Output

At each (for the time being manually determined) trigger, the system re-injects a past phrase into the present of the improvisation. The selection process is based on a contrast factor γ . The system randomly selects a feature vector \mathbf{x}_k among the $\lfloor \gamma(T-1) + 1 \rfloor$ ones that are closest (in Euclidian terms) to the estimated \mathbf{x}_{T+1} . The past phrase re-injected is then s_k to which the feature vector s_k corresponds. In the context of the present work, the material was replayed literally or slightly transformed through timber distortion, but without impeding its recognition.

4. Preliminary testing

Once in place, the system was tested with a professional saxophone player in a series of improvisation sessions. The system was controlled by one of the authors of this paper. A total of six short sessions, as described in Table 1, were played and analyzed. The improviser was only informed on how the system works after the 3rd session was completed.

As hinted in Table 1, the system's parameters were kept fixed during each the first four sessions and varied freely during the last two. In the fourth session, the saxophonist was instructed to avoid providing the progression with a clear direction and to provide maximum contrast at all time. For the last two sessions no particular instruction were given. The first four and the last two are now considered in turn.

Piece	Description
1	Prediction is equated to last phrase, minimum contrast
2	Prediction based on forecast, maximum contrast
3	Prediction based on forecast, minimum contrast
Musician is informed on how the system works	
4	Prediction based on forecast, minimum contrast, saxophonist asked to intentionally avoid providing the progression with a clear direction
5	All parameters freely determined on-the-fly
6	All parameters freely determined on-the-fly

Table 1: Description of the pieces used for testing.

4.1. Leaving the parameters fixed

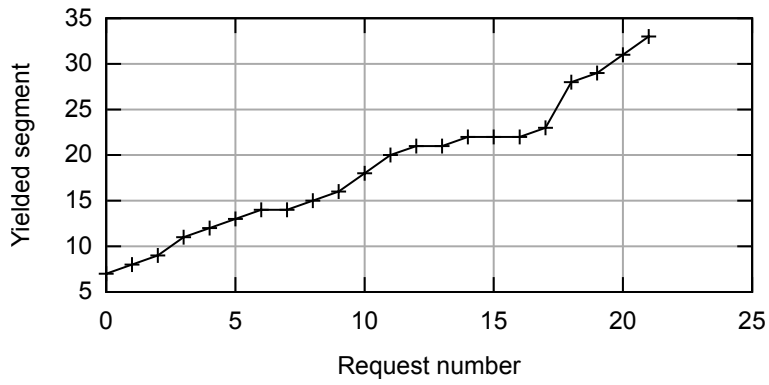
To observe the general behavior of the system the paths it took through past segments were plotted. The results are shown in Figure 1, whereby the y-axis represents to the (chronologically indexed) segments from the improviser and the x-axis represents to the (chronological) succession of triggers.

Figure 1a presents an entirely predictable behavior in which the last completed segment is systematically repeated. On the contrary, Figure 1b displays a highly erratic behavior, in which it is hard to detect a pattern. In both cases, the observed behaviors are precisely the ones that the settings used would have us expect.

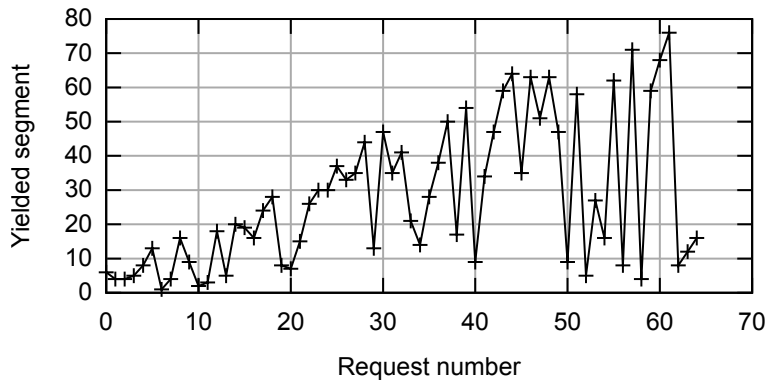
When the system is set to find a close match to the continuation it extrapolates, its behavior is neither erratic nor entirely predictable, as shown in Figure 1c and 1d. In most cases, the system tends to present re-configurations of segments taken from the relatively recent past of the improvisation but without displaying the same predictability observed in the first session. This type of behavior is dominant in the first half of Figure 1c and alternates in Figure 1d. This behavior could be traced back to the fact that the improviser tended to project similar or gradually evolving musical ideas over time spans that included several consecutive segments.

Under this same configuration, the system also retrieved segments from a more distant past. It did so, however, less often than in the second session. Also, it was observed that these segments presented certain coherence with the current musical context. Figure 2 illustrates this by showing spectrograms of the saxophonist's and of the system's outputs (see Table 1). This example corresponds to segments 27 and 28 in Figure 1c.

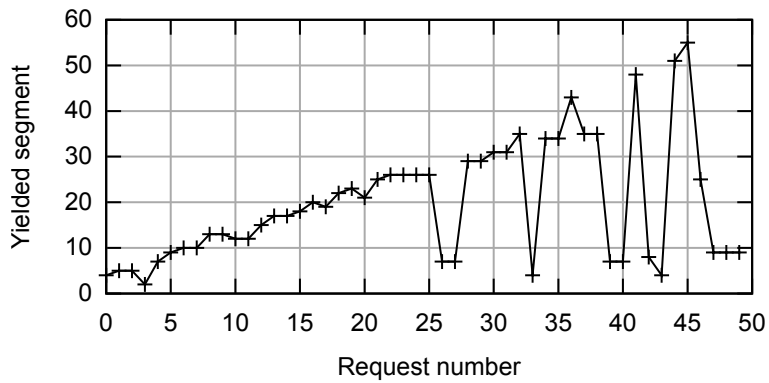
Two visually distinguishable spectral textures can be seen in both images. The first is formed by sustained notes played in the higher register while the second is formed by notes in the lower register repeated in rapid successions. The later is produced simultaneously by the saxophonist and by the system. The one played by the system, however, had been retrieved from a more distant past and corresponds to the 7th phrase played by the musician.



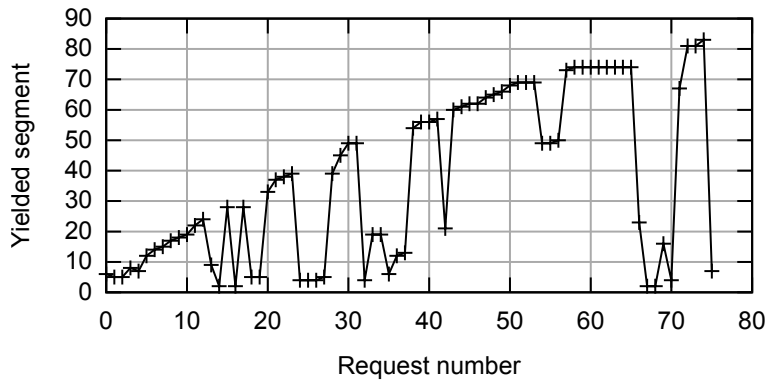
(a) Session 1 (last segment, follow)



(b) Session 2 (prediction, contrast)



(c) Session 3 (prediction, follow)



(d) Session 4 (musician creates contrast)

Figure 1: Re-injected segment for each request in sessions 1 to 4.

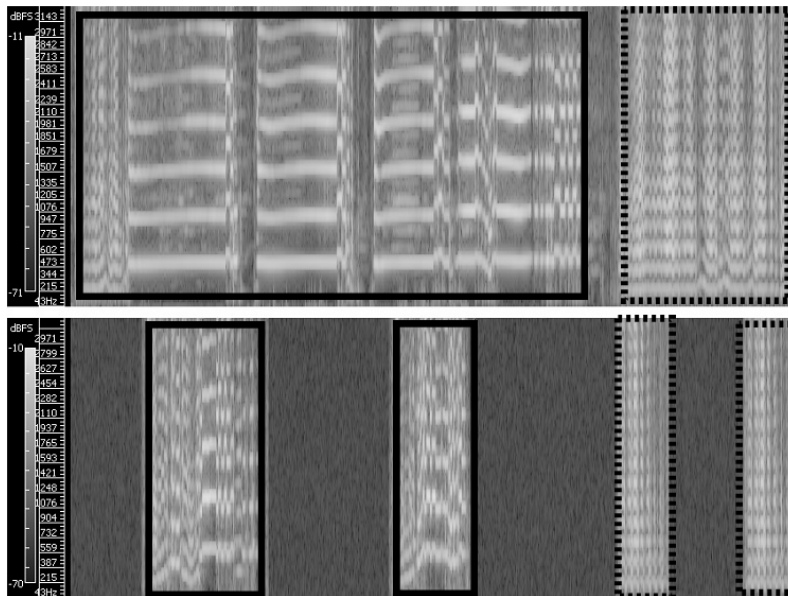
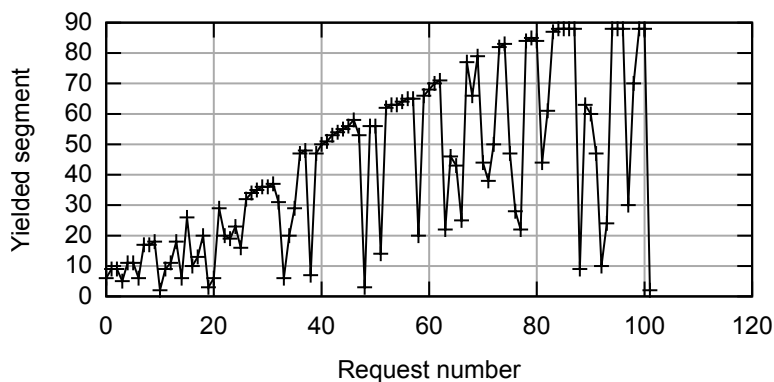


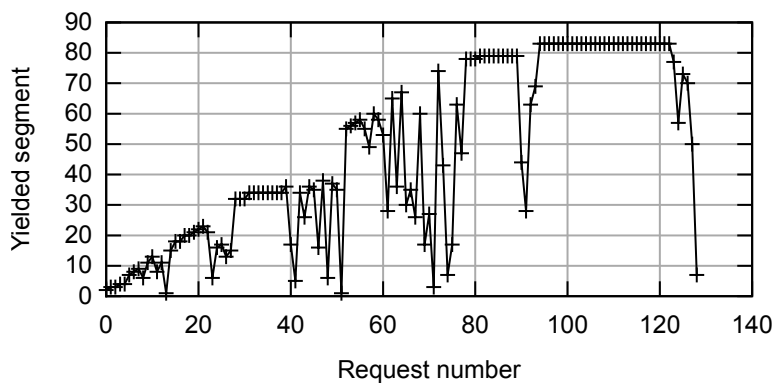
Figure 2: spectrogram

4.2. Changing parameters on-the-fly

Figure 3 illustrates the system's behavior in sessions 5 and 6, as described in Table 1. In these sessions, the system's parameters were changed on-the-fly by a human musician. It is possible to see some stable and unstable parts, combining the behaviors observed in Figure 1.



(a) Session 5



(b) Session 6

Figure 3: Re-injected segment for each request in free improvisation sessions.

In Figure 3a, the system's reactivity can be particularly clearly observed between 101

requests 26 and 45. The system presented a very stable behavior until request 30, when it started behaving more erratically. Figure 4 depicts both the re-injected segments and the instrument parameters during this excerpt. As can be seen, the contrast parameter was null and, raised at request 30, caused the system to select segments that are more distant from the current one. After that, the contrast parameter varied, maintaining a similar behavior. After request 40, when the contrast was set back to null, the system returned to stability.

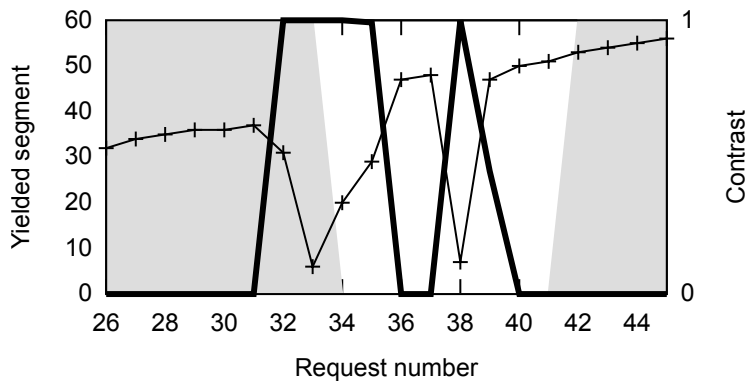


Figure 4: Re-injected segments (line with dots), contrast parameter (thick line) and reference for selection (gray background for prediction, white background for last sample).

A similar phenomenon can be observed in Figure 3b between requests 80 and 100. As shown in Figure 5, the unstable gap is created between two evidently stable parts by changing the value of the contrast parameter.

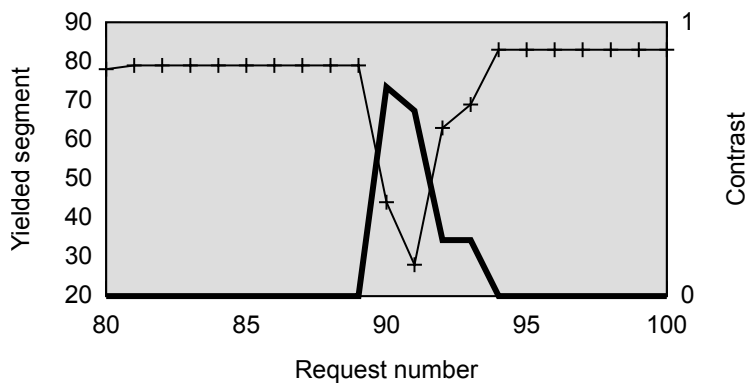


Figure 5: Re-injected segments (line with dots), contrast parameter (thick line) and reference for selection (gray background for prediction, white background for last sample).

The reactivity of the system can also be visualized in Figure 3b between requests 30 and 43. As shown in Figure 6, the user breaks the stability by raising the contrast level. During the stable part of this example, the parameters were set as in session 1.

These analyses show that the user can directly induce the system to respond in certain ways depending on his perception of the ongoing session. While the linear prediction seeks a coherent continuation and may reinforce a stable behavior for the interaction, the contrast factor can serve to re-inject musical ideas more removed from the present contexts, potentially inducing new directions to the musical progression.

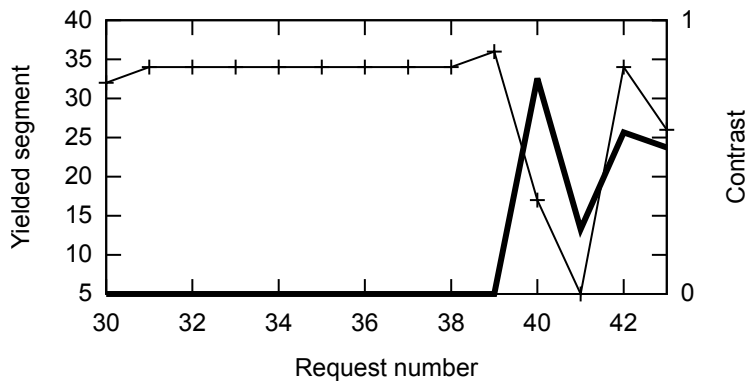


Figure 6: Re-injected segments (line with dots), contrast parameter (thick line) and reference for selection (gray background for prediction, white background for last sample).

5. Evaluation and Concluding remarks

The observations made during the tests just described showed the system to be both responsive and, in a certain narrow sense at least, coherent. An aspect that is much more difficult to translate into objective terms, but which was clearly present during these tests, is the playful quality the musical exchanges between the improviser and the person operating the system could be endowed with. Such observations were particularly encouraging when considered against the system's relative simplicity.

Concerning the two principles on which the development of the system was premised: the non-reliance on a fixed alphabet and the use of anticipation as a basis from which to determine outputs; both can be considered to having been met. The question remains open as to their true contribution to the quality of exchange just mentioned. As in any system that is supervised in real time, it is unclear to what extent success (or failure) is due to the human intervention or to the system's orderly behavior.

This problematic, whether answerable or not, directly suggests two directions in which future developments could be taken: that of a greater autonomy of the system and that of a more detailed and controlled anticipatory procedures. For the first, some principle determining the calling of new phrases (triggers) would have to be devised as well as criteria for the automatic selection of values for the contrast factor. For the second, the exploration of further sets of features, with the possibility to control their impact through weighting or through their partitioning into distinct sets could be envisioned. The same can be said of the mechanism underlying the anticipation procedure itself. Indeed, nothing impedes combining several, possibly concurrent, principle at the same time.

Each of these developments, from their prototyping to their eventual integration, can be undertaken in separate steps and accompanied by live testing.

Finally, yet another type of development can be considered. As the system's external controls are both simple and intuitive, setups can be envisioned in which the performer himself is in control through some form or another of augmented performance environment.

6. Acknowledgements

The authors would like to thank Prof. Rogério Costa, the saxophonist (and scholar of non-idiomatic improvisation) who has generously donated some of his time and lend his skills to the benefit of this project. Thanks also go to the institutions that have supported this research, FAPESP; CNPq and CAPES.

References

- [1] Pressing, J. (1998) "Psychological Constraints on Improvisational Expertise and Communication", in *In the Course of Performance: Studies in the World of Musical Improvisation*. Ed. Nettl B. and Russel, M. University of Chicago Press, Chicago, p. 53-74.
- [2] Lewis, G. (2000) "Too Many Notes: Computers, Complexity and Culture in VoyagerÓ", *Leonardo Music Journal*, v.10,p. 33-39.
- [3] Bown , O. (2011) "Experiments in Modular Design for the Creative Composition of Live AlgorithmsÓ. *Computer Music Journal*, 35:3, pp. 73-85.
- [4] Bown , O., Eldriedge A., McCormack, J.(2009) "Understanding Interaction in Contemporary Digital Music from Instruments to Behavioural ObjectsÓ. *Organised Sound*. 14:2, pp. 188-196.
- [5] McLean, A., Wiggins, G. A. (2010) "Bricolage Programming in the Creative ArtsÓ In: <http://yaxu.org/writing/ppig.pdf>. Accessed 15/10/2013.
- [6] Hiller, L.A, Isaacson. L. M. (1959) "Experimental Music: Composition with an Electronic ComputerÓ. McGraw-Hill Book Company, New York.
- [7] Conklin, D. and Witten, I. H. (1995-2002) "Multiple Viewpoint Systems for Music Prediction", *Journal of New Music Research*, Vol 24/1,1995, p. 51-73, (revised version 2002).
- [8] Cope, D. (2004) *Virtual Music: Computer Synthesis of Musical Style*, MIT Press.
- [9] Pachet, F. (2002) "The Continuator: Musical Interaction With Style", In *Proceedings of the ICMC*.
- [10] Assayag, G. and Bloch, G. and Chemillier, M. and Cont, A. and Dubnov, S. (2004), "Omax Brothers: a Dynamic Topology of Agents for Improvization Learning", In *Workshop on Audio and Music Computing for Multimedia*.
- [11] Biles, J. A. (1994) "GenJam: A Genetic Algorithm for Generating Jazz Solos", In *Proceedings of the ICMC*.
- [12] Franklin. J.A. (2004), "Predicting reinforcement of pitch sequences via lstm and tdÓ. In: *Proceedings of International Computer Music Conference (ICMA)*, Miami.
- [13] Meyer, Leonard B.(1956). "Emotion and Meaning in Music. Chicago University Press, Chicago.
- [14] Huron, D. (2006) *Sweet Anticipation: Music and the Psychologie of Expectation*. MIT Press, Cambridge.
- [15] Pressing, J. (1988) "Improvisation: Methods and Models" In, Sloboda, J. *Generative Processes in Music*, Clarendon, Oxford, pp. 129-178.
- [16] Conklin, D. (2003) "Music Generation from Statistical Models", In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, Aberystwyth, Wales,pp. 30-35.
- [17] Cont, A., Dubnov, S. and Assayag, G., (2007) "Anticipatory Model of Musical Style Imitation Using Collaborative and Competitive Reinforcement Learning", In Butz, M. V., Sigaud, O, Pezzulo, G, and Baldassarre, G., Eds *Anticipatory Behavior in Adaptive Learning Systems*, Springer-Verlag, Berlin, Heidelberg, p. 285–306.
- [18] Puckette, M. S. and Brown, J.C. (1998) "Accuracy of frequency estimates using the phase vocoder", In *Speech and Audio Processing, IEEE Transactions on*, v. 6:2,pp. 166-176.

Implementação Orientada a FPGA's de um Oscilador Digital Multiplexado

Mateus Henriques Negrelli, Márcio Brandão

Departamento de Ciência da Computação – Universidade de Brasília (UnB)
Brasília – DF – Brasil

negrelli@aluno.unb.br, brandao@unb.br

***Abstract.** This paper describes an FPGA-oriented implementation of a multiplexed digital oscillator. The multiplexation of oscillators is done through a pipeline of the tasks involved. The ultimate goal of this work is to provide an audio synthesis framework for an FPGA.*

***Resumo.** Este artigo descreve uma implementação orientada a FPGA's de um oscilador digital multiplexado. A multiplexação dos osciladores é feita usando um pipeline das tarefas necessárias. O objetivo final desse trabalho é fornecer um framework de síntese de áudio em FPGA's.*

1. Introdução

Aplicações de síntese de áudio digital, em sua grande maioria, valem-se de osciladores em maior ou menor escala. Seu funcionamento é simples e conhecido: a partir de um conjunto de entradas de frequência, amplitude e fase, um circuito contendo um ciclo de onda em memória gera como saída a forma periódica desejada [Roads 1996]. Alia-se a isso o fato demonstrado por Fourier [Lathi 1998, Spiegel 1974] de que qualquer sinal pode ser obtido a partir da combinação de ondas senoidais. Essa idéia é a base da técnica de síntese aditiva de áudio: um conjunto de osciladores senoidais operando em paralelo, com frequências diferentes e amplitudes variáveis no tempo, fornece uma maneira versátil e flexível de gerar formas de onda mais complexas [Jansen 1991]. A síntese aditiva não é a única aplicação desta idéia, claro, mas é uma das mais difundidas.

Contudo, há maneiras mais eficientes e elegantes de se obter o mesmo resultado. Snell (1991) propõe um oscilador multiplexado em hardware que pode gerar, em tempo real, um formato de onda complexo através da soma ponderada de um grande número de componentes senoidais. A proposta depende das velocidades de *clock* disponíveis em processadores, muito superiores às frequências padrão de amostragem de áudio: subdivide-se o ciclo de amostragem em segmentos, e em cada um deles obtém-se a amostra de áudio relativa a uma componente. Ao final, as várias amostras são somadas, gerando-se assim a onda composta.

Este trabalho apresenta a implementação de um oscilador multiplexado em FPGA, com capacidade de operar em tempo real, visando a criação de um *framework* para o desenvolvimento de aplicações envolvendo diferentes técnicas de síntese de áudio em dispositivos reconfiguráveis com poucos recursos.

1.1. Oscilador Simples

A idéia central de um oscilador simples baseado na técnica de busca em tabela é o envio para um conversor digital/analógico (DAC), a uma taxa de amostragem fixa $f_s (= 1/T)$, amostras de um ciclo de uma senóide. Para armazenar o ciclo de onda necessário, usa-se normalmente uma tabela circular de comprimento L (Figura 1).

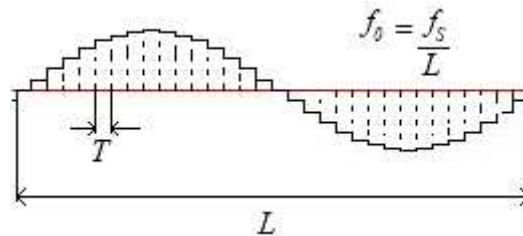


Figura 1. Ciclo amostrado de senóide.

Introduz-se aqui também a idéia do incremento I , o número de amostras percorridas na tabela a cada ciclo T . Caso o incremento de indexação da tabela seja dobrado, o novo sinal terá metade das amostras em um ciclo, correspondendo ao dobro da frequência. A equação abaixo mostra a expressão geral da frequência f_0 da senóide resultante em função do incremento I , do comprimento L da tabela do seno, e da frequência f_s de amostragem.

$$I = f_0 \cdot L / f_s$$

A Figura 2 apresenta a arquitetura básica de um oscilador simples baseado na técnica de busca em tabela. Verifica-se a presença de:

- uma memória com L palavras, contendo as amostras da senóide;
- um registrador armazenando o valor do incremento de ângulo de fase;
- outro registrador, para armazenar e atualizar o valor do ângulo de fase;
- um somador que completa o acumulador de ângulo de fase;
- um DAC que conecta o oscilador ao resto do circuito.

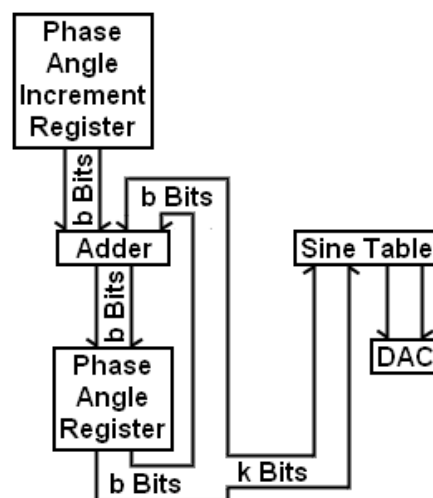


Figura 2. Estrutura geral do oscilador simples.

Vale notar que o registrador de incremento atende à representação de números reais em ponto fixo, com seus k bits mais significativos representando a parte inteira. O endereço de memória, portanto, passa por um processo de truncamento dos $b - k$ bits correspondentes à parte fracionária do registrador de incremento. Com a escolha de um comprimento adequado para a tabela de seno, é possível se obter uma relação sinal-ruído adequada para sinais de áudio, conforme descrito em Moore (1991).

1.2. Pipeline

Na arquitetura de computadores, a técnica de *pipeline* é fundamental para tornar processadores rápidos, e sua essência se baseia em aumentar o *throughput* do sistema através do funcionamento em paralelo de suas partes, geralmente denominadas estágios [Patterson e Hennessy, 2005].

Num processador sem *pipeline*, quando uma instrução passa de um estágio para o seguinte, os estágios anteriores ficam inativos até a finalização da instrução atual, quando a próxima finalmente entra no início do caminho de dados. Num processador com *pipeline*, contudo, ocorre a sobreposição de múltiplas instruções na execução. Isso quer dizer que, assim que uma instrução passa do primeiro estágio para o segundo, a instrução seguinte começa a ser processada.

Observa-se, então, o chamado paradoxo do *pipelining*: o tempo de execução total de cada instrução não se altera nesse novo paradigma. O aumento de velocidade é gerado pela paralelização das etapas, e está associado ao aumento de *throughput* mencionado acima.

Na Figura 3, vê-se a execução de três instruções num sistema com *pipelining* implementado. Um programa que normalmente levaria 9 ciclos de relógio para ser executado agora leva apenas 5. Assumindo-se uma duração semelhante para os estágios, o processador com *pipeline* seria três vezes mais rápido que o padrão caso o número de instruções tendesse a infinito.

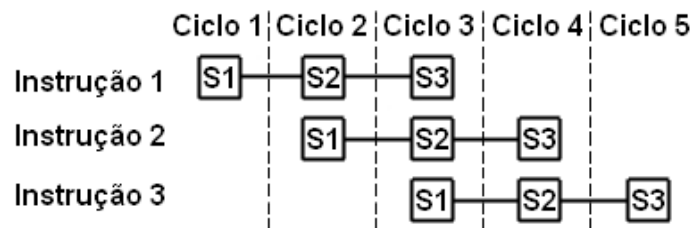


Figura 3. Exemplo de fluxo de instruções num *pipeline* de três estágios.

2. Oscilador Multiplexado

A Figura 4 apresenta a arquitetura do oscilador multiplexado em FPGA, baseada na implementação em hardware apresentada por Snell (1991).

Como mencionado na Seção 1, a base dessa proposta é a relativamente baixa taxa de amostragem de áudio quando comparada com a frequência dos *clocks* de processamento computacional atualmente disponíveis: durante um único período de amostragem de áudio, podem-se realizar diversas operações de período curto, sincronizadas com o que chamamos de *clock* de componente. No caso do oscilador multiplexado, essas operações envolvem o cálculo das amostras associadas às várias

componentes senoidais de uma forma de onda complexa. Desse modo, faz-se necessário apenas um dispositivo para realizar uma tarefa normalmente associada a múltiplos osciladores simples.

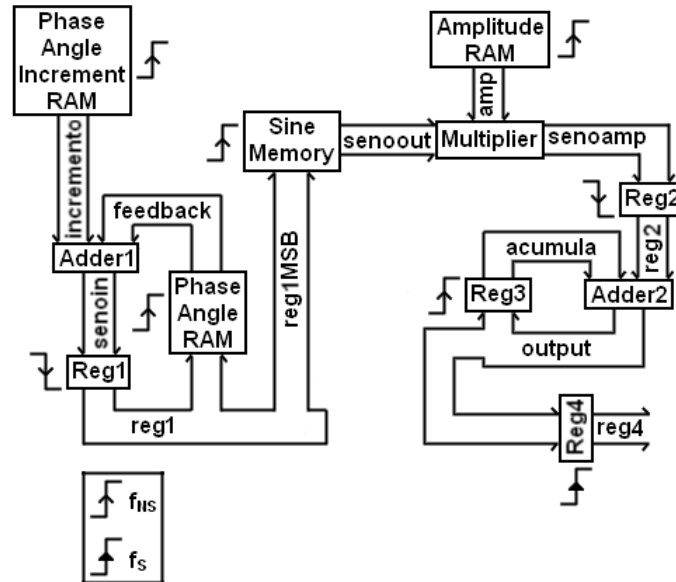


Figura 4. Estrutura geral do oscilador multiplexado.

O circuito é controlado por dois *clocks* principais (ver Seção 3.5), um associado à amostragem (ClkSamp, frequência f_s) e o outro ao cálculo de componentes senoidais (ClkComp, frequência f_{NS}). A taxa f_{NS} é o produto da frequência f_s pelo número n de componentes (Figura 5).

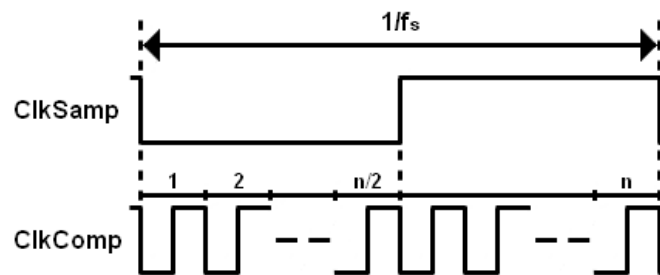


Figura 5. Comparação entre as frequências f_s e f_{NS} .

O processo é dividido num *pipeline* de três estágios, separados pelos registradores Reg1 e Reg2. O primeiro estágio é a soma progressiva, para cada componente senoidal, de incrementos cujo valor base depende da frequência desejada ($I = f_0 \cdot L / f_s$). O segundo estágio usa os valores de incremento calculados como endereços para acessar um ciclo de senoide em memória. A amplitude de cada componente senoidal é determinada, a cada período de amostragem, pelo multiplicador localizado imediatamente antes de Reg2, em conjunto com os valores internos da Amplitude RAM. O terceiro estágio, por fim, acumula as amostras das várias componentes em Reg3, preparando a saída para o *clock* de amostragem.

2.1. Memórias

A principal diferença entre o oscilador simples e o multiplexado é a presença, neste, de três memórias a mais. Num único ciclo de componente, todas as componentes são processadas uma vez, e seus valores associados devem permanecer armazenados para serem utilizados nos cálculos um período mais tarde. Entram em ação aí a Phase Angle Increment RAM, a Phase Angle RAM e a Amplitude RAM, memórias com número de palavras igual ao número de componentes senoidais, cujo valor máximo é dependente dos gargalos do circuito (ver Seção 2.4).

A quarta memória, Sine Memory, por outro lado, tem o mesmo papel que no oscilador simples, que é o de armazenar amostras de um ciclo completo de uma senóide para permitir a busca em tabela.

2.2. Endereçamento

Um circuito de endereçamento faz com que as três memórias passem sequencialmente por todas as suas palavras uma vez por ciclo de amostragem, sendo atualizados todos os valores no oscilador. A memória do seno é endereçada no circuito principal pelos bits mais significativos dos valores de ângulo de fase das componentes.

2.3. Registradores

Há registradores com duas funções principais no oscilador multiplexado. A primeira delas é, como mencionado anteriormente, a divisão do circuito em três estágios de *pipeline*. Isso se dá de maneira simples: os registradores Reg1 e Reg2 armazenam valores em um ciclo, passando-os adiante quando os três estágios estão prontos para processar a próxima componente. Reg4 está no final do terceiro estágio, o último do *pipeline*, e apenas trabalha como porta de saída do circuito.

A segunda função é a construção de acumuladores. As saídas de Reg1 e Reg3 são barramentos que realimentam um ponto anterior do oscilador, no intuito de manter somas atualizadas. No caso de Reg1, que desempenha simultaneamente as duas funções principais descritas, há a acumulação dos valores de ângulo de fase para cada componente senoidal, que posteriormente endereçam a Sine Memory. Reg5 acumula os valores finais de cada componente, gerando-se ali o valor final da amostra em si para cada período de amostragem.

2.4. Componentes Adicionais

Há somadores simples nos acumuladores, que produzem, a todo ciclo de componente, a adição do valor acumulado com o novo valor recebido para a componente atual. Há também um multiplicador que altera a saída da memória de seno por um fator de amplitude, tendo cada componente senoidal seu próprio fator armazenado na Amplitude RAM. Esse processo de multiplicação é o gargalo teórico do circuito, ou seja, a velocidade de funcionamento do oscilador é mais severamente limitada pelo segundo estágio do *pipeline*.

3. Implementação

Para implementar o oscilador multiplexado em FPGA, utilizou-se a versão 9.1 do programa Quartus II e o kit lógico DE2-70 [Altera, 2013]. Os componentes necessários

foram desenvolvidos com uma combinação de diagramas de blocos e programação VHDL [Mealy e Tappero, 2013].

Para esta implementação específica, escolheu-se construir o oscilador com capacidade para lidar com 384 componentes senoidais simultâneas, operando a uma frequência de amostragem de 48 kHz, o que faz com que o *clock* de componente tenha que operar a 18,432 MHz.

Cada uma das partes constituintes do oscilador descritas na Seção 2 foi implementada em separado e testada isoladamente com respeito à sua resposta temporal. Desse modo, foi possível verificar e adequar os atrasos individuais de cada bloco para a operação de *pipeline*.

3.1. Memórias

A Sine Memory foi construída com uma simples memória ROM de 1024 palavras, contendo um ciclo completo de senoíde. Esse comprimento de tabela possibilita a obtenção de uma relação sinal-ruído adequada para sinais de áudio [Moore, 1991].

Phase Angle Increment RAM, Phase Angle RAM e Amplitude RAM são memórias RAM comuns, com escrita e leitura, e número de palavras igual ao número máximo de componentes escolhido.

3.2. Endereçamento

O circuito de endereçamento utiliza um contador que, durante um ciclo de amostragem, conta de 0 a 383, para que as três memórias acessem os valores adequados a cada ciclo de componente. Ao se processar a última componente, a contagem é reiniciada.

3.3. Registradores

Com exceção de Reg3, que possui também uma entrada de *clear* assíncrono (explicada na Seção 3.5), todos os registradores trabalham com os pinos básicos de *clock* e entrada/saída de dados.

Além dos registradores mencionados previamente, a implementação prática do *pipeline* demandou um registrador auxiliar para atrasar em um ciclo de componente o valor de endereçamento. Esta medida se justifica pois, em um dado ciclo de componente n , o estágio 1 do *pipeline* processa a componente senoidal. Enquanto isso, o estágio 2 ainda processa a componente $n-1$, devendo lidar com a componente n somente no ciclo $n+1$. Dessa maneira o registrador auxiliar, para endereçar a Amplitude RAM, é endereçado pelo valor do ciclo de componente $n-1$.

3.4. Componentes Adicionais

Os somadores, por apresentarem um pequeno tempo de resposta quando comparado ao do multiplicador, são considerados componentes pouco críticos, e implementados com somadores paralelos padrão. Já para a implementação do multiplicador, levou-se em conta que multiplicadores construídos com unidades lógicas de uma FPGA podem ser bastante lentos se comparados a circuitos dedicados de multiplicação, estreitando o gargalo teórico mencionado anteriormente. Essa alternativa é disponibilizada em alguns kits lógicos de FPGA's, e é a que se escolheu usar nesta implementação.

3.5. Esquema de Clock

Em implementações práticas de circuitos que devem funcionar em tempo real, um grande problema que surge ao se sair do âmbito teórico ideal são as restrições temporais dos tempos de *setup*, atrasos de processamento e necessidades de sincronização. O circuito do oscilador multiplexado não é exceção, exigindo cuidado em todos esses aspectos.

Primeiro, e mais evidente, o ciclo de componente deve ser longo o suficiente para comportar os atrasos de processamento de todos os blocos. Como mencionado anteriormente, todos os blocos foram testados individualmente, podendo-se tirar conclusões rápidas sobre a viabilidade do número de componentes senoidais escolhido.

Além disso, devem ser levados em consideração os tempos de *setup* dos vários blocos, principalmente das memórias. Como os registradores entre estágios de *pipeline* devem passar adiante os valores finais de cálculo, não os iniciais, não é possível usar um *clock* unificado para todos os componentes. Optou-se, então, por dividir as tarefas do oscilador entre as bordas de subida e descida do *clock* de componente: na borda de descida são ativados os registradores 1, 2 e auxiliar, além do circuito de endereçamento. Desse modo, as entradas de cada estágio estão prontas na borda de subida do relógio.

Por fim, o processo de amostragem em si gerou a necessidade de se criar um novo bloco, chamado Amostra Enable. O oscilador multiplexado não precisa receber como entrada um relógio na taxa de amostragem, apenas o *clock* de componente. Assumindo-se que a frequência deste foi definida corretamente, o intervalo de tempo entre dois cálculos da mesma componente senoidal tem o valor $T = 1/f_s$. Segue-se desse raciocínio que, para corretamente amostrar a onda complexa, basta se extrair o valor no terceiro estágio do *pipeline* assim que for processada a última componente senoidal. A função de Amostra Enable é detectar a presença desta componente no último estágio do *pipeline* e, quando os cálculos forem concluídos, ativar a passagem de dados por Reg4, simultaneamente zerando o valor de Reg3. Em seguida, reinicia-se o ciclo de processamento para gerar a próxima amostra de áudio.

4. Sincronização com o Kit Lógico

Com a estrutura do oscilador multiplexado estabelecida, o próximo passo na implementação do *framework* de síntese foi a construção da interface entre o oscilador e o codec de áudio da FPGA, o WM8731. A DE2-70, em seus manuais, já fornece alguns blocos pré-prontos que facilitam esse processo, ficando a cargo do implementador a geração dos *clocks* necessários, e do canal serial de passagem de dados.

4.1. Clock de Entrada

A DE2-70 possui dois *clocks* internos, um de 50 MHz e outro de 28 MHz. Os blocos pré-prontos mencionados acima exigem um *clock* de 18,432 MHz, assim como o oscilador multiplexado. Desse modo, escolheu-se utilizar em sua entrada uma PLL para criar, a partir do *clock* de 50 MHz, um *clock* com o valor adequado.

4.2. Serialização de Dados

O modo de funcionamento padrão do codec da DE2-70 é o chamado *Left Justified*, ilustrado na Figura 6. Os valores das amostras de áudio são passados serialmente, do MSB para o LSB, divididos em dois canais. O *Left Right Clock* (LRC) separa os canais: uma borda de subida indica o começo do canal esquerdo, e uma borda de descida, o começo do canal direito. O *Bit Clock* (BCLK) sinaliza a passagem de dados do oscilador para o WM8731: uma borda de subida indica que o codec deve receber o valor atual do canal de dados (DAT).

Vale notar que o LRC tem frequência f_s , de 48 kHz. Assim, BCLK tem frequência igual a n vezes f_s , onde n é o número de bits nas palavras que representam as amostras de áudio. Os valores padrão para n são 16, 20 ou 24, todos abaixo da ordem de grandeza do número de componentes senoidais no oscilador multiplexado. Isso significa que o BCLK não é um gargalo do circuito, com bastante tempo disponível para a realização de suas operações associadas.

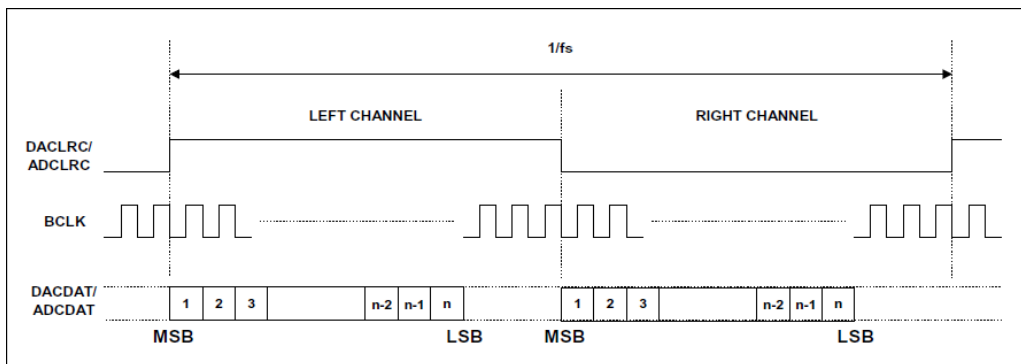


Figura 6. Left Justified Mode do codec WM8731.

O oscilador multiplexado, em todos os seus estágios, trabalha com dados paralelos. Para possibilitar a interação com o codec, adicionou-se um novo registrador após Reg4, que recebe sua saída e serializa o valor obtido, gerando DAT. Esse registrador opera com *Bit Clock* negado, de modo que a relação entre BCLK e DAT aconteça como na Figura 6. Além disso, ele usa um sinal de *load*, descrito na seção seguinte, para carregar o valor paralelo a ser serializado.

4.3. Bloco de Sincronização

Para cuidar da geração síncrona dos *clocks* associados ao Left Justified Mode, construiu-se um novo bloco. Ele recebe como entrada a saída da PLL, de frequência 18,432 MHz e, contando os ciclos desse relógio, gera as bordas de subida e descida para LRC e BCLK.

Adicionalmente, o bloco de sincronização usa o contador associado a LRC para prever as trocas de nível e enviar adequadamente o sinal de *load* para o registrador de serialização. Sendo assim, a interface entre o oscilador multiplexado e o WM8731 é realizada de maneira ordenada, sem risco de offsets temporais que comprometam o funcionamento do circuito completo.

5. Normalização

Com a estrutura descrita até aqui, já é teoricamente possível gerar sons na FPGA. Contudo, propõe-se ainda um novo ponto de preocupação: cada componente senoidal possuirá uma amplitude máxima, proveniente da quantidade de bits nas saídas das memórias de seno e amplitude. Esse máximo individual de cada componente será apenas 1/384 do valor total com o qual o oscilador está preparado para lidar.

Em outras palavras, o problema é que, para entradas com poucas componentes, a saída terá amplitudes reduzidas. A solução dada foi substituir-se Reg4 por um bloco de normalização, que permite ao usuário amplificar o sinal acumulado.

Esse bloco funciona do mesmo modo que Reg4, com a adição de um fator de *shift*. Sua função é realizar *shift lefts* aritméticos sobre o valor final da amostra a cada ciclo, de maneira que somas de poucas componentes possam aproveitar melhor a faixa dinâmica disponibilizada. O usuário pode escolher a execução de zero a nove *shifts*, dependendo de ele estar usando 384 componentes, 256, 128, 64, e assim por diante.

É importante notar que cabe ao próprio usuário saber quantas componentes estão em uso, e a escolha correspondente do número de *shifts*. O mau uso dessa função, com excesso de movimentações à esquerda, pode corromper os valores finais das amostras, deteriorando a relação sinal-ruído do sinal de áudio.

6. Testes Iniciais

Após as considerações feitas até agora, decidiu-se confirmar na prática os resultados previstos, com um teste simples que utilizou um número reduzido de componentes senoidais correspondentes a notas da escala de Dó maior. Estas podiam ser acionadas por meio de chaves do kit lógico, de modo que se pudesse verificar as frequências das diversas senóides, os intervalos musicais entre elas, e a ação do mecanismo de normalização.

Instanciado o circuito em modo de testes na FPGA, os resultados verificados se encaixaram no esperado: o kit lógico produziu som com sucesso, e confirmou-se auditivamente que as notas foram geradas de maneira correta, tanto melódica quanto harmonicamente. O processo de normalização também desempenhou seu papel como previsto, permitindo controlar a amplitude do sinal de saída satisfatoriamente para diversas configurações de componentes.

7. Discussão

A intenção inicial desse trabalho, como mencionado na Seção 1, era implementar uma ferramenta capaz de operar em tempo real, processando formas de onda elaboradas para síntese de áudio em dispositivos com capacidade de processamento restrita.

O primeiro ponto a ser mencionado é a operação em tempo real, verificada com sucesso. O oscilador multiplexado atualiza e acumula valores constantemente a partir de suas memórias, calculando-se novas amostras a cada ciclo, sendo que o usuário pode interferir ativamente na geração desses resultados.

Em seguida, fala-se sobre o aspecto da complexidade das formas de onda passíveis de geração. Como descrito na Seção 6, até agora testes práticos foram

realizados com no máximo oito componentes senoidais. Todavia, foi verificada numericamente, via simulação no Quartus, a capacidade de paralelização de 384 componentes do oscilador multiplexado. Os trabalhos futuros mencionados na Seção 8 serão essenciais para a confirmação auditiva dos resultados de dezenas, ou até centenas de componentes sendo processadas simultaneamente, com valores de incremento e amplitude cuidadosamente calculados para gerar timbres elaborados.

Consideram-se, assim, alcançados os objetivos iniciais propostos. No entanto, avaliamos que o sistema atual ainda pode receber melhorias, tornando-se mais versátil, e capaz de executar operações mais complexas. Basicamente, a estrutura do oscilador multiplexado ainda deve ser atualizada antes de poder ser usada como o cerne de aplicações variadas de síntese de áudio.

8. Trabalhos Futuros

O próximo passo nesse trabalho é integrar o oscilador multiplexado em um sistema completo de síntese aditiva na placa DE2-70. Os componentes a serem desenvolvidos são, a princípio, geradores de envoltória tanto de amplitude quanto de frequência, e um sistema de reescrita em tempo real que permita atualizar a Phase Angle Increment RAM e a Amplitude RAM com os valores das envoltórias associadas a cada componente. Além disso, deve ser implementada alguma forma de interface com o usuário.

Os próximos esforços serão voltados à implementação dos geradores de envoltórias de amplitude com reescrita da Amplitude RAM em tempo real. Esses avanços poderão ser usados para um funcionamento análogo com as envoltórias de frequência e a Phase Angle Increment RAM. Quanto à interface, um desenvolvimento natural seria o controle do sistema via dispositivos MIDI. Além disso, o *framework* do oscilador multiplexado pode ser aplicado a outras técnicas de síntese além da aditiva.

Referências

- Altera, DE2-70 Development and Education Board. Disponível em <http://www.altera.com/education/univ/materials/boards/de2-70/unv-de2-70-board.html>. Julho, 2013.
- Altera, Quartus II Web Edition Software. Disponível em <http://www.altera.com/products/software/quartus-ii/web-edition/qts-we-index.html>. Julho, 2013.
- Jansen, C., “Sine Circuitu: 10,000 high quality sine waves without detours”, Proceedings of the ICMC, 1991, pp. 222-225.
- Lathi, B. P., “Signal processing and linear systems”, Berkeley-Cambridge Press, 1998.
- Mealy, B. and Tappero, F., “Free Range VHDL”. Disponível em http://www.freerangefactory.org/dl/free_range_vhdl.pdf. Julho, 2013.
- Moore, F. R., “Table lookup noise for sinusoidal digital oscillators”, Foundations of Computer Music, MIT Press, Cambridge, 1991, pp. 326-334.
- Patterson, D. A. and Hennessy, J. L., Computer Organization and Design, 3rd ed., Elsevier, 2005, pp. 367-453.
- Roads, C., “Computer Music Tutorial”, MIT Press, Cambridge, 1996.

Snell, J., "Design of a digital oscillator that will generate up to 256 low-distortion sine waves in real time", *Foundations of Computer Music*, MIT Press, Cambridge, 1991, pp. 289-325.

Spiegel, M. R., "Schaum's outline of Fourier analysis with applications to boundary value problems", McGraw-Hill, 1974.

Detecção de Movimentos Auxiliares de Clarinetistas Através de Visão Computacional

Rodrigo A. Seger^{1*}, Alessandro L. Koerich^{1,2}, Marcelo M. Wanderley³

¹Departamento de Engenharia Elétrica
Universidade Federal do Paraná
81531-970 Curitiba, PR, Brasil

rodrigo.seger@gmail.com, alessandro.koerich@ufpr.br

²Programa de Pós-Graduação em Informática
Pontifícia Universidade Católica do Paraná
80215-901 Curitiba, PR, Brasil

alekoe@ppgia.pucpr.br

³Centre for Interdisciplinary Research in Music Media and Technology
Schulich School of Music
McGill University
H3A 1E3 Montreal, QC, Canada

marcelo.wanderley@mcgill.ca

Abstract. *In this paper we discuss the use of computer vision techniques as an alternative method for the detection of ancillary movements produced by clarinetists during performances. Ancillary gestures are movements spontaneously made by musicians and do not produce sound, but they help in the creation of music. In order to optimize the current analysis procedures, we propose the use of computer vision algorithms to detect movements of clarinet, knees and backs, three characteristic ancillary gestures of these musicians. The techniques were applied to nine videos of clarinetists in a controlled studio and the three proposed gestures were detected successfully.*

Resumo. *Neste artigo discute-se o uso de técnicas de visão computacional como um método alternativo para a detecção de movimentos auxiliares produzidos por clarinetistas durante apresentações musicais. Movimentos auxiliares são gestos produzidos espontaneamente pelos instrumentistas e que não geram som, mas que acompanham a execução musical. Para otimizar os atuais procedimentos de análise propôs-se o uso de algoritmos de visão computacional para detectarem movimentos de clarinete, joelhos e costas, três gestos auxiliares característicos destes músicos e que ocorrem com frequência. As técnicas foram aplicadas em nove vídeos de clarinetistas em ambiente controlado e os três movimentos foram detectados com sucesso.*

1. Introdução

Há cerca de 30 anos estuda-se como o ser humano se interage com outras pessoas e com máquinas. Esta interação advém basicamente da fala e da produção de ges-

*Esta pesquisa contou com suporte parcial da CAPES/Fulbright processo BEX1770/712-9, da Fundação Araucária, processo 203/12 e do CNPq processos 306.703/2010-6 e 472.238/2011-6.

tos. Pesquisas em áreas como engenharia, computação, música e psicologia procuram compreender qual o significado real de cada gesto produzido e como reproduzi-lo ou interpretá-lo [Aggarwal and Cai, 1999] [Gavrila, 1999] [Mitra and Acharya, 2007] [Dahl and Friberg, 2004] [Davidson, 1993] [Jensenius et al., 2010].

Centrado no estudo dos movimentos executados por instrumentistas, estes se dividem em três grandes grupos: os gestos efetivos, os gestos de acompanhamento e os gestos meramente figurativos [Cadoz, 1998]. Em uma analogia com um clarinetista, os gestos efetivos são os movimentos que efetivamente produzem som, como, no caso, os dedos pressionando as teclas do clarinete. Os gestos de acompanhamento, também conhecidos como gestos auxiliares, espontâneos ou não óbvios [Wanderley, 1999], são os movimentos que acompanham o processo de produção musical, como movimentos dos ombros e da cabeça. Por fim, os gestos figurativos são todos aqueles que são percebidos pelo som em si, e não pelos movimentos - normalmente gerados por mudanças na melodia ou da entonação.

Os movimentos auxiliares, justamente por não possuírem significado lógico ou motivo de existência, são o principal foco de estudos atuais [Wanderley, 1999] [Wanderley, 2002]. No universo dos clarinetistas, estes gestos são basicamente o movimento completo cima/baixo do clarinete, movimento circular completo do clarinete, movimento cima/baixo de cabeça, movimento cima/baixo de ombros, movimento de ondulação das costas, movimento de abertura/fechamento de braços ("bater asas"), movimento de inclinação da cintura, movimento de dobra de joelhos, movimento de inclinação e batida de pés, deslocamento de peso (balanço do corpo) para a esquerda ou para a direita [Wanderley et al., 2005]. Dentre esta listagem, alguns movimentos aparecem com mais frequência. É o caso dos gestos de cima/baixo do clarinete, ondulação das costas, dobra dos joelhos e deslocamento de peso. Estes são movimentos com alto grau de distinção entre si e amplamente utilizados em estudos desta natureza [Wanderley, 2002] [Wanderley et al., 2005].

Atualmente, a avaliação destes movimentos é feita através de observações de vídeos gravados em ambientes controlados e da análise de informações obtidas via sensores de movimento presos aos músicos. Estes dados normalmente são de grande quantidade, onerando o processo de análise, ao passo que as observações dos vídeos são entediadas e subjetivas (a detecção de movimentos é extremamente dependente do avaliador) [Wanderley, 2002] [Wanderley et al., 2005] [Teixeira, 2010] [Verfaillie et al., 2006].

Baseado nesta problemática que se propôs o uso de técnicas de visão computacional aplicadas a vídeos de clarinetistas para criar uma rotulação para os mesmos, ou seja, indicar - via esforço computacional - instantes nos vídeos onde há a execução dos movimentos auxiliares de cima/baixo do clarinete, ondulação das costas e dobra dos joelhos. Dessa forma, seria possível substituir as observações dos vídeos e poder concentrar os posteriores estudos de dados dos marcadores apenas nos intervalos sugeridos pelo algoritmo, otimizando o processo.

A base de dados de estudo é uma compilação de nove vídeos de quatro clarinetistas executando um trecho de cerca de 1 minuto do Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo, fornecidos pelo *Input Devices and Music Interaction Laboratory* do Setor de Tecnologia Musical da Universidade McGill, Canadá. Os vídeos foram gerados por uma simples câmera fixa, em um estúdio com iluminação constante.

Este artigo é composto inicialmente por uma breve contextualização e os objetivos da pesquisa, já mencionados. Dando prosseguimento, será apresentado o método de trabalho proposto, com os resultados obtidos e suas respectivas avaliações. Por fim, as conclusões obtidas e todas as referências utilizadas.

2. Método Proposto

Para poder analisar a base de vídeos e a partir destes mensurar os três movimentos não óbvios neles contidos, foi criado um método composto por três partes fundamentais: pré-processamento dos vídeos, detecção de movimentos e avaliação dos resultados.

Qualquer detecção de movimentos a partir dos vídeos originais é muito complexa. Por isso, inicialmente fez-se necessário adequar a base para, então, viabilizar a detecção de movimentos. Esta etapa foi denominada pré-processamento e tem como principal objetivo segmentar o músico, ou seja, eliminar a influência do cenário no vídeo e corrigir problemas de iluminação durante as filmagens.

A técnica de visão computacional escolhida para a eliminação do cenário foi a de subtração de fundo [Gonzalez and Woods, 2007]. Este procedimento foi escolhido à iluminação ser constante - não há mudança brusca de quantidade de luz - e o ambiente ser estático, porque o cenário é fixo, apenas o músico se desloca, enquanto que a câmera e o cenário em si permanecem imóveis.

Para a utilização da técnica, faz-se necessário o armazenamento de um quadro de referência que será subtraído dos quadros em análise. Como o intuito é eliminar o cenário, ou seja, tudo além do músico, o quadro do vídeo contendo apenas o ambiente, sem o músico, é a referência ideal. O método aplicado encontra-se resumido na Figura 1.

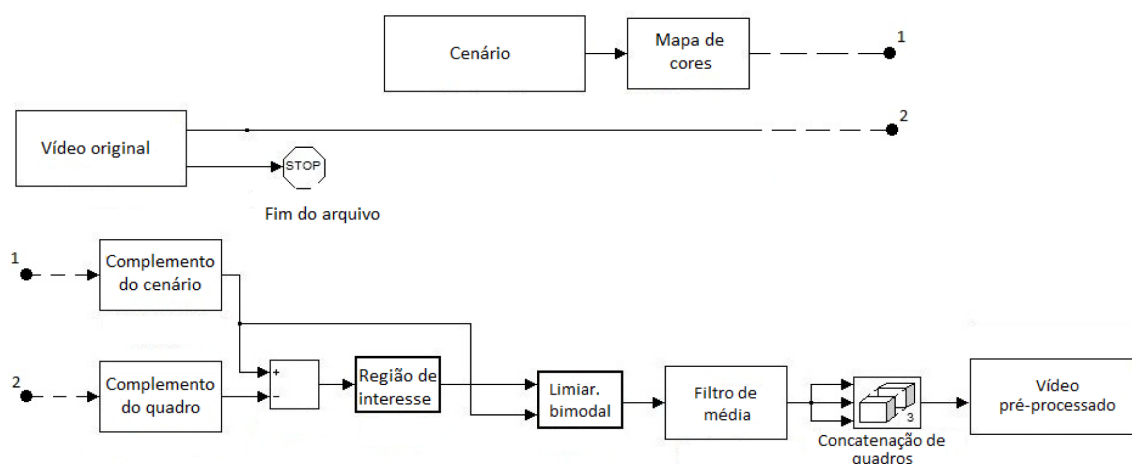


Figura 1: Etapas do pré-processamento.

A Figura 2 ilustra o pré-processamento aplicado. Primeiramente suprimiram-se as informações de cores transformando os vídeos em níveis de cinza com oito bits de profundidade (Figura 2c). Em seguida, fez-se a subtração do quadro-referência de cenário (Figura 2b) de todos os demais quadros dos vídeos (Figura 2d). A próxima etapa foi delimitar a região de interesse (Figura 2e). Como o músico passa quase que a totalidade de duração da gravação na região central, dividiu-se o vídeo em quatro partes iguais (no sentido da largura da imagem), e eliminaram-se as duas extremidades. Realizado o corte, aplicou-se uma limiarização da imagem, ou seja, através de um simples algoritmo condicional, níveis medianos de cinza foram reduzidos para preto e os demais elevados a branco, tornando a imagem binária: o músico e seu clarinete em preto e o cenário inteiramente branco. Um filtro de média foi ainda utilizado no intuito de suavizar os contornos da imagem resultante (Figura 2f). A Figura 3 ilustra o resultado obtido para vídeos dos quatro clarinetistas.

Os três movimentos espontâneos propostos tiveram procedimentos particulares para sua detecção. O movimento auxiliar de cima/baixo de clarinete foi o mais evidente

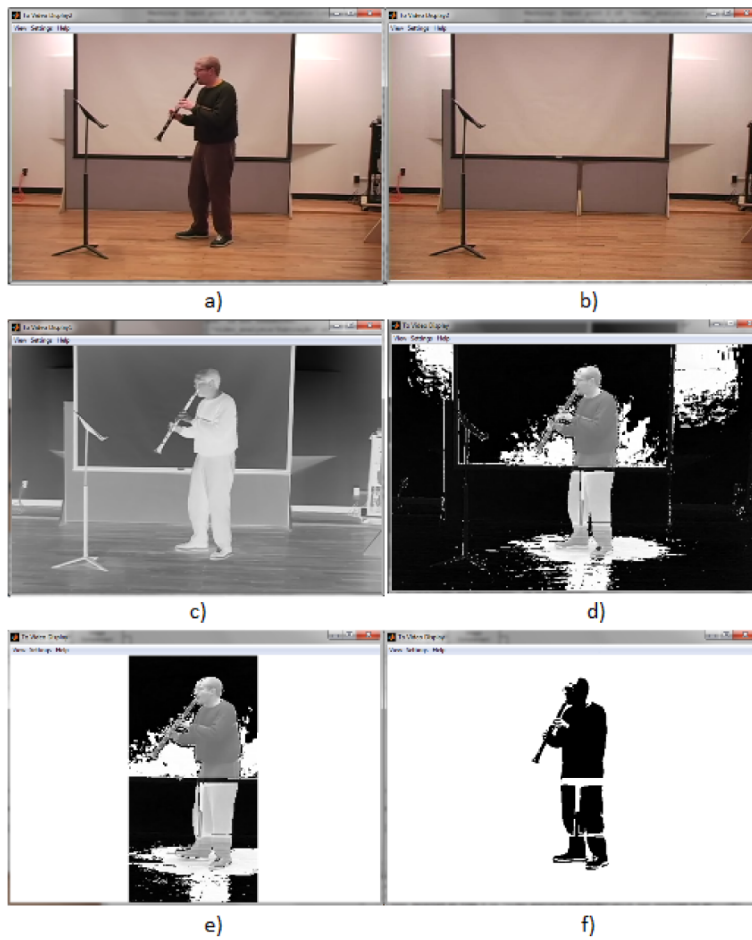


Figura 2: Resultados das etapas do pré-processamento.



Figura 3: Resultados finais do pré-processamento.

por se tratar de um instrumento anexo ao corpo do clarinetista. A detecção de movimentos baseou-se no rastreamento da trajetória de um marcador virtual fixado na extremidade do instrumento. Este rastreamento foi viabilizado através de um simples algoritmo em linguagem C, com o auxílio da biblioteca OpenCV [Bradski and Kaehler, 2008]. No entanto, ter o valor puro das coordenadas do marcador não foi o suficiente para indicar ou não a presença do movimento. Fez-se necessário relativizar a medição, baseado em um referencial interno ao músico. Era preciso que fosse interno, pois, caso não fosse, os deslocamentos do instrumentista atrapalhariam na manutenção da confiabilidade do referencial. Por isso, determinou-se o centro de massa de cada músico, em cada vídeo. Agora, referenciando a trajetória do marcador virtual ao centro de massa (quadro a quadro), pôde-se gerar a curva de deslocamento. Limiarizando esta curva, foi possível inferir os instantes dos vídeos onde ocorreram movimentos auxiliares de clarinete. Estes resultados estão exemplificados na Figura 4.

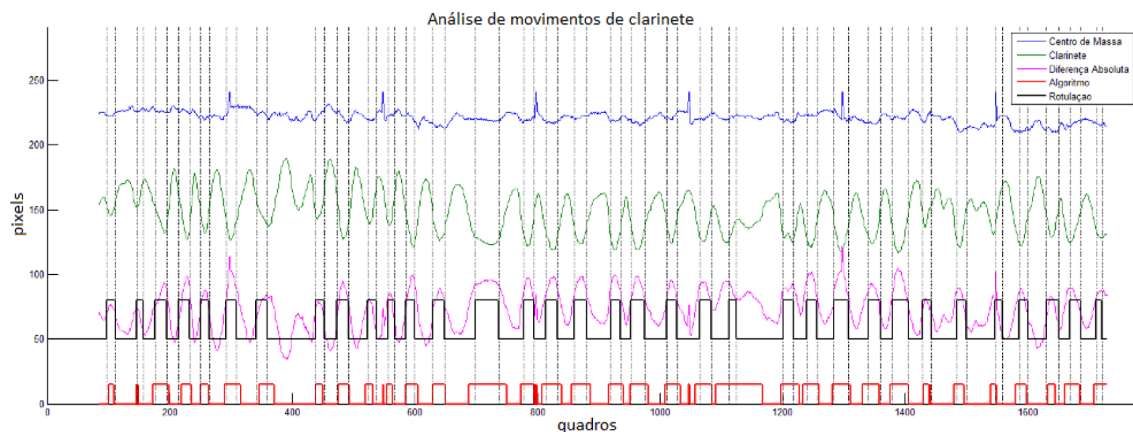


Figura 4: Análise de movimentos auxiliares de clarinete.

Para a detecção de gestos de dobra de joelhos o procedimento foi um pouco mais trabalhoso. Não se tem como rastrear pontos neste caso, uma vez que os joelhos se sobrepõem e que não há região destacada para que se possa rastrear coerentemente. Por isso, optou-se pelo emprego da técnica de projeção de perfil [14]. Gerou-se o perfil horizontal nos nove vídeos e analisou-se o comportamento da contagem de pixels na região, conforme ilustra a Figura 5a. Redução no valor de contagem de pixels indicaria os intervalos dos vídeos onde houve movimentos de dobra de joelhos (Figura 5b). Para relativizar esta medida, usou-se novamente o centro de massa, porém não sua coordenada original, mas sim um deslocamento. Originalmente o centro de massa localiza-se atrás dos joelhos, ou seja, a redução da contagem de pixels nunca se confundiria com esta referência. Assim, calculou-se um percentual do centro de massa, que seria análogo a um deslocamento à frente deste ponto. Agora, todas as vezes em que a região dos joelhos na projeção de perfil horizontal atingisse ou ultrapassasse o percentual de centro de massa, indicaria os intervalos de tempo contendo movimento. A Figura 6 exemplifica este resultado. Cada linha do gráfico representa a linha da imagem, enquanto que cada coluna é equivalente ao frame analisado. Quando a contagem de pixels é inferior ao limitador, gera-se o aspecto escuro no gráfico, ou seja, o movimento.

Por fim, programou-se a detecção de movimentos de ondulações nas costas. Similar à metodologia dos gestos de dobra de joelhos utilizou-se a técnica de projeção de perfil [Zramdini and Ingold, 1993], porém agora na vertical. Desta vez, a técnica foi aplicada apenas nos trechos da imagem onde o músico se fez presente. Isso foi possível através da determinação de bounding boxes. A bounding box nada mais é que um retângulo gerado pelas extremidades da imagem. Neste caso as extremidades são os braços, a cabeça e os pés do músico, ou seja, a pessoa sempre estará circunscrita à bounding box. Assim,

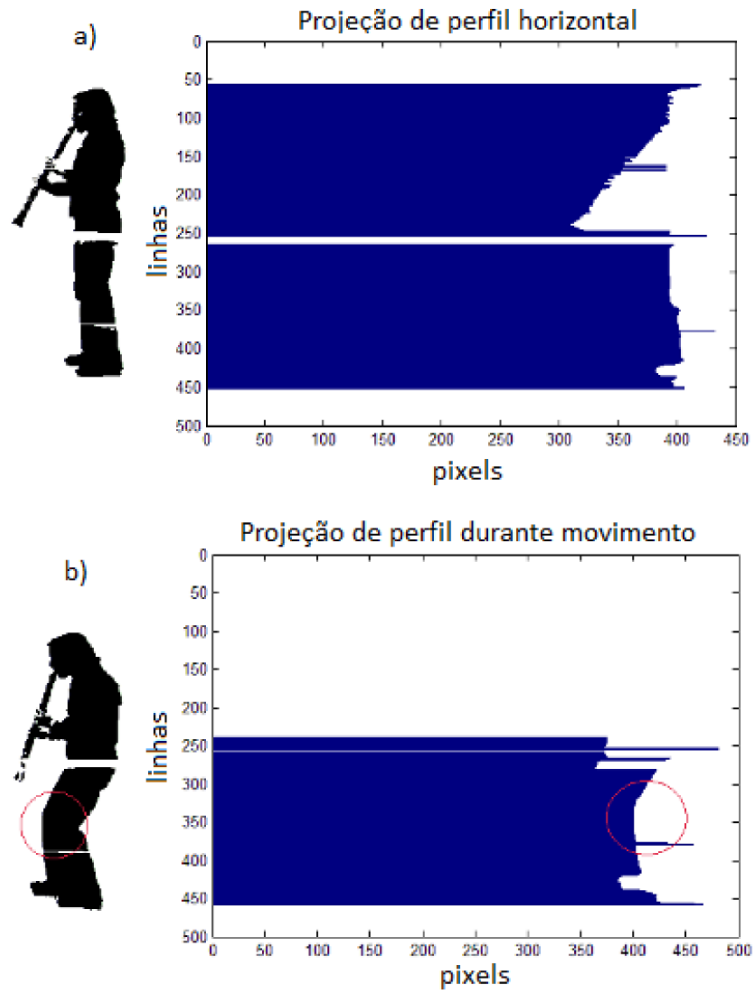


Figura 5: Projeção de perfil horizontal para detecção de movimentos auxiliares de dobra de joelhos.

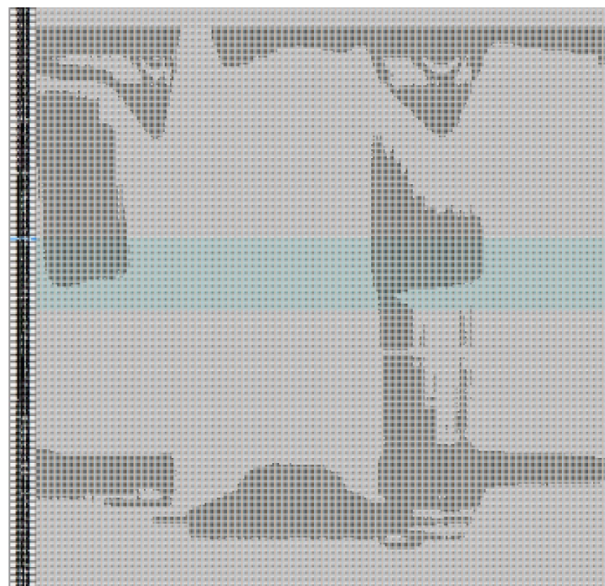


Figura 6: Análise de movimentos auxiliares de dobra de joelhos.

aplicando a projeção de perfil vertical dentro destes retângulos, gera-se o perfil vertical do músico em todos os quadros dos vídeos, conforme a Figura 7a. O movimento de ondulação nas costas pode ser resumido a variações nas barras de contagem de pixels limitadas pelo retângulo e pela cabeça do músico. Transições de pequenas para grandes áreas sugerem movimentos. Para determinar essas variações, calcularam-se dois pontos de interesse no perfil: um na cabeça e um relativo à última barra da contagem de pixels (Figura 7b). Calculando a área formada por esses dois pontos e um terceiro co-linear a ambos, pôde-se inferir a área caracterizadora do movimento (Figura 7c). Análogo ao movimento de clarinete, limiarizando os valores de área obtidos, teve-se como indicar os intervalos de ocorrência do movimento auxiliar de ondulações nas costas, conforme ilustra a Figura 8.

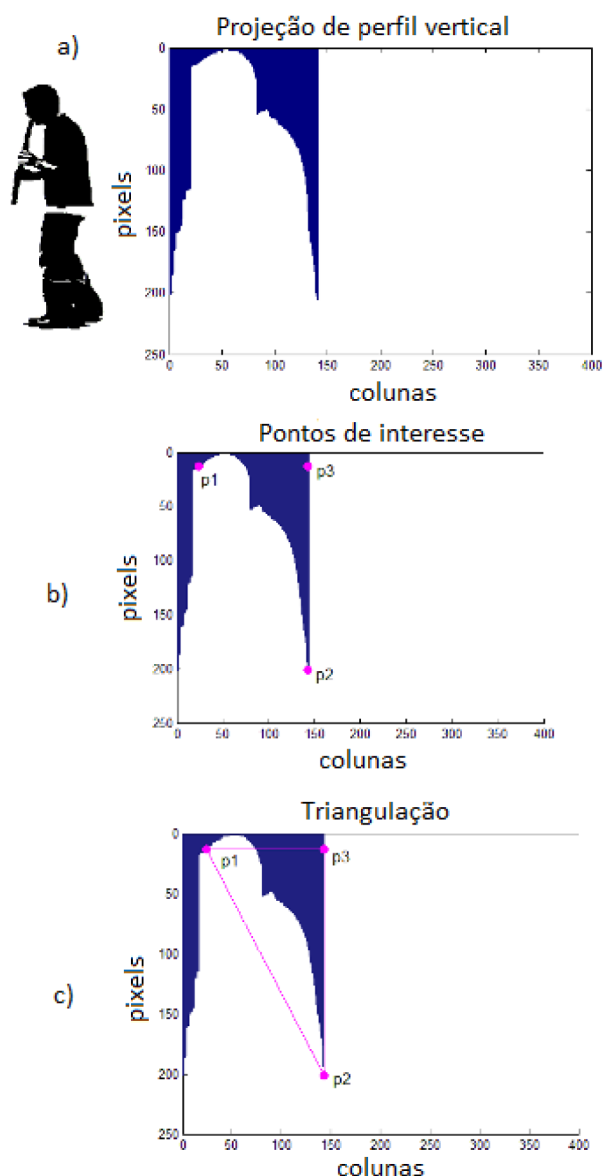


Figura 7: Projeção de perfil vertical para detecção de movimentos auxiliares de ondulações nas costas.

3. RESULTADOS

O método proposto foi avaliado em nove sequências de vídeos de quatro clarinetistas executando o Segundo Movimento das Três Peças de Stravinsky para Clarinete Solo que

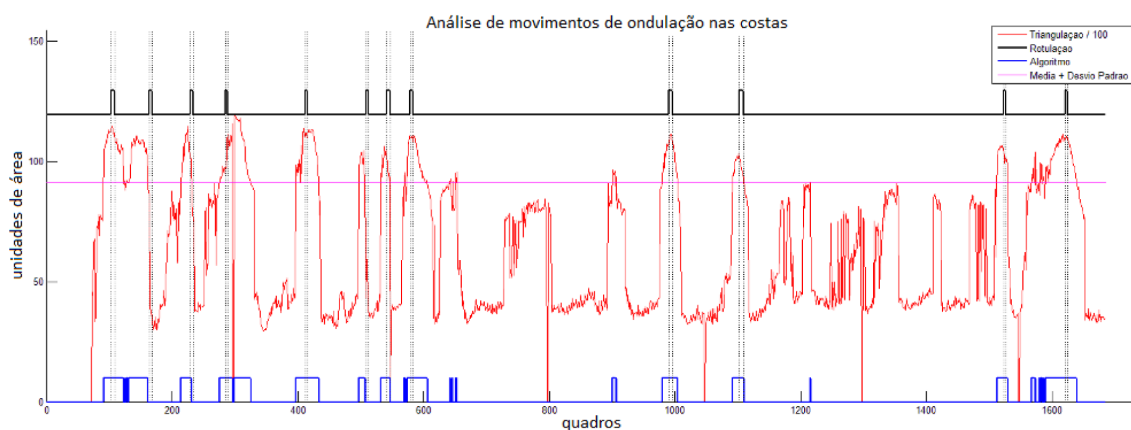


Figura 8: Análise de movimentos auxiliares de ondulações nas costas.

foram gravados no Input Devices and Music Interaction Laboratory da McGill University, Canada. Os vídeos foram gravados com uma única câmera em um estúdio com uma iluminação constante. Em particular, esta peça foi escolhida em razão da sua estrutura não-métrica; a ausência de uma métrica consistente foi utilizada para filtrar alguns movimentos rítmicos que são omnipresentes em música métrica [Wanderley et al., 2005]. Esta peça também é ideal para este estudo por que é sem acompanhamento; a representação mental do músico para o trabalho não inclui outros eventos sonoros, senão aqueles presentes em seu próprio som. Finalmente, esta peça faz parte do repertório padrão de clarinetistas avançados, o que permite a replicação ou extensão de pesquisas prévias com músicos de diferentes escolas.

Os vídeos foram capturados em um estúdio usando uma câmera de vídeo digital instalada em paralelo com o músico. A iluminação foi a do próprio ambiente, assim como o fundo. A resolução dos vídeos variam de 720x480 à 400x300 pixels e a taxa de quadros varia entre 15 e 25 quadros por segundo. Cada vídeo possui entre 1.300 e 1.700 quadros. Além disso, cada vídeo foi rotulado de acordo com a presença ou ausência dos três movimentos auxiliares de interesse. Esta rotulação é necessária para avaliar posteriormente a precisão do método proposto. Os vídeos rotulados totalizam 335 movimentos auxiliares, sendo 151 movimentos de clarinete, 54 movimentos de dobra de joelhos e 130 movimentos de ondulações nas costas. Já os algoritmos aqui propostos totalizaram 486 detecções de movimento, sendo 245 movimentos auxiliares de clarinete, 75 movimentos não óbvios de dobra de joelhos e 166 movimentos de acompanhamento de ondulações nas costas.

Dentre os 486 intervalos de tempo contendo movimento, 82 foram falsos positivos, ou seja, o algoritmo indicou os quadros, no entanto não existiam movimentos de fato. Além disso, outros 39 movimentos que ocorreram não foram detectados, ou seja, falsos negativos.

4. CONCLUSÃO

É possível reparar que o método de detecção proposto identificou cerca de 20% a mais de movimentos auxiliares que a tradicional observação dos vídeos. Além disso, aproximadamente 75% dos movimentos de clarinete, 70% dos movimentos de dobra de joelhos e mais de 80% dos movimentos de ondulações nas costas foram detectados corretamente.

A maioria das falhas de detecção (falsos negativos + falsos positivos) foi de falsos positivos e devido a falhas de pré-processamento e, principalmente, de erros de tomada de decisão baseada na limiarização das curvas - cerca de 70% dos erros. Este resultado é tolerável, uma vez o método proposta visa a substituir apenas as observações. Como os

resultados dos algoritmos são uma espécie de filtragem dos vídeos para análises posteriores de dados obtidos via marcadores, os falsos positivos implicariam apenas em análises desnecessárias. Já os falsos negativos sim são problemáticos, uma vez que representam perda de dados.

Enfim, os resultados aqui expostos são condicionados à base estudada, não sendo possível sua generalização, entretanto, foram conclusivos e comprovam que é possível detectar movimentos auxiliares de clarinete, de dobra de joelhos e de ondulações nas costas através de simples arranjos de algoritmos de visão computacional, simplificando o método de avaliação.

Para se melhorar estes resultados, faz-se necessário aprimorar as técnicas de pré-processamento dos vídeos e estudar formas alternativas de tomada de decisão, ou seja, substituir a proposta de limiarização das curvas por outros arranjos menos suscetíveis a erros, para então compilar os métodos em um único software detetor dos três gestos auxiliares propostos.

Referências

- Aggarwal, J. K. and Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding*, 73:428–440.
- Bradski, G. and Kaehler, A. (2008). Learning opencv: Computer vision with the opencv library. *O'Reilly*, 1:556.
- Cadoz, C. (1998). Instrumental gesture and music composition. *Proceedings of the International Computer Music Conference*, 1:1–12.
- Dahl, S. and Friberg, A. (2004). Expressiveness of musicians' body movements in performances on marimba. *Gesture-Based Communication in Human-Computer Interaction*, 1:479–486.
- Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21:103–113.
- Gavrila, D. M. (1999). The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73:82–98.
- Gonzalez, R. C. and Woods, R. E. (2007). Digital image processing. *Prentice Hall*, 3:976.
- Jenseni, A. R., Wanderley, M. M., Godoy, R. I., and Leman, M. (2010). Musical gesture: Concepts and methods in research. *Musical Gestures: Sound, Movement and Meaning*, 1:12–35.
- Mitra, S. and Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man and Cybernetics*, 37(3):311–324.
- Teixeira, E. C. F. (2010). Análise da expressividade musical com base em medidas acústicas e do gesto físico. *Dissertação de Mestrado*, 1:46.
- Verfaille, V., Quek, O., and Wanderley, M. M. (2006). Sonification of musicians' ancillary gestures. *Proceedings of the 12th International Conference on Auditory Display*, 1:20–23.
- Wanderley, M. M. (1999). Non-obvious performer gestures in instrumental music. *Gesture-based Communication in Human-Computer Interaction*, 1:37–48.
- Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. *Gesture and Sign Language in Human-Computer Interaction*, 1:241–253.

- Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., and Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34(1):97–113.
- Zramdini, A. and Ingold, R. (1993). Optical font recognition from projection profiles. *Electronic Publishing*, 6(3):249–260.

Impacto Regional do SBCM: Caso de estudo em Vitória-ES

Leandro Costalonga¹, Marcus Neves²

¹ CEUNES - Universidade Federal do Espírito Santo (UFES),
Rodovia BR 101 Norte, Km. 60, Bairro Litorâneo, CEP 29932-540, São Mateus – ES

² Departamento Teoria das Artes e Música - Universidade Federal do Espírito Santo (UFES), Cemuni V, Centro de Artes, Av. Fernando Ferrari, 514, Goiabeiras, CEP 29075-910, Vitória – ES

{leandro.costalonga,marcus.neves}@ufes.br

Abstract. *The heritage that a scientific gathering may leave to the host city is not often analysed. This paper presents the legacy that the SBCM 2011 has left to the city of Vitória - ES. The conference was held in the state of Espírito Santo (Brazil). At the time, the local Universities had no scientific production in Computer Music. Two years has passed since the event and it is fair to state that some progress have been made locally. As a direct result, it is possible to list the following results: a) two graduation works (final project); b) regular offer of field-related modules; c) musical and artistic performances; d) institutional partnership focused on game development; e) research group foundation; f) a program and three extension projects with 5 researcher and 8 graduate students.*

Resumo. *O legado que um congresso científico deixa na cidade sede e instituições organizadoras não é rotineiramente analisado e relatado. Este artigo aborda o impacto que SBCM 2011 teve em solo espírito-santense, naquele momento, um dos poucos estados brasileiros que não possuíam produção em Computação Musical. Portanto, apontaremos as orientações de trabalho de conclusão de curso, ofertas regulares de disciplinas nos cursos de graduação, as apresentações artísticas realizadas e as parecerias interinstitucionais, assim como a criação do grupo de pesquisa NESCoM, do programa de extensão “Artes Digitais” e as perspectivas futuras.*

1. Introdução

Há dois anos (2011) ocorria em Vitória-ES o último Simpósio Brasileiro de Computação Musical. Acreditava-se que a realização do SBCM em solo capixaba pudesse favorecer a criação de um canal de comunicação entre as principais instituições de ensino superior do estado, possibilitando a consolidação de um grupo de pesquisa ativo na área, além de despertar o interesse da sociedade pelos meandros que permeiam as produções e performances da computação musical. Entre as Instituições de Ensino Superior organizadoras estavam a FAESA, FAMES e UFES.

Com mais de três décadas de existência, a FAESA constitui-se num marco de referência no cenário educacional e cultural do Estado do Espírito Santo. Na sua Unidade de Engenharia, Computação e Sistemas (UECS) abarca um centro de

excelência na formação de profissionais da computação através dos cursos de: Engenharia da Computação, Ciência da Computação, Sistemas de Informação, Tecnologia em Redes de Computadores, Tecnologia em Análise e Desenvolvimento de Sistemas, e o pioneiro curso de Tecnologia em Jogos Digitais. Soma-se a esses os cursos de Engenharia de Automação e Controle e Engenharia Elétrica.

Com mais de 50 anos, Faculdade de Música do Espírito Santo Maurício de Oliveira (FAMES), atualmente autarquia do governo estadual vinculada à Secretaria de Educação do estado do Espírito Santo, foi a primeira instituição a oferecer curso superior de música no estado. Além de ser um centro acadêmico, a FAMES mantém em suas dependências o Centro de Formação Musical do Espírito Santo, uma unidade que ministra cursos permanentes de extensão, de longa, média e curta duração, objetivando dar formação musical a crianças, jovens, adolescentes e adultos. Esta unidade de ensino oferece o Curso de Musicalização Infantil e Curso de Formação Musical, nos gêneros erudito e popular. Os cursos superiores oferecidos pela FAMES ofertam anualmente 90 vagas, sendo 60 vagas direcionadas para o Curso de Licenciatura em Música e 30 para o Curso de Bacharel em Música com habilitação em instrumento/canto, com várias habilitações oferecidas.

Atualmente, Universidade Federal do Espírito Santo (UFES) também oferece o curso de Licenciatura em Música desde 2000. Em 2010, através do REUNI, a UFES começou a ofertar também o curso de Bacharelado em Música com Habilitação em Composição e Ênfase em Trilha Musical. Com entrada anual através de processo seletivo – assim como a Licenciatura –, os dois cursos somam mais de 250 alunos matriculados no semestre 2013/2.

Conclui-se, na época, que a carência de conhecimento em Computação Musical no Estado advinha da falta de comunicação entre as instituições de ensino superior e que, ao sediar o SBCM, as IES teriam a chance de dialogar e estabelecer uma colaboração interinstitucional em nível regional, nacional, e até internacional. O SBCM 2011 teve a missão de, mais do que propiciar à comunidade acadêmica uma chance para trocar experiências, de levar esse conhecimento à sociedade capixaba, principalmente aos profissionais tocados pela interdisciplinaridade da área.

Esse artigo analisa o impacto que o SBCM propicia em comunidades de Computação Musical ainda embrionárias, como era o caso do Espírito Santo. A seguir, relata-se as atividades que ocorreram no estado nos anos subsequentes ao evento bem como as perspectivas futuras nas principais instituições.

2. Registro de Atividades Pós-SBCM

O SBCM 2011 superou as expectativas em termos de qualidade dos trabalhos, tutorial e palestras apresentados. No entanto, o fato que mais chamou a atenção foi número recorde de inscrições: 249 inscritos (em grande parte alunos oriundos das instituições promotoras). No entender da organização isto reflete, não somente um acerto no processo de divulgação local, mas também um crescente interesse nesta área no Brasil. Mais que isso: evidenciou-se um interesse latente dos alunos e acenou com repercussões positivas junto aos mesmos.

2.1 Integração dos cursos de Desenvolvimento de Jogos Digitais e Bacharelado em Música

Conforme previamente mencionado, a FAESA oferece o curso de Desenvolvimento de Jogos Digitais. Na grade curricular do curso há uma disciplina de *Design de Som* voltada a produção efeitos sonoros e trilhas musicais. Uma dificuldade que constantemente permeava as atividades didáticas era a heterogeneidade da turma: alguns alunos eram bons programadores, outros bons desenhistas ou roteiristas - uma minoria tinha conhecimento musical.

A UFES, através do curso de Bacharelado em Música com Habilitação em Composição e Ênfase em Trilha Musical, passou a fortalecer o diálogo interinstitucional enviando à disciplina de Desenvolvimento de Jogos Digitais da FAESA três alunos da turma mais avançada do seu curso - a saber, pertencentes à primeira turma de ingressantes - naquele momento para um trabalho em conjunto. Os alunos escolhidos cursavam a disciplina de *Design de Som* junto com os alunos do curso de Desenvolvimento de Jogos e foram envolvidos, desde o início, no processo de concepção dos jogos. Como exemplo de sucesso dessa parceria pode-se citar o fato de um desses alunos oriundos do curso de Música da UFES ser atualmente o *Sound Designer* da desenvolvedora de jogos digitais *Extintogamers* (2011), cujos fundadores foram alunos do curso em questão da FAESA. Um dos jogos que possui trilha sonora assinada aluno em questão é *Esfarelante* [Extintogamers 2013], que tem cerca de 1000 acessos semanais. Outros trabalhos fruto dessa iniciativa podem ser visto no site do artista [Martins 2013].

2.2 Trabalho de Conclusão de Curso

No intervalo de um ano, dois trabalhos de conclusão de curso foram defendidos. O primeiro, intitulado “Computação Musical e as Novas Tecnologias da Informação”, foi desenvolvido pelo aluno do curso de Bacharelado em Música da UFES. O objetivo do trabalho era discutir algumas das várias possibilidades de aplicação das tecnologias da informação e da computação musical como ferramentas didáticas nos cursos de graduação em música. Utilizando a classificação de Eloy Fritsch [2004] entre softwares musicais pedagógicos e musicais não-pedagógicos, apresentou-se as principais características dessas duas classes de softwares no intuito relatar seu campo de aplicação em música, apresentando o computador e as tecnologias da informação como ferramentas didáticas que facilitam e/ou melhoram a execução de determinada atividade pelos alunos e professores.

O segundo trabalho foi desenvolvido por alunos do curso de Ciência da Computação da FAESA. O trabalho intitulado “Produção Algorítmica de Trilhas Sonoras Voltada a Jogos Digitais Baseada em Emoções” propõe uma biblioteca desenvolvida para a geração de música, tendo definições abstratas, como emoções ou sentimentos, como única forma de entrada, retirando a necessidade de qualquer conhecimento musical de quem for utilizá-la. Foi dado foco especial na criação de músicas para jogos eletrônicos, mas o algoritmo final não é de forma alguma restritivo a este universo.

2.3 Apresentações e Obras Musicais

Em paralelo à realização do SBCM 2011, foi criado na UFES o GEXS - Grupo de Experimentação Sonora, liderado pelo Prof. Marcus Neves, desenvolvendo em seu âmbito dois trabalhos: a) *Atari Sound Performance* (2011-2013), em parceria com o

então professor substituto e artista sonoro Herbert Baioco. Trata-se de performance compartilhada envolvendo instrumentos derivados do *circuit bending* em diálogo com o processamento em tempo real do som, através de *patch* construído em *Pure Data*, tendo também o público como *performer* que joga um *video game* Atari 2600 durante a apresentação, produzindo material sonoro primário que é enviado aos artistas e processado. Produz-se então uma nova trilha para o jogo, mixada em tempo real, constituindo assim o que os artistas chamam de fluxo sonoro contínuo; b) a instalação *(en)rolo* (2012), que se utiliza de reprodutores de fita K-7 para explorar, através de técnicas de *circuit bending*, questões ligadas não só à reprodução das fitas, mas a conceitos que são disparados a partir da elaboração da obra, como memória e afetividade. Este trabalho contou com efetiva participação de três alunos da Licenciatura em Música da UFES, primeiros discentes integrantes do GEXS.

Atari Sound Performance foi realizada, entre 2011 e 2013, no Plano B/RJ [2011], uma das principais casas de música experimental do país, no Festival de Música de Londrina/PR, Projeto Multissensorial/ES, Festival Música Livre/ES, em 2011, e Virada Cultural/SP, em 2013.

A obra *(en)rolo* é fruto direto do SBCM 2011 visto que as alunas que participaram do processo foram monitores do evento, tendo, em sua maioria, o primeiro contato direto com o campo da tecnologia aplicada à música. O GEXS, em 2012, trabalhou em colaboração com o projeto de extensão "Música para o Teatro do Inconsciente" (UFES), coordenado pelo Prof. José Eduardo Costa Silva, na composição, a partir dos materiais sonoros gerados na *Atari Sound Performance*, da trilha do espetáculo *Abram-se os histéricos!*, da Companhia Inconsciente em Cena, do Rio de Janeiro, trilha esta dirigida e composta pelo docente parceiro.

2.4 Oferta de Disciplinas Específicas

Em 2012/1, foi realizado o curso de 20 horas de Programação Musical em *Pure Data* oferecido aos alunos da Licenciatura e Bacharelado em Música da Universidade Federal do Espírito Santo, como *intermezzo* entre as disciplinas de Laboratório de Estilos Musicais: Música eletroacústica I e II, ministradas, assim como o curso, pelo Prof. Marcus Neves. O curso teve como objetivos: a) proporcionar contato inicial com a programação musical no software *Pure Data*; b) estudar aspectos acústicos e psicoacústicos do som; c) realizar sínteses diversas através da escrita no software em uso. O curso possibilitou os alunos a criarem pequenas peças como produto final.

Em 2013/1, o Prof. Leandro Costalonga passou a integrar o quadro de professores efetivos do Centro Universitário Norte do Espírito Santo [CEUNES] da Universidade Federal do Espírito Santo. Imediatamente, propôs a oferta da disciplina de Tópicos Especiais em Programação I cuja a ementa era direcionada ao ensino de *Pure Data* como uma forma de facilitar o entendimento de conceitos básicos de programação. É fato notório e conhecido a grande retenção de alunos nas disciplinas básicas de programação dos cursos de Ciência e Engenharia da Computação [Deters 2008]. A procura pela disciplina foi alta, sendo necessário dobrar o número de vagas.

2. Fundação do Núcleo Espírito-santense de Computação Musical

A criação do Núcleo Espírito-santense de Computação Musical (NESCoM) é resultado direto da parceria interinstitucional promovida para a organização SBCM. O registro do

NESCoM junto aos CNPq foi recentemente formalizado com 5 docentes da FAESA e da UFES nas áreas de: Ciência da Computação, Engenharia da Computação, Engenharia Elétrica, e dois pesquisadores da Música. No momento, 8 alunos estão cadastrados trabalhando com as pesquisas do grupo.

A principal linha de pesquisa do grupo é em Sistemas Multimídia Interativos, principalmente Instalações Artísticas e Artefatos Culturais Automatizados por Computador. O suporte propiciado aos artistas pelos especialistas em computação gráfica, computação musical e eletrônica possibilitam a realização de trabalhos inovadores que avançam tanto no estado da arte da computação como na parte artística.

Atualmente o NESCoM oferece suporte ao programa e projetos abaixo descritos:

3.1 Programa de Extensão em Artes Digitais (2013)

Arte Digital é um termo geral para classificar uma variedade de trabalhos e práticas artísticas que usam a tecnologia digital (software, hardware, e mídia) como uma parte essencial do processo criativo e de apresentação. Desde da década de 1970 vários nomes tem sido usados para descrever esse processo incluindo arte computadorizada e arte multimídia. A própria Arte Digital esta sob o escopo das Novas Artes Midiáticas ou Arte Contemporânea [Paul 2003].

O impacto da tecnologia digital tem transformado as já conhecidas atividades artísticas tais como a pintura, desenhos, escultura, e a música. Ainda há novas formas de práticas artísticas como as instalações multimídias, realidade virtual, realidade aumentada, composição algorítmica [Lieser 2009].

Como principal objetivo, o programa visa oportunizar o encontro e diálogo entre engenheiros e artistas visando o desenvolvimento de novas tecnologia e produção cultural. Outros objetivos são: organização de palestras semestrais de divulgação do programa; acolhida e estímulo a participação em projetos de pesquisa e extensão que desenvolvem ou usem novas tecnologias voltadas às artes digitais, principalmente no campo da Eletrônica, Computação Gráfica e Computação Musical. Estão ligados ao programa de extensão referido as seguintes atividades: a) eventos de extensão como exposições artísticas, concertos, seminários; b) projetos de extensão como a Orquestra de Computadores e o Campus Vivo; c) Desenvolvimento de novas tecnologias em conjunto com o Núcleo Espírito-santense de Computação Musical e outros de reconhecida competência.

3.2 Projeto Campus Vivo (2013)

Artefatos Culturais Automatizados por Computador (CACA) e/ou Instalações Multimídia Interativas são objetos oriundos do choque entre de visões de mundo muitos diferentes. Enquanto as Engenharias focam em funcionalidade, otimização, e desempenho, as Artes tem seu foco na estética, diálogo, e bem estar. O projeto Campus Vivo promove o encontro de indivíduos desses dois mundos e entrega o resultado dessa cooperação a comunidade acadêmica, principalmente dos campi do interior do Espírito Santo, que sofrem com escassez de cultura e atividades artísticas por ainda estarem em fase de consolidação. Assim, proporcionar o contato e a imersão desta comunidade à instalações e máquinas que, se não pudessem ser construídas em um ambiente acadêmico com fins educacionais, dificilmente seriam viabilizadas economicamente somente para sua utilidade intrínseca – esse é principal objetivo do Campus Vivo.

Do ponto de vista acadêmico, o projeto visa estimular o pensamento criativo de estudantes dos cursos de engenharia e ciência da computação através da aplicação do conhecimento tecnocientífico na concepção e construção de instalações interativas e artefatos culturais computacionais. Tais projetos não necessariamente possuem um propósito funcional sob o prisma da engenharia; bastam que essas instalações/obras/artefatos, de caráter artístico, possam despertar a curiosidade da comunidade acadêmica que, ao interagir com as mesmas, tornam o campus um lugar mais "vivo" e divertido – o campus se tornaria uma espécie de museu de artes modernas a céu aberto.

3. Projeto Orquestra de Computadores

O marco inicial do desenvolvimento da música eletrônica foi a invenção do telefone por Alexander Graham Bell em 1876. Importante avanços sucederam até que em 1957 a música encontra a computação pelas mãos de Max Mathews no Bell Laboratories. Mathews criou um programa de computador que manipulava dados musicais. O programa chamado Music I tinha uma única voz, uma forma de onda triangular, não possuía ADSR e só controlava a afinação, intensidade e duração dos sons. O Music I deu origem a uma série de programas musicais como Music II, Music III e Music IV. Estes programas abriram espaço para uma “avalanche” de novos programas musicais de todas as categorias que foram criados a partir de 1976 com a difusão dos microcomputadores e a utilização de linguagens de programação de alto nível e grande portabilidade [Miletto 2004]. Hoje, o computador é considerado um importante instrumento musical capaz de superar as limitações dos instrumentos acústicos.

O projeto Orquestra de Computadores visa apresentar a comunidade o computador como um instrumento musical de grande versatilidade e relativa facilidade de manuseio através de concertos abertos ao público. Um grupo heterogêneo formado por alunos e professores dos cursos de computação (engenharia e ciência) e de música, realizam ensaios periódicos onde exploram novas técnicas e tecnologias para síntese sonora, produção musical, e técnicas performáticas. Concertos interativos são realizados nos 3 campi da UFES.

4. Leitura Solidária

Os livros falados (*audiobooks*) têm ganhado popularidade com disseminação e barateamento de reprodutores de áudio digital integrado aos dispositivos de computação e comunicação móveis de uso pessoal como, por exemplo, os *smartphones*. Até então, deficientes visuais tinham como única opção de “leitura” as restritas publicações em Braille. Entretanto, seria errôneo pensar que Livros Falados são dirigidos exclusivamente a indivíduos com deficiência visual – uma gama de pessoas com dificuldade de leitura, por quaisquer que sejam os motivos, tem nos livros falados uma oportunidade de conhecer novas obras. Infelizmente, o acervo disponível em português de livros falados ainda é muito restrito. No projeto Librivox.org (2005), que disponibiliza livros falados gratuitamente em várias línguas, dos quais 160, de aproximadamente 6.500 livros, são em português, ou seja 2.5% do acervo total.

Esse projeto, apesar de não estar no escopo do programa de artes digitais e de não avançar o estado da arte da Computação Musical, é de interesse do NESCoM pois possibilita a disseminação de técnicas de gravação, sonoplastia, edição sonora, além do desenvolvimento e aplicação de filtros e efeitos específicos sob demanda.

3. Conclusão e Perspectivas Futuras

Tanto os projeto Campus Vivo como a Orquestra de Computadores ainda estão em estágio inicial – é muito prematuro falar de resultados nesse momento. Pode-se, no entanto, afirmar que a procura dos alunos tem sido muito positiva.

Infelizmente, a FAMES, apesar de ter sido grande apoiadora do SBCM 2011, não tem participado ativamente da iniciativas da área junto as outras IES. Aparentemente, falta um professor que tenha interesse acadêmico e artístico pela área e queira servir de pivô para a integração da instituição com os trabalhos do NESCoM e dos Programa extensionista em Artes Digitais. Esse é um ponto a ser trabalhado pelo grupo, reaproximação e inclusão da Faculdade de Música do Espírito Santo nas atividades.

No geral, pode-se afirmar que o SBCM deixou bons frutos para o estado do Espírito Santo. Nesse dois anos após o SBCM 2011 temos com resultados: a) orientação de dois trabalhos de conclusão de curso tanto na Computação como na Música; b) oferta regular de disciplinas e tanto nos cursos de Computação como na Música; c) produção de apresentações musicais e artísticas; d) parceria interinstitucional para o desenvolvimento de Jogos Digitais; e) criação do grupo de pesquisa NESCoM; e f) criação de um programa de extensão; g) elaboração de três projetos de extensão que contam com 5 docentes e 8 alunos de graduação. Esses resultados ainda são tímidos frente nossa ambição, mas o grupo julga ser um primeiro e importante passo na direção certa.

Referências

- Deters, J. I., da Silva, J. M. C., de Miranda, E. M., & Fernandes, A. M. R. (2008). O Desafio de Trabalhar com Alunos Repetentes na Disciplina de Algoritmos e Programação. Simpósio Brasileiro de Informática na Educação.
- ExtintoGamer, (2011). <http://extintogamer.com.br/wp/jogos/>. Acesso em Agosto 2013.
- ExtintoGamer, (2013). <http://www.scirra.com/arcade/addicting-action-games/5471/the-esfarelante>. Acesso em Agosto 2013.
- Fritsch, E. F.; Miletto, E. M.; Costalonga, L. L. (2004) Uma Proposta de Método para o Ensino de Técnicas de Composição de Música Eletrônica por Computador. In: ANAIS DO SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, Manaus. Anais eletrônicos... Disponível em: < <http://www.br-ie.org/pub/index.php/sbie/article/view/370/356>>. Acesso em: 30 de abr. 2012.
- Librevox (2005). <http://librivox.org/>. Acesso em: Agosto 2013
- Lieser, W., & Baumgärtel, T. (2009). Digital art. hf ullmann.
- Martins, D. (2013) .<http://deyvidmartins.wordpress.com/> Acesso em: Agosto 2013
- Miletto, E. M., Costalonga, L. L., Flores, L. V., Fritsch, E. F., Pimenta, M. S., & Vicari, R. M. (2004). Introdução à computação musical. In IV Congresso Brasileiro de Computação.
- Paul, C. (2003). Digital art (p. 29). London: Thames & Hudson.

O conceito de música experimental no ciberespaço: uma pesquisa etnográfica em comunidades virtuais de música

Ramiro Galas¹

¹Programa de Pós-Graduação Música em Contexto da Universidade de Brasília (UnB)

ramiro.galas@gmail.com

Resumo. *Este trabalho utiliza a etnografia virtual e a análise de conteúdo em comunidades virtuais em música com objetivo de delinear o que é “experimental” nesse contexto. Para tanto, mapeou-se as diversas definições de experimental, separando-as em grupos semânticos, posteriormente comparadas ao conceito de música experimental presente na academia. Como resultado, observa-se a distância conceitual entre esse termo na academia e nos escritos críticos e musicológicos e a maneira como esse é entendido pelas comunidades virtuais de música “experimental”.*

Abstract. *This paper uses virtual ethnography and content analysis in virtual communities of music in order to delineate what is "experimental" in this context. To this end, it mapped to the different experimental settings, separating them into semantic groups, which are compared to the concept of experimental music at the Academy. As a result, there is the conceptual gap between this term in the academy and critical writings and how it is understood by virtual communities of experimental music.*

1. Introdução

Atualmente, há diversas redes sociais que servem para fazer circular música. O Soundcloud é uma rede social, com sede na Alemanha, que ficou disponível em 2007. Hoje, já possui mais de 10 milhões usuários, conforme a *newsletter* (boletim informativo) de janeiro de 2012. Essa rede destaca-se pelo limite maior de músicas compartilhadas e por ser uma interface mais simples e intuitiva – que não contém *chat* e não armazena fotos (somente o avatar). Cria-se uma visualização gráfica de cada faixa compartilhada, na qual é possível fazer comentários e, a partir da “própria música”, interagir com outros ouvintes e produtores (ver Figura 1).

No Soundcloud, além de criar seu perfil pessoal, no qual o usuário pode tanto compartilhar suas produções como ouvir e interagir com outros produtores, há a possibilidade de criar grupos, que são de criação espontânea, cujo o objetivo é agrupar pessoas de semelhantes interesses musicais, seja este um gênero, época, equipamento, instrumento etc. Neste trabalho, por meio do mapeamento das descrições de todas as comunidades (grupos) da rede Soundcloud que contém a palavra-chave “experimental”, criou-se, para rigor da análise, sete categorias (grupos semânticos), podendo cada grupo estar aliado a um ou mais dessas categorias. Feito isso, busca-se, por meio da revisão de literatura, o conceito de música experimental presente na academia, para, então, traçar um paralelo entre as ideias sobre música experimental na academia nas descrições dos grupos de música experimental da rede Soundcloud.

2. Música e a rede social *Soundcloud*

O ciberespaço, e as novas possibilidades de circulação da música que ela proporcionou, repaginou a maneira como consumimos música. A crescente popularização da internet significa “uma guinada [...] de novas possibilidades de exposição e compartilhamento” [Pinheiro 2009, p. 200] e ainda procura-se entender o que dessa revolução vai permanecer, sendo adotada como padrão no futuro. As redes sociais, conforme Boyd e Ellison (2007), são serviços de internet que permitem aos usuários a: *i*) montar um perfil público ou semipúblico dentro de um sistema conectado; *ii*) articular uma lista de outros usuários com os quais eles dividem uma conexão; e *iii*) ver e atravessar a sua lista de conexões e aquelas feitas pelos outros dentro do sistema. Essas redes são hoje mais um instrumento de divulgação e compartilhamento de música disponibilizado pela rede. Nelas, produtores e consumidores criam diversas cadeias de pessoas interessadas em música.

O Soundcloud possibilita gratuitamente uma conta para cada usuário, na qual terá um limite de minutos de música para compartilhar e poderá também a qualquer momento criar grupos¹. É justamente nesses “grupos” que esta observação se conduzirá. Sendo cada grupo de criação livre pelos usuários, existem hoje muitos grupos, cada um buscando criar um nicho específico, servindo de filtro para usuários que buscam pontualmente certos tipos de música e direcionando usuários para os conteúdos de interesse. Um desses nichos que interessa aqui é o formado pelos grupos que se identificam com o conceito de música experimental.

O que realmente seria “experimental” no contexto dessas comunidades virtuais? O que marca a “experimentação” das obras musicais nesse contexto? Este artigo irá buscar as diversas ideias que constroem esse conceito, por meio de uma observação das comunicações mediadas por computador (CMC) (HINE, 2000) e da análise de conteúdo (BARDIN, 1977) dos textos disponibilizados pelos usuários criadores de cada grupo. Objetiva-se, desse modo, investigar de que forma o conceito “experimental” é entendido pelos usuários dessa rede. A partir da coleta e da categorização dos textos dos grupos que contêm essa palavra no título, pode-se montar um panorama do que se pensa sobre o “experimental” no contexto dessa rede e as suas relações com o conceito de experimental na música presente na academia. Após a leitura de todos os textos coletados, as categorias (campos semânticos) foram criadas no sentido de indicar, ou não, a presença de certos significados na definição de música experimental. Pretende-se também, a partir dos mapeamento das definições dessa comunidades, elencar alguns fatores que, embasados pelas análises dos dados, descrevem o que significa ser “experimental” nesse contexto.

3. Metodologia e análise dos dados

Usa-se aqui a metodologia etnográfica da pesquisa em música, considerando, dessa forma, o ciberespaço como o local no qual se insere o observador e os grupos da rede Soundcloud como as “tribos” nas quais se concentrará a observação. Ao considerar essas comunidades como grupos “que lidam com música das mais diferentes maneiras no ciberespaço”, pode-se compreender que “o homem contemporâneo [...] apesar de ‘transitar’ num ‘espaço’ considerado virtual”, afirma Caroso (2009), “é tão real quanto a quantidade incomensurável de *hardware* que suporta a dita virtualidade”. Ou seja, apesar de imersos em contextos virtuais, cada

usuário é real. Lida-se aqui, portanto, com pessoas reais e situações de criação e interação entre ouvintes e criadores reais. Nesse sentido, porém, o campo de atuação do pesquisador encontra-se diluído pelas características virtuais dessas comunidades, por isso usa-se a etnografia virtual (HINE, 2000) para lidar com as comunicações mediadas por computador (CMC) e utilizá-las como *corpus*.

Dessa forma, percebe-se essas comunidades virtuais como grupos sociais que emergem da rede e desenvolvem discussões e ideias públicas durante um tempo notável, mobilizando sentimentos humanos necessários para formar redes de relacionamentos pessoais no ciberespaço (GARRIDO, 2005). Ideia corroborada também por Katz (2004, p. 172), ao destacar que, apesar da escuta musical na internet não congregar pessoas em um mesmo espaço físico, e muitos dos seus membros nem se conhecem, uma comunidade funciona muito bem para agregar pessoas de semelhantes interesses musicais.

Nesta pesquisa, os textos disponibilizados pelos usuários da rede são considerados dados para análise, em um processo metodológico semelhante ao trabalho de Ferreira (2004), no qual o autor analisa os textos disponibilizados pelos usuários da rede, buscando sintetizar as ideias presentes nos membros dessa comunidade. Dessa forma, a filiação dos usuários aos grupos e suas definições nos servem como dado para a tentativa de generalização e delimitação do que seria o “experimental” no contexto em que esta pesquisa se desenvolve, ou seja, na rede social Soundcloud.

Fez-se inicialmente um mapeamento de todos os grupos que contém a palavra “experimental” em seu título (ex.: “Mad Experimental”, “ethno-experimental”). Registrou-se também a definição de cada grupo, pois, ao serem criados, um espaço para a descrição textual do grupo é disponibilizado, e é nesse espaço que a observação se concentrará, buscando classificar essas descrições com base em campos semânticos (ideias relacionadas a um mesmo sentido)².

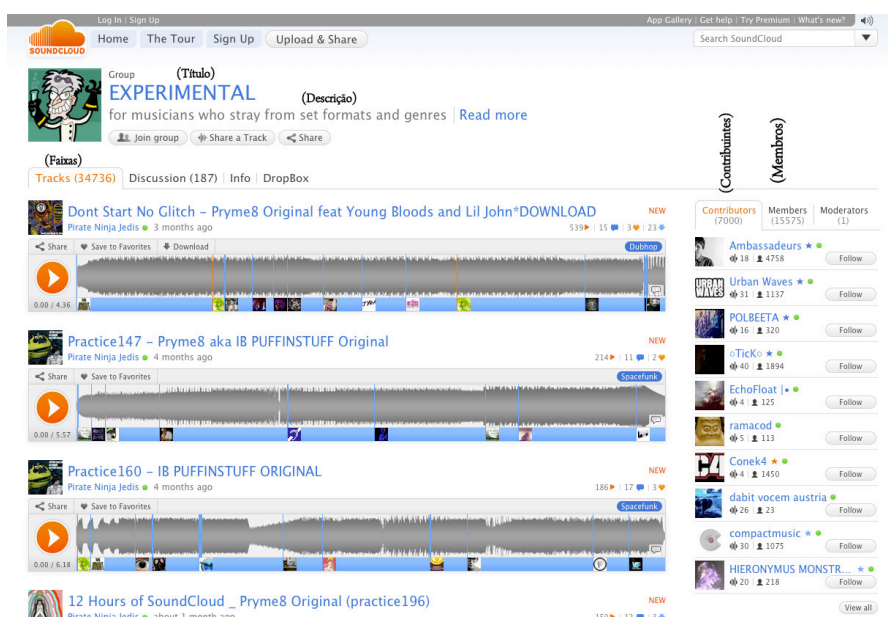


Figura 1. Visualização da página de um grupo: título, descrição, faixas compartilhadas, membros e contribuintes

4. A música experimental

Desenvolve-se aqui uma revisão de literatura sobre o que seria a música experimental. Brian Eno, no prefácio do livro *Experimental music: Cage and beyond*, resume algumas ideias que definem essa manifestação. Inicialmente, muitos dos artistas, que eram também professores de Artes, e alunos chamavam atenção para uma música que mais interessada “no processo do que no produto”. Ou seja, incluía-se a música como algo sensitivo, físico, “livre das narrativas estruturais, livre para a pura experiência sonora” e também como algo sofisticado, intelectual, “na qual podíamos exercitar e testar proposições filosóficas e intrigantes procedimentos de jogo” (ENO, 2009, p. xi-xii). Enfim, um movimento de constante de autoquestionamento sobre o que seria música.

Nyman (2009, p. 1-30), em sua busca pela definição da música experimental, detalha as seguintes características, destacando a performance da peça *4'33''*, de John Cage, como ponto de referência: *composição*, mudança radical da notação musical, não definição de um *objeto-tempo*, utilização de processos de criação pelo acaso, processos humanos, contextuais, de repetição, eletrônicos, de identidade e tempo; *performance*, acrescentando as ideias de tarefas, elemento do jogo, as regras e suas interpretações subjetivas, o instrumento total, música/silêncio, relativização de quem é o executor; *escuta*: com o conceito de *focus* e a constatação da total presença de sons ao nosso redor. Cox e Warner (2004, p. 207-208) chamam atenção para a filiação o experimental à contracultura dos anos 1960, o que o diferencia das vanguardas, ligadas a estruturação musical altamente sofisticada do serialismo. Esse autor também destaca como essa estética experimental influenciou o pop, o rock, o hip hop e a música eletrônica dos dias atuais.

Dessa forma, é possível delimitar a estética experimental como aquela que relativiza as condições tradicionais europeias do estatuto da obra de arte e do autor, usa diferentes meios para composição e privilegia a criação artística musical não como a finalidade de criar um objeto, mas sim de desenvolver um processo.

A seguir, observa-se, por meio da análise dos dados, quais as diferenças e similitudes entre essa definição fornecida pela literatura e o mapeamento das descrições de cada comunidade experimental pesquisada.

5. Análise dos dados e resultados

Inicialmente fez-se um mapeamento de todas as comunidades que incluem no título a palavra “experimental”, suas definições e os números de faixas, contribuintes e membros de cada grupo. Dessa forma, ao observar o número de usuários que se identificam com essas categorias, é possível delinear numericamente quais as ideias que caracterizam o termo “experimental” nesse contexto atraem mais usuários, podendo estes compartilhar faixas ou apenas tornarem-se membros. A seguir estão alguns exemplos:

Tabela 1. Exemplos dos grupos, suas definições, número de contribuintes e faixas compartilhadas

Nome	Definição	Contribuintes	Membros	Faixas
Mad Experimental ³	A group for anything and everything that doesn't fit into a well-defined genre.	38	67	87
ethno-experimental ⁴	a world fusion of all our collective pasts, nows & possible futures...	73	167	166
Experimental Electronic ⁵	group dedicated to the testing of experimental electronic music in all its forms of expression and artistic.	581	1121	2190

Observa-se que, em sua peculiaridade, cada definição destaca uma ou outra ideia sobre o que é música experimental. As ideias mais recorrentes foram consideradas como campos semânticos, que seriam atribuídos às descrições década grupo. Ou seja, os grupos de filiam às categorias semânticas pelo conteúdo presente na descrição, classificador pelo pesquisador.

Dessa forma, é possível, a partir do mapeamento de mais de 70 comunidades que contém a palavra-chave “experimental”, sintetizar os campos semânticos que cada definição destaca.

Os campos semânticos criados para a análise são os seguintes:

- (1)*experimental como liberdade* (sem regras, criação com liberdade);
- (2)*experimental como não gênero* (definição pela impossibilidade de classificação em um gênero já consolidado);
- (3)*experimental como diversidade* (se define pela pluralidade, abre espaço para “vários”, “quaisquer” tipos de música);
- (4)*experimental como vanguarda* (utilizam a palavra “vanguarda”; falam de quebra de “limites” e “fronteiras”)
- (5)*experimental como subgênero* (dentro de cada gênero há um subgênero experimental)
- (6)*experimental como um gênero* (consolidado, geralmente não trazem uma definição);
- (7)*experimental como algo “estranho/diferente”* (aqui se englobam definições que trazem ideias como “novo”, “diferente”, “estranho”);
- (8)*sem definição* (grupos que não apresentam definição).

Na tabela acima, por exemplo, as definições desses três grupos foram respectivamente categorizadas como estando dentro dos seguintes campos semânticos: *i)* não gênero, *ii)* diversidade e *iii)* diversidade e gênero consolidado.

Após observar os gráficos gerados pela coleta e classificação das definições, apresenta-se uma síntese do termo experimental utilizado na rede social Soundcloud, comparado ao resultado da pesquisa com as definições dadas pela revisão de literatura. Montando um gráfico com base no número de ocorrência dos “campos semânticos” definidos, tem-se o seguinte gráfico:

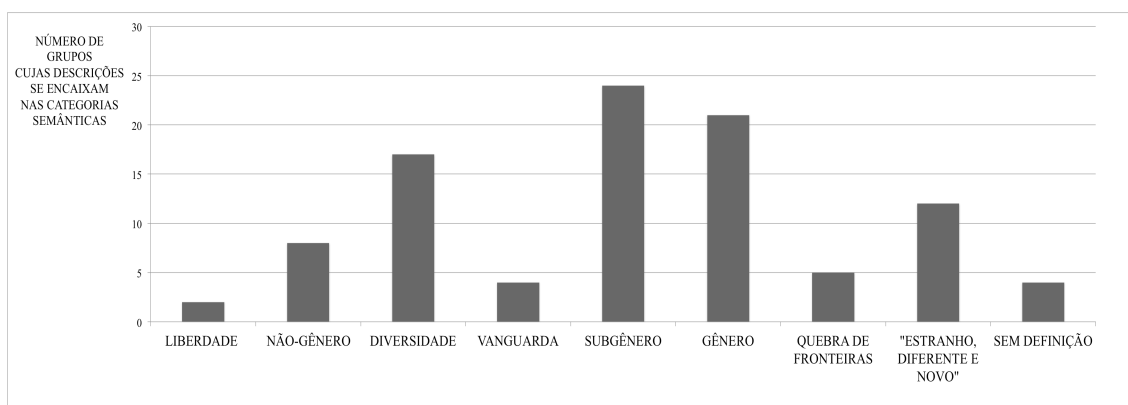


Figura 2. Número de grupos cujas definições se encaixam em cada grupo semântico.

Nota-se então que, conforme vê-se no gráfico, os campos semânticos aos quais as definições de cada grupo mais se associam são: nesta ordem, (5) *experimental como subgênero*, (6) *experimental como gênero consolidado* e (3) *experimental como diversidade*. Poucos são os grupos que abrem mão da descrição, o que indica uma preocupação conceitual por parte dos membros.

Ao observar não as descrições em função do número de grupos, mas sim as descrições das comunidades que mais contêm faixas compartilhadas, vê-se que aquilo que se caracteriza como experimental está nos campos semânticos da (3) *diversidade*, seguida pelo (2) *não gênero* e pelo da (4) *vanguarda* e do (7) *diferente*.

Observa-se também que os grupos que mais atraíram faixas foram aqueles cujas descrições estão, em ordem, no campo semântico da (3) *diversidade* e do (6) *experimental como um gênero*. Seguido pelo campos semântico do (2) *não gênero*, e depois o da (4) *vanguarda* e do (7) *diferente*.

6. Considerações finais

A partir da análise dos dados coletados nos grupos virtuais (grupos da rede social Soundcloud), conclui-se que não há uma definição clara e única para a estética experimental nesse contexto. Como já concebido pela própria natureza das redes, lida-se aqui com um conteúdo fluído, em constante transformação (cf. LÉVY, 1999). Por isso, são muitas as descrições desses grupos de teor não excludentes, inclusivos, que abrangem um amplo espectro de manifestações musicais.

Observa-se que há uma variação nas descrições oferecidas pelo usuários do experimental entendido como um gênero consolidado – quando assim tratado pelas comunidades, faltam descrições que o caracterizam como tal – e como um subgênero, como se toda manifestação musical (seja rock, pop, sinfônica, minimalista) tivesse o seu viés experimental. Conclui-se então que, nesse contexto, o termo experimental pode se referir a essas duas acepções. A ideia do experimental como algo de vanguarda, que quebra limites e barreiras, possibilitando liberdade, também está presente nessas comunidades, além da comum referência a esse tipo de música como “diferente” e “nova”.

Após observar a maneira como o conceito de experimental é utilizado nas comunidades virtuais, é possível indicar que há certa distância entre a acepção do

termo na academia e nos escritos teórico-musicais. Na rede social de música investigada, a ideia de música como processo fica esquecida. E aquele movimento de constante reflexão sobre o que é uma obra musical dá espaço para um conceito de experimental como algo diferente, novo, definido, na maioria das vezes, como um subgênero ou como uma classificação para criações que não se encaixam bem nos gêneros já consolidados.

Esta é uma pesquisa inicial que faz parte de um todo de pesquisas sobre as relações entre música e mídias no Programa de Pós-Graduação Música em Contexto da Universidade de Brasília. Espera-se que aqui tenha-se alcançado, além dos resultados presentes na análise dos dados, mais um recurso metodológico para investigar as práticas musicais imersas no mundo virtual da rede mundial de computadores.

Referências

- Bardin, L. 1977. *Análise de conteúdo*. São Paulo: Edições 70.
- Boyd, D. M.; Ellison, N. B. 2007. Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication*, v. 13, n. 1. Disponível em: <<http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>>. Acesso em 20 dez.
- Caroso, L. 2009. Por uma etnomusicologia no ciberespaço. In: CONGRESO DE LA CIBERSOCIEDAD, 4.
- Cox, C.; Warner, D. (Eds.). 2004. *Audio culture: readings in modern music*. New York: Continuum.
- Eno, B. Foreword. In: Nyman, Michael. 2009. *Experimental music: Cage and beyond*. Cambridge: Cambridge University Press.
- Ferreira, P. P. 2004. O analógico e o digital: a politização tecnoestética do discurso dos DJs. In: XXIV REUNIÃO DA ANPOCS. Caxambu, MG. *Anais...* Caxambu, MG.
- Garrido, A. G. 2005. Globalización y comunidades virtuales a través del fenómeno de Internet y los programas de P2P. In: CONGRESO INTERNACIONAL DE LA SOCIEDAD IBÉRICA DE ETNOMUSICOLOGÍA. 8. Zaragoza, 2005. *Actas...* Institución Fernando El Católico (CSIC), Zaragoza.
- Hine, Christine. 2000. *Virtual Ethnography*. London: Sage.
- Katz, Mark. 2004. *Capturing Sound: How Technology Has Changed Music*. Berkeley: University of California Press.
- Lévy, Pierre. 1999. *Cibercultura*. São Paulo: Editora 34.
- Nyman, Michael. 2009. *Experimental music: Cage and beyond*. Cambridge: Cambridge University Press.

Pinheiro, Chico. 2009. Mudança dos ventos à vista, In: O futuro da música depois da morte do CD, Organizado por I. F. Perpétuo e S. A. Silveira, . Momento Editorial, São Paulo.

1 Disponível em: <www.soundcloud.com/groups>. Acesso em 8 dez. 2011

2 Coleta de dados realizada entre os dias 7 e 11 de dezembro de 2011.

3 Disponível em: <www.soundcloud.com/groups/mad-experimental>. Acesso em 8 dez. 2011

4 Disponível em: <www.soundcloud.com/groups/ethno-experimental>. Acesso em 8 dez. 2011

5 Disponível em: <www.soundcloud.com/groups/experimental-music-1>. Acesso em 8 dez. 2011

Influência do Gênero Musical na Aquisição de Músicas Digitais

Irapuru H. Flório¹, Daniel F. Sampaio¹, Maria A. Cunha², Miriam S. Mazo²

¹Programa de Pós-Graduação em Informática

²Programa de Pós-Graduação em Administração

Pontifícia Universidade Católica do Paraná

Rua Imaculada Conceição, 1155 CEP 80.215-901

Curitiba – PR – Brazil

{irapuru,daniel.sampaio}@ppgia.pucpr.br, {alexandra.cunha,miriam.mazo}@pucpr.br

Abstract: *The importance and contribution of music in digital format in the entertainment industry makes it one of the main media consumption in this segment. The way that a music title is chosen, processed and the influence of gender in this procedure are study objects of this work. The important factors for the choice of the songs are addressed in experiment questionnaire with heterogeneous groups of people from the internet community. A careful analysis is used to verify if the musical genre is a determinant factor where in the choice of unknown music titles and displayed randomly.*

Resumo: *A importância e participação da música em formato digital na indústria do entretenimento mundial fazem dela uma das principais mídias de consumo neste segmento. A forma de escolha de um título musical e a influência do gênero associado a esse título são objetos de estudo deste trabalho. Os fatores preponderantes para a escolha das músicas são abordados experimentalmente, por meio da aplicação de questionários com pessoas de grupos heterogêneos da comunidade da Internet. Uma análise criteriosa é empregada para verificar se o gênero musical é um fator determinante na escolha dos títulos musicais desconhecidos e mostrados aleatoriamente.*

1. Introdução

A partir da segunda metade da década de 90, a disseminação de software de compressão de dados abriu possibilidades para o compartilhamento de músicas on-line e para a conversão de dados de música em CDs. De certa maneira, o impacto do desenvolvimento tecnológico neste campo sobre a indústria da música não estava previsto, sendo até subestimado. No entanto, parte da indústria estava aberta às mudanças tecnológicas e começou a observar não só os riscos da Internet, mas também as suas potenciais oportunidades [Dolata, 2011].

Com o advento da Internet, mais especificamente, do comércio eletrônico de músicas digitais, mudou a forma como os clientes fazem aquisição de suas músicas. Esse procedimento tornou-se mais dinâmico, fácil e, de certo modo, mais acessível economicamente pelo ganho em escala. O acesso a produtos de forma *on-line* ou *off-line* desafia as empresas a construir e disponibilizar soluções inovadoras para o acesso e

para a escolha de títulos musicais. Recentes avanços da tecnologia e principalmente de dispositivos móveis podem redefinir as dinâmicas do comportamento dos usuários com relação ao entretenimento.

A aceitação deste canal de venda, o comércio eletrônico, pelo consumidor brasileiro nos últimos anos é verificada pelo 28º relatório da Webshoppers elaborado pela empresa e-bit divulgado em agosto de 2013. O documento aponta que comércio eletrônico faturou R\$ 12,74 bilhões no primeiro semestre de 2013, valor 24% maior que o faturamento registrado no mesmo período de 2012 [E-bit, 2013]. A pesquisa TIC Domicílios e TIC empresas [CGI, 2013], publicada pelo Comitê Gestor da Internet no Brasil, também mostra o crescente uso de computadores e Internet nos lares e empresas brasileiras, bem como um aumento de uso do comércio eletrônico por indivíduos e empresas.

O crescimento populacional e econômico, a expansão da classe média, o aumento do uso de dispositivos e redes sociais são fatores importantes para o mercado de música digital no Brasil. A quantidade de músicas gravadas cresceu 11,2% no primeiro semestre de 2012 e 8,6% no mesmo período de 2011. As vendas digitais decolaram em 2012, após o lançamento do iTunes no final de 2011, com as receitas praticamente dobradas em relação aos 12 meses anteriores. A renda dos direitos autorais também tem crescido no Brasil, tendo um aumento de 29,6% em 2012, a maior taxa dos últimos oito anos, conforme relatório da Digital Music Report (2013).

O gênero musical está associado à música digital. Ele é, talvez, o descritor mais óbvio que vem à mente e a forma mais utilizada de descrição e identificação de uma música. A escolha de gêneros musicais é um problema que pode ser inserido em diferentes contextos, desde a escolha de uma música para ouvir em seu computador até o desenvolvimento de uma aplicação comercial [Lee e Downie, 2004][Pampalk, et al 2002]. Este artigo investiga o Gênero musical como fator de escolha de músicas digitais e a pesquisa quer resposta à questão: “Quando há informação sobre o gênero musical, diminui o tempo dedicado à escolha de músicas?”. Pressupõe-se que o gênero seja um elemento decisivo na escolha de músicas digitais, diminuindo o tempo dedicado à escolha. Para comprovar esta afirmação, foi realizado um experimento em setembro de 2013, por meio da aplicação de um questionário on-line. Utilizou-se a escala de marketing de Novak e Hoffman (2009), para investigação do tipo de estilo a que a decisão de escolha de uma música está ligado. O artigo está organizado como se segue. A seção 2 apresenta a revisão de literatura sobre o estilo de pensar numa situação de escolha específica, a *Situation-Specific Thinking Style*, e a seção 3 descreve a metodologia aplicada. Os resultados experimentais obtidos são detalhados na seção 4. As conclusões e perspectivas de trabalhos futuros são apresentadas na última seção.

2. A escala *Situation-Specific Thinking Style* e sua aplicação ao contexto da escolha de uma música na web

Anos de pesquisa demonstram fortes evidências de que os consumidores processam a informação de duas maneiras distintas e qualitativamente diferente, denominadas racional e experimental [Epstein 1994, 2003; Hogarth 2005; Kahneman e Frederick 2002; Sloman 1996; Smith e DeCoster 2000]. No entanto, pouca pesquisa tem medido explicitamente como a situação influencia diretamente o estilo de pensamento e seu impacto e também não houve tentativas de medir simultaneamente as duas dimensões

de pensamentos específicas para cada situação e validar esta medida em um contexto mais amplo.

Acredita-se que o conhecimento significativo em relação ao comportamento real do consumidor pode ser obtido medindo o estilo de pensamento usado em uma determinada atividade do consumidor como uma medida do processo. Os resultados contribuem para a literatura sobre o ajuste, congruência e efeitos que envolvem estilos de pensamento, e também defendem a situação específica do estilo de pensamento que está sendo medido de forma rotineira, não só para o controle de manipulação, mas também como uma variável de processo quando se considera o desempenho em uma ampla gama de atividades de consumo [Novak e Hoffman, 2009]. Novak e Hoffman (2009), desenvolveram uma proposta de estilo de pensamento, *Situation-Specific Thinking Style - STSS*, adotado por um consumidor numa situação específica. Há dois estilos de pensar que operam de maneiras diferentes. O primeiro estilo, o “Pensando Experimental” numa situação específica, é associativo, emocional, de baixo esforço, rápido para implementar mas lento para mudar, paralelo, imediatista, orientado a resultados, holístico, pré-consciente e experimentado passivamente, com o processo opaco para o indivíduo. O “Pensando Racional”, por outro lado, é lógico, baseado em causa e efeito e em regras, hierárquico, sequencial, orientado a processo, mais lento para implementar mas mais rápido para mudar, de alto esforço, orientado em direção a ação retardada, consciente e experimentado ativamente, com o indivíduo ciente do controle do processo [Novak e Hoffman, 2009].

Uma comparação detalhada das distinções entre as duas teorias de processo, observa-se que as diferenças entre as teorias podem ser substanciais. Por exemplo, na teoria reflexiva - impulsivo de Strack e Deutsch (2004), o sistema é impulsivo se acionado apenas pela percepção imediata e funciona como uma rede associativa simples de difusão através da ativação. Esta é uma conceituação muito mais limitada se comparada a maioria das outras teorias de duplo processo, uma vez que os resultados complexos, como efeito, a intuição e a emoção são vistos como produtos conjuntos de interações entre os sistemas reflexivos e impulsivo, e não como o produto do sistema experimental. Espera-se que neste trabalho a escala de medida de marketing de Novak e Hoffman (2009), possa produzir elementos que possibilitem a avaliar a influência desta medida na escolha das músicas e a correlação do gênero musical proporcionada pelo experimento desenvolvido.

O novo contexto da distribuição de música eletrônica e a sistemática exploração de grandes bases de dados musicais cria a necessidade de descrições simbólicas de títulos musicais. O gênero musical apesar de ser intrinsecamente mal definido, e as tentativas de defini-lo com precisão têm uma forte tendência de “ficar em círculos”, é certamente um elemento relacionado à classificação, organização ou escolha de músicas. Segundo Aucoeur e Pachet (2003) atribuir um gênero a uma música é realmente uma maneira útil de descrever o que aquele item compartilha com outros itens do mesmo gênero. A gênese do gênero pode, portanto, ser encontrada em nossa tendência natural e irresistível de classificar.

De forma geral, os gêneros musicais são rótulos criados e utilizados por seres humanos para categorizar e descrever o vasto universo da música. Os gêneros musicais não possuem definições rigorosas e limites, pois surgem através de uma complexa interação entre o público, marketing, história e fatores culturais. Esta situação levou

alguns pesquisadores a sugerir a definição de um novo esquema de classificação de gênero musical para fins da *Musical Information Retrieval* (MIR). Os atuais gêneros musicais são determinados por certas características tipicamente relacionados com a instrumentação, estrutura rítmica e conteúdo da música.

Como a Internet é uma significativa e importante fonte de músicas com vários sites dedicados à divulgação, distribuição e comercialização de música. Tornou-se necessário neste contexto, procedimentos automáticos capazes de lidar com grandes quantidades de música em formatos digitais e a MIR tornaram-se uma destacada área de pesquisa da Inteligência Artificial. Uma das tarefas focadas pela MIR é o problema da classificação automática do gênero musical (CAGM). Em gêneros musicais os rótulos categóricos são a sua essência, criados por especialistas humanos, a fim de identificar o estilo da música [Silla, Koerich e Kaestner, 2008].

Fortes tendências apontam que a CAGM será um dos serviços que os fornecedores de distribuição de conteúdo musical vão usar para atrair clientes. Outra indicação da crescente importância da distribuição da música digital é a atenção legal que as empresas estão recebendo pela indústria da música. Isso se deve ao fato de que o gênero musical é um dos argumentos utilizados com mais frequência em buscas de títulos musicais [Tzanetakis & Cook, 2002][Downie, J. S. e Cunningham, 2002].

3. Metodologia

Este trabalho tem como objetivo verificar se a existência do atributo gênero musical está associado ao tempo de resposta do usuário e também verificar a associação com o estilo da tomada de decisão nesta situação específica. É utilizada a escala de marketing *Situation-Specific Thinking Style* (STSS). O estudo desenvolvido fundamentou-se no método experimental, pode-se dizer que é um quase experimento, pois, como efetuado on-line na Web, não se tem controle sobre todas as variáveis. A seguir são apresentados o desenho, o ambiente, os sujeitos experimentais e o controle deste quase experimento.

3.1 Desenho, ambiente, sujeitos experimentais e controle

O delineamento, de inspiração experimental, foi considerado adequado para estudar a problemática proposta. O desenho experimental foi implementado construindo-se um questionário on-line, onde eram apresentadas músicas em dois questionários diferentes. Uma aplicação escrita em PHP apresentava os questionários alternadamente, ou seja, a aplicação mostrava automaticamente o questionário tipo 1 para o primeiro acesso e o tipo 2 para o segundo acesso e assim sucessivamente. A diferença entre os questionários era a informação do gênero musical, o questionário 1 possuía-a, o questionário 2 não. O motivo pelo qual foi colocada ou subtraída a informação foi a necessidade de verificação do tempo de resposta do preenchimento nos dois tipos de questionário.

A tarefa experimental consistiu na seleção de 5 músicas, por se tratar de um número ideal para não demandar um tempo demasiado, possuindo como alternativas 100 trechos de músicas diferentes entre si e, como informações adicionais, atributos relevantes destas: o título da música, o nome de artista, gênero musical (presente somente no formulário 1), álbum a que pertence a música e um link para ouvi-la, em um pequeno trecho. Conforme o Tabela 1, foram utilizados na aplicação do experimento 25 gêneros musicais e aproximadamente 10 mil títulos de músicas distribuídas entre

estes gêneros, extraídas do *Subset do Million Song Dataset (MSD)* produzido pelo laboratório LabRosa da Universidade de Columbia, Nova York, EUA.

Tabela 1 - Gêneros e Títulos Musicais

Gêneros Musicais	Títulos de Musicas	Média de Título x Gêneros
25	10241	410

Também foram coletadas informações sobre o respondente e sobre a dimensão racional ou experimental do seu estilo nesta situação específica de seleção de música. A Tabela 2 apresenta as 22 questões, na ordem em que surgiam no questionário.

Tabela 2 – Conteúdo do Questionário

Campo	Descrição
Nome	Nome da pessoa que respondeu ao questionário (Opcional)
Idade	Idade da pessoa (em anos)
Escolaridade	Formação escolar (ensino médio, graduação ou pós-graduação)
Sexo	Sexo da pessoa (masculino ou feminino)
Música	Nome da Música
Artista	Nome de artista
Gênero	Gênero musical (presente somente no formulário 1)
Álbum	Álbum a que pertence a música
Preview	Link para ouvir a música
Perguntas 1 a 10 (ver Tabela 3)	Dimensão racional
Perguntas 11 a 20 (ver Tabela 3)	Dimensão experimental
Pergunta 21	Dificuldades para a escolha da musica
Pergunta 22	Por favor, descreva a forma como você escolheu as 5 músicas
Duração	Tempo de resposta do questionário (gerado automaticamente pela aplicação, em quantidade de segundos)

A primeira tela do questionário, na web, apresentava as instruções e as questões relacionadas às informações pessoais do respondente. A segunda tela, ou segunda etapa, apresentava 100 títulos de músicas de gêneros musicais variados, contendo os seguintes dados: título, artista, álbum, gênero (formulário1) e um link para ouvir um trecho da música, cuja duração varia de 30s a 60s (Figura 1).



Figura 1 - Tela de Opções de Escolha de Músicas

A escolha de 5 músicas desconhecidas deve-se ao fato de forçar e eliminar o aspecto emocional. Estas músicas foram selecionadas aleatoriamente de uma de uma base resumida de 10 mil músicas da MSD [Bertin-Mahieux et al, 2011]. A motivação para que o usuário escolhesse as músicas foi a pergunta: “Você deve escolher cinco músicas desconhecidas para tocar na festa de um amigo”. A situação “festa” mencionada procura avaliar a escolha dentro do contexto do grupo social.

A terceira parte do questionário compreendia perguntas (Tabela 3) elaboradas de acordo com a escala STSS para verificação do estilo de pensamento nesta situação de escolha das músicas, nas dimensões racional ou experimental. As respostas foram atribuídas numa escala de 1 a 5 com os significados: (1) Definitivamente Falso, (2) Quase Falso, (3) Indeciso, (4) Quase Verdadeiro e (5) Completamente Verdadeiro [Bearden e Netemeyer, 2010]. Adotou-se manter a escala original por ter sido utilizada e testada em outro contexto.

Tabela 3 - Dimensões segundo a escala de Novak e Hoffman (2009)

Dimensão	Estímulos do questionário
Racional	<ol style="list-style-type: none"> 1. Eu fundamentei as respostas cuidadosamente 2. Eu respondi esta pesquisa de forma sistemática 3. Eu respondi esta pesquisa de forma lógica 4. Eu respondi esta pesquisa de forma analítica 5. Eu estava muito concentrado(a) nas etapas desta pesquisa 6. Eu utilizei regras precisas para selecionar as respostas 7. Eu estava muito concentrado(a) no que estava fazendo para chegar às respostas 8. Eu estava muito ciente da forma como eu estava pensando 9. Eu cheguei nas minhas respostas analisando cuidadosamente as informações que me foram passadas 10. Eu usei regras claras
Experimental	<ol style="list-style-type: none"> 1. Eu usei a minha intuição 2. Eu respondi o que me fez sentir bem 3. Eu confiei em meu palpite 4. Eu me apoiei no meu senso de intuição 5. Eu me apoiei nas minhas impressões 6. Eu usei os meus instintos 7. Eu usei o meu coração para me guiar nas escolhas 8. Eu tive momentos de reflexão 9. Ideias surgiram na minha cabeça 10. Cada ideia me levava à próxima

Participaram da pesquisa como sujeitos experimentais pessoas que têm acesso à internet e presentes nas listas da rede social do Facebook e das listas de e-mails dos pesquisadores. O tempo em que o questionário permaneceu aberto foi de quatorze dias consecutivos, em agosto e setembro de 2013 e houve 192 respostas, das quais 169 foram consideradas válidas.

Antes de se iniciar o processo experimental, foi realizada a validação do questionário, com os alunos de uma turma de pós-graduação em Administração e Informática. O questionário também recebeu as sugestões de uma professora de Gestão de TI e outra de Marketing. A partir dos resultados dessa validação foram realizados ajustes para aplicação do questionário. Fizeram parte do controle experimental a aleatoriedade na alocação dos questionários 1 e 2, as instruções por escrito aos indivíduos no início do questionário, e a duração de somente 14 dias para a realização de todo o preenchimento.

3.2 Análise dos Dados

De posse dos dados coletados realizou-se uma validação dos questionários, antes da fase de processamento dos resultados. Foram descartados os casos encerrados em menos de 40 segundos e acima de 850 segundos, e aqueles que, na segunda etapa, assinalaram a mesma resposta a todas as perguntas. Para a análise estatística utilizou-se o software estatístico da IBM *Statistical Package for the Social Sciences* (SPSS), para o processamento da estatística descritiva, verificação da normalidade, correlações e testes estatísticos das amostras, descritos a seguir [Corder e Foreman, 2009].

4. Resultados

Após a análise dos dados no aplicativo SPSS, com a aplicação de estatística descritiva descreve-se neste item a amostra e discutem-se os resultados.

4.1 Descrição da Amostra

Conforme segue abaixo na (Figura 2 e Tabela 4) percebe-se que a distribuição das idades dos usuários participantes do experimento assim como o tempo de resposta para preenchimento dos questionários são bem heterogêneos.

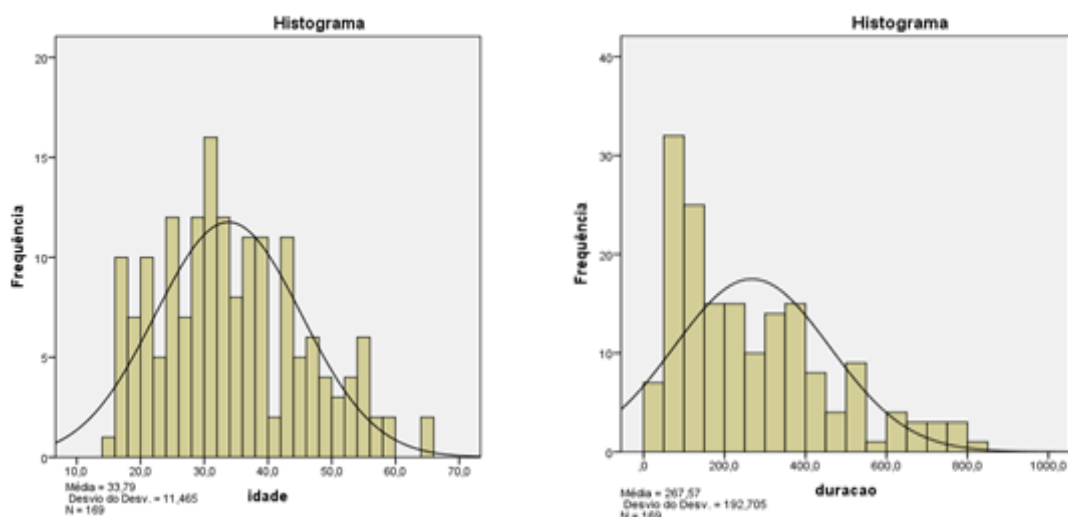


Figura 2. Distribuição da Idade dos Usuários e Duração do Tempo de Resposta

A distribuição dos usuários que responderam por gênero de sexo é também distribuído desigualmente. Para o sexo feminino houve uma frequência de 65 correspondendo a 38,5% da amostra e para o sexo masculino um frequência de 104 correspondendo a 61,5% da amostra.

Tabela 4 - Estatística descritiva Idade x Tempo Resposta

Campo	N	Mínimo	Máximo	Média	Desvio padrão
Idade	169	15	65,0	33,79	11,46
Tempo de Resposta (s)		42	839	267,57	192,70

4.2 Discussão

O gênero musical mostrou-se no experimento como decisivo na escolha do título das músicas. Foi perguntado aos respondentes como eles haviam feito sua escolha. Conforme a questão 22 do questionário, o gênero musical tem um percentual de 27,2 em relação ao total de 169 ocorrências (Tabela 5). Remarca-se que a metade dos questionários não possuía o gênero musical, já que esta era a variável independente.

Tabela 5 - Estatística descritiva da escolha x Motivo

Motivo da Escolha	Com e Sem Gênero		Formulário Com Gênero		Formulário Sem Gênero	
	Quantidade Ocorrências	Percentual (%) de Ocorrências	Quantidade Ocorrências	Percentual (%) de Ocorrências	Quantidade Ocorrências	Percentual (%) de Ocorrências
Função	6	3,6	2	1,2	4	2,4
Artista	12	7,1	4	2,4	8	4,7
Título	12	7,1	6	3,6	6	3,6
Outros	14	8,3	4	2,4	10	5,9
Aleatória	25	14,8	11	6,5	14	8,3
Gênero	46	27,2	37	21,9	9	5,3
Indeterminado	54	32,0	23	13,6	31	18,3
Total	169	100	87	51,5	82	48,5

A distribuição da duração do tempo para as respostas dos formulários sem gênero e com gênero demonstrado pelo gráfico de dispersão (Figura 3) mostrou que a média da duração dos tempos de respostas dos usuários que tinham a opção de gênero na tela foi de 263,2 segundos sendo maior que da opção contrária que foi de 242,8 segundos. Não há indicação de que o tempo de decisão sobre a música tenha diminuído em função da existência do gênero musical na tela.

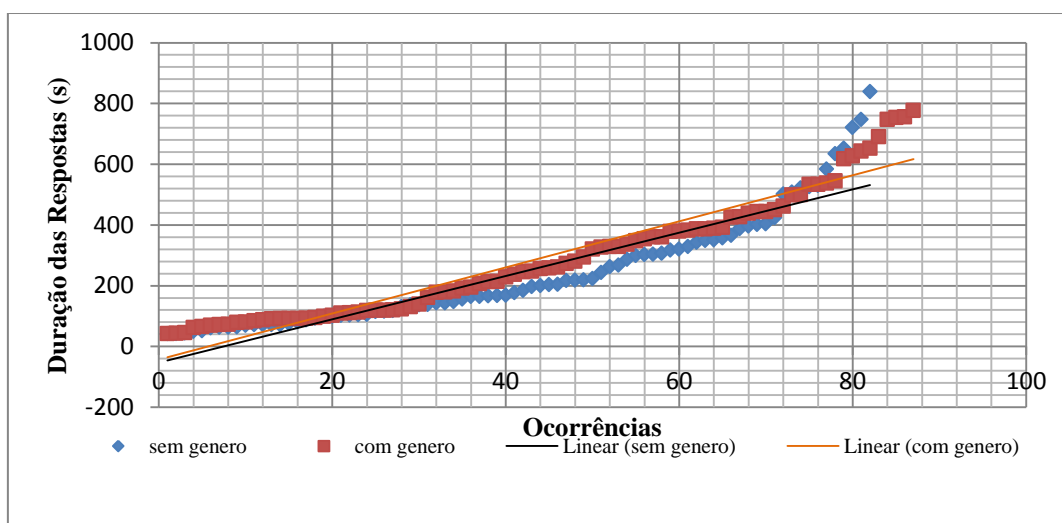


Figura 3 - Distribuição dos Tempos de Resposta x Tipo de Formulário

Outra situação visualizada com a forma do preenchimento dos questionários foi a correlação do tempo de resposta com a escala de marketing STSS, demonstra-se pelo gráfico da (Figura 4) que a forma de pensamento racional e experimental quase não houve variação entre a média dos dois tempos 276,1 segundos e 259,4 segundos respectivamente.

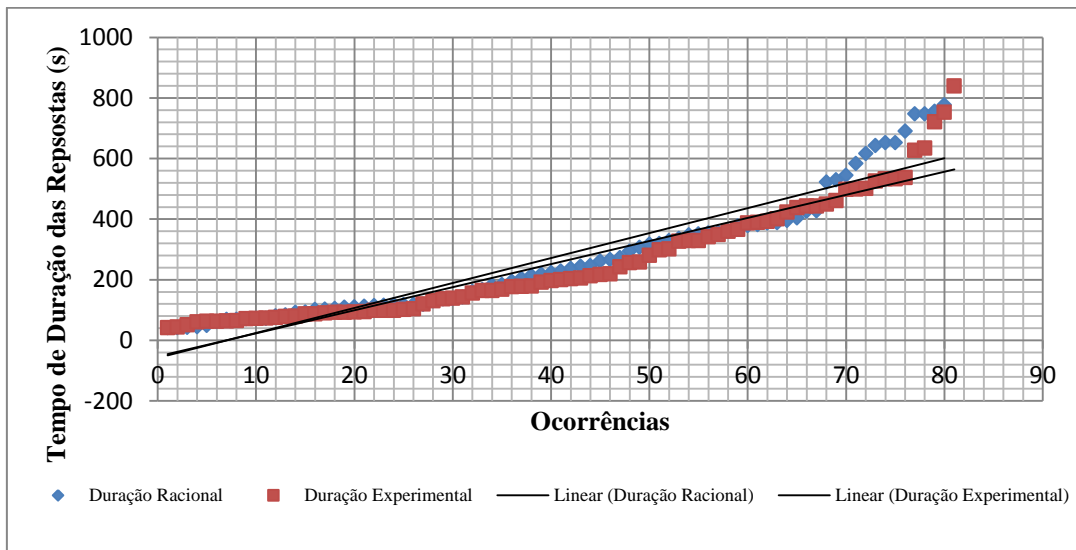


Figura 4 - Distribuição dos Tempos de Resposta x Tipo de Pensamento

Mas, quando se desdobra os dados dos usuários que responderam com e sem a opção de gênero musical em relação ao tipo de pensamento, experimental ou racional, e a duração do tempo de resposta, verifica-se que:

- a) O tempo de resposta dos usuários sem a opção de gênero e utilizando um pensamento experimental é menor, ver (Figura 5); e

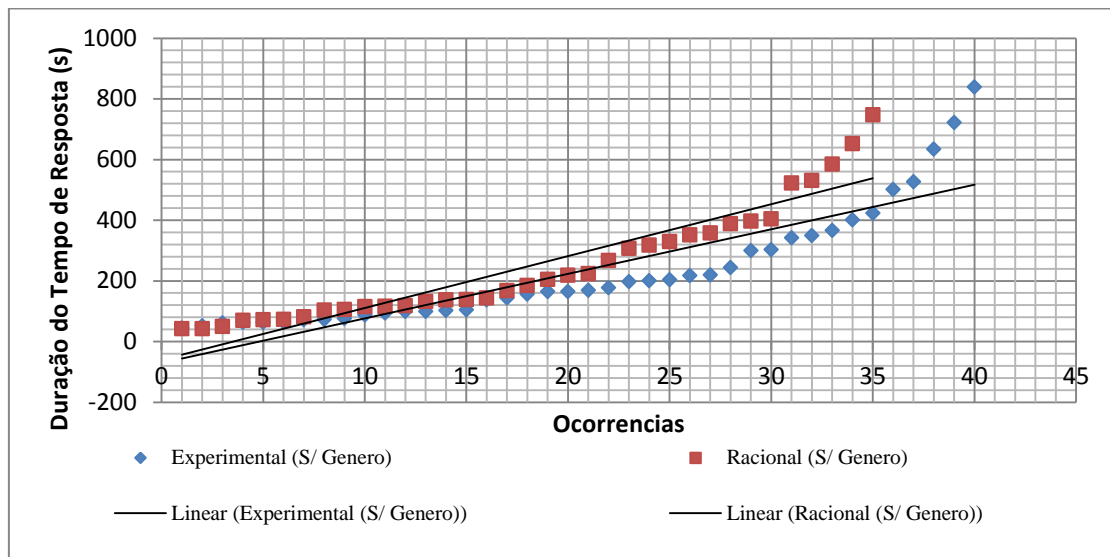


Figura 5 - Distribuição dos Tempos de Resposta Sem o Gênero x Tipo de Pensamento

- b) Os usuários que escolheram as músicas por um pensamento racional, ao terem a opção de gênero disponível, decidem, no entanto num tempo menor que os demais (Figura 6).

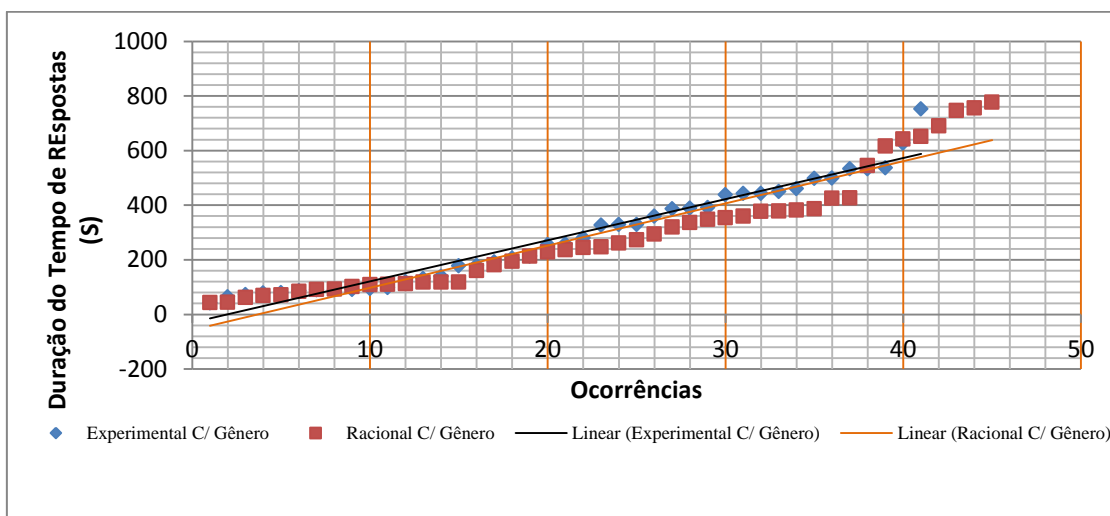


Figura 6 - Distribuição dos Tempos de Resposta Com o Gênero x Tipo de Pensamento

5. Conclusão

É possível afirmar que o número de pessoas que acessam a internet em sua residência e trabalho aumenta a cada ano e que aumenta também o acesso ao comércio eletrônico [CGI, 2013]. Os consumidores no Brasil encontram um novo canal para realizarem suas compras. Este trabalho, um quase experimento, utilizou amostragem por adesão, e na amostra obteve-se uma elevada quantidade de jovens, 3,4% dos respondentes estão entre 15 e 40 anos. No relatório e-bit [2013], é preponderante a participação deste público no comércio eletrônico. Existe um enorme potencial de aquisição de músicas digitais para este consumidor e há oportunidade para concentração de esforços no segmento.

Quis-se investigar se o gênero musical é um elemento decisivo na escolha de músicas digitais, diminuindo o tempo dedicado à escolha. Pode-se concluir, ao final deste trabalho, que há evidências de que a existência do gênero musical entre as informações disponíveis facilitou o processo da escolha das músicas, quando o usuário se depara com títulos e obras desconhecidas. As pessoas que, ao escolherem músicas, adotam um “pensar experimental”, escolhem mais rápido do que as que adotam um “pensar racional”.

No entanto, com a informação de gênero disponível, inverte-se, o pensamento racional torna-se mais rápido na decisão. Como trabalhos futuros deve-se pesquisar mais as opções de escolhas de músicas com outras ferramentas de escala de marketing comparativamente a fatores determinantes nestes processos.

Referências

- Aucouturier, J. e Pachet, F. (2003) Representing Musical Genre: A State of the Art. *Journal of New Music Research*, vol. 32, no. 1, pp. 83-93.
- Bearden, W. O. e Netemeyer, R. G. (2010) *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Sage Publication.

- Bertin-Mahieux, T. Ellis, Daniel P.W. Whitman, B. e Lamere, P. (2011) The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011).
- CGI. TIC domicílios e TIC empresas (2013). Pesquisa sobre o uso das Tecnologias da Informação e Comunicação no Brasil. Acesso em: 09/09/2013. Disponível em: <<http://op.ceptro.br/cgi-bin/cetic/tic-domicilios-e-empresas-2012.pdf>>.
- Corder, Gregory W. e Foreman, Dale I. (2009) Nonparametric Statistics for Non-Statisticians: A Step-by-Step. Approach Published by Wiley.
- Dolata, Ulrich. (2012) The Music Industry and the Internet: A Decade of Disruptive and Uncontrolled Sectoral Change. University of Stuttgart. Stuttgart – DE 2011.
- Digital Music Report (2013) 10º Edition IFPI (International Federation of the phonographic Industry) Acesso em: 09/09/2013. Disponível em: <<http://www.ifpi.org/content/library/DMR2013.pdf>>.
- Downie, J. S. e Cunningham, S. J. (2002) Toward a theory of music information retrieval queries: system design implications. Proc. 3rd International Conference on Music Information Retrieval, pp. 299-300, 2002.
- E-bit (2013) 28º Edição do Relatório de Comércio Eletrônico. Acesso em: 28/8/2013. Disponível em: <<http://img.ebit.com.br/webshoppers/pdf/WebShoppers2013.pdf>>.
- Epstein, S. (1994) Integration of the cognitive and the psychodynamic unconscious. Journal American Psychologist, Vol. 49(8), Aug 1994, 709-724.
- Hogarth, Robin M. Betsch, Tilmann; Haberstroh, Susanne (2005). Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought. The routines of decision making. The routines of decision making. , (pp. 67-82). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, xxx, 392 pp.
- Kahneman, D. e Frederick. S. (2002), “Representativeness Revisited: Attribute Substitution in Intuitive Judgment.” In. T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment. New York: Cambridge University Press.
- Lee, J. H. and Downie, J. S. (2004) Survey of music information needs, uses, and seeking behaviours preliminary findings. Proc. 5th International Conference on Music Information Retrieval, Barcelona, Spain, pp. 441-446.
- McKay, C., e I. Fujinaga. (2006). Musical genre classification: Is it worth pursuing and how can it be improved? Proceedings of the International Conference on Music Information Retrieval. 101–6.
- Novak, Thomas P. e Hoffman, Donna L. (2009). "The Fit of Thinking Style and Situation: New Measures of Situation-Specific Experiential and Rational Cognition," Journal of Consumer Research, University of Chicago Press, vol. 36(1), pages 56-72.

- Pampalk, E. Rauber, A. e Merkl, D. (2002) Content-Based organization and visualization of music archives. Proc. ACM Multimedia, Juan-les-Pins, France, pp. 570-579.
- C.N. Silla Jr., A.L. Koerich & C.A.A. Kaestner. (2008) A Machine Learning Approach to Automatic Music Genre Classification. Journal of the Brazilian Computer Society, Vol.14, No.3, pp.7-18, September 2008. ISSN 0104-6500.
- Sloman, Steven A. (1996), "The Empirical Case for Two Systems of Reasoning," Psychological Bulletin, 119 (1), 3-22.
- Smith, Elliot R. e Jamie DeCoster (2000), "Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems," Personality and Social Psychology Review, 4 (2), 108-131.
- Strack, F. e R. Deutsch (2004), "Reflective and Impulsive Determinants of Social Behavior" Personality and Social Psychology Review, 8, 220-247.
- Tzanetakis G., e Cook, P. (2002) Musical genre classification of audio signals Speech and Audio Processing, IEEE transactions on 10 (5), 293-302.

Avaliando Interfaces Gestuais para Prática de Instrumentos Virtuais de Percussão

Jaziel Vitalino Souza e Silva¹, Filipe Calegario¹, Giordano Cabral², Geber Ramalho¹

¹ Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7851 – 50.732-970 – Recife – PE – Brasil

{jvss,fcac,glr}@cin.ufpe.br

² Departamento de Informática e Estatística (DEINFO) – Universidade Federal Rural de Pernambuco (UFRPE)

Rua Dom Manoel de Medeiros, 52.171-900 – Recife – PE – Brasil

{giordanorec}@gmail.com

Abstract. *In this paper, we analyze and evaluate the interaction of a set of gestural interfaces in order to identify technical limitations and to correlate them with the low adoption of these technologies in practice. Through literature researches, we made a mapping of gestures used in percussion instruments, and conducted an experiment where professional drummers evaluated the interaction offered by two interfaces: the Microsoft Kinect Sensor and the Nintendo Wii remote.*

Resumo. *Neste trabalho analisamos e avaliamos a interação de algumas interfaces gestuais atuais, no intuito de identificar limitações técnicas e de que forma estas explicam a baixa adoção de tais tecnologias na prática. Através de pesquisas bibliográficas, fizemos um mapeamento dos gestos mais utilizados nos instrumentos de percussão e realizamos um experimento onde profissionais percussionistas avaliaram a interação oferecida por duas interfaces: o Sensor Microsoft Kinect e o Nintendo Wii remote.*

1. Introdução

No século XX, fatores como o avanço do conhecimento sobre aspectos físicos e cognitivos do som, o advento dos instrumentos elétricos e a aplicação das tecnologias eletrônica e digital na geração sonora artificial mudaram a relação entre música e tecnologias eletrônicas e digitais (Iazzetta, 1997). Pesquisas acerca do desenvolvimento de novas interfaces digitais têm crescido principalmente nas últimas décadas e isto pode ser comprovado com o crescimento de conferências como o NIME (New Interfaces for Musical Expression). Instrumentos virtuais, construídos a partir de acelerômetros e câmeras com sensores de profundidade, permitem a interação do usuário através de gestos e a produção do som através de algoritmos de síntese sonora. Boulezz (1977) menciona que estes dispositivos não foram criados essencialmente para aplicações musicais, mas a sua utilização leva a mudanças que podem resultar em novas soluções.

Apesar de tanta tecnologia empregada na criação de novos instrumentos, por que não é frequente vermos músicos profissionais utilizando essas tecnologias em performances reais? Por que a maioria dos músicos não usam as novas interfaces disponíveis? Esta é a questão que esta pesquisa se dispõe a investigar.

2. Interfaces Digitais

Os instrumentos musicais digitais (*digital musical instruments*, DMIs) são sistemas digitais compostos por um módulo de controle de entrada, que são interfaces que capturam a ação do músico, e um módulo de produção sonora, que efetivamente realiza a síntese sonora. Fazendo a ligação entre a entrada e a saída dos DMIs, temos as estratégias de mapeamento.

Dentre os DMIs mais conhecidos, estão os que simulam instrumentos acústicos existentes, como por exemplo: um saxofone MIDI; uma bateria eletrônica, geralmente formada por um conjunto de *pads*, montados sobre um *rack*, na mesma disposição de uma bateria convencional.

Entretanto, há também DMIs com dispositivos de entrada alternativos, que não se baseiam em instrumentos existentes (MIRANDA e WANDERLEY, 2006, p. 19). Tais DMIs não possuem padrão de tamanho físico ou forma, podendo ser desenvolvido para funcionar em iPads ou Smartphones, softwares e sintetizadores que simulam teclados, violões, guitarras, contrabaixos e outros instrumentos virtuais. Por exemplo, o Reactable¹ é uma interface em que o usuário interage através de objetos do mundo físico (blocos). Tais objetos possuem diferentes funções na execução musical, podendo também ser utilizados por vários usuários ao mesmo tempo, como exibido na Figura 1.



Figura 1 – Interface do Reactable

3. O Gesto Musical

Pode-se assumir que o gesto é um elemento essencial na execução musical, já que os músicos tocam instrumentos por meio de gestos. Por exemplo, são os gestos dos dedos do violonista que excitam as cordas do violão, provocando ondas sonoras que são amplificadas pelo corpo (caixa acústica) do instrumento, para eventualmente serem depois captadas por microfones ou outros sensores e amplificadas eletronicamente.

A tecnologia de construção de instrumentos sempre levou em consideração: a ergonomia, levando em conta a facilidade ou dificuldade com que os músicos interagem com os instrumentos; e a geração sonora, preocupando-se com a forma com que o som é gerado, ou seja, com a qualidade sonora do instrumento.

A preocupação com gestos acontece não só na criação de instrumentos acústicos, mas também na chamada luteria digital (JORDÀ, 2005). Enquanto os instrumentos acústicos tentam evoluir em qualidade física e timbrística, os instrumentos digitais tentam ganhar espaço, por vezes procurando criar instrumentos musicais com modelos alternativos que busquem novas formas de expressão musical.

¹ Disponível em: <<http://mtg.upf.edu/project/reactable>>. Acesso em: 06 Ago. 2013.

4. Interfaces Gestuais em Computação

Baseando-se em conceitos da área de Interação Humano-Computador (IHC), Wanderley (2006) preconiza que sistemas de geração sonora controlados por gestos podem ser utilizados em vários contextos da interação musical. Desenvolvedores de software tem dado cada vez mais importância a estudos relacionados a gestos e movimentos do corpo (Corrêa et al, 2005). A motivação para isto é poder “herdar” um grande repertório de gestos que os seres humanos desenvolveram ao longo dos anos fazendo com que a interação deles com o computador seja mais natural.

Temos visto se intensificarem pesquisas sobre interfaces gestuais, com aplicações em diversas áreas. Dentre estas interfaces gestuais, algumas merecem maior destaque por sua popularidade:

- **Tela Multi-toque** - é uma tecnologia de interação que reconhece múltiplos contatos simultâneos na tela e possibilita a manipulação de sistemas através do toque de algum objeto ou dos dedos. Exemplos: Tablets, Smartphones e Mesas Multi-toque.
- **Novos Controladores de Jogos** – O Nintendo Wii trouxe o conceito de Nunchuck, um novo tipo de controlador de jogo (como foram os *joysticks* e *joypads*). É segurado com as mãos e se utilizam os dedos para pressionar os botões de comando, mas a inovação vem por parte do uso de acelerômetros que permitem novas formas de controle no espaço. Exemplos: Wii Remote e PlayStation Move.
- **Câmera de Detecção de profundidade** – Dispositivo que cria novas maneiras de executar tarefas dispensando a necessidade de contato físico com o instrumento de trabalho. Exemplo: Microsoft Sensor Kinect.

O número de possibilidades é imenso e os aplicativos musicais que utilizam tais interfaces surgem em grande quantidade. Mas uma dúvida perdura: por que não é com frequência que vemos músicos profissionais utilizando essas tecnologias em performances musicais? A maioria dos músicos não usam essas novas interfaces. Isto tem a ver com a falta de qualidade, preconceito, aspectos culturais?

Pode haver inúmeros motivos, porém o objetivo desta pesquisa é analisar os aspectos técnicos, identificando as limitações dos sistemas existentes na tentativa de levantar possíveis hipóteses sobre a não utilização destas interfaces mais frequentemente pelos músicos profissionais.

Na próxima seção faremos um mapeamento de gestos instrumentais, procurando relaciona-los com as possibilidades de gestos reconhecíveis por algumas interfaces.












5. Mapeamento de Gestos

Visto que os sensores das interfaces digitais ainda não estão suficientemente refinados e não conseguem captar movimentos curtos com precisão (por exemplo, os dos dedos), neste trabalho foram escolhidos para a análise, alguns instrumentos musicais convencionais que utilizassem gestos mais largos, particularmente os de percussão.

5.1 Gestos Instrumentais

No intuito de relacionar gestos humanos, possivelmente reconhecíveis pelas interfaces e os gestos usados na prática de cada instrumento, foi feita uma tabela descrevendo os gestos de maior relevância, usados na prática de alguns dos instrumentos de percussão (Tabela 1). Os gestos mapeados foram escolhidos por serem utilizados em grande parte dos instrumentos de percussão e possuírem enorme importância na performance instrumental para o músico.

Tabela 1 – Tabela descritiva de instrumentos e gestos

Instrumento Musical	Movimento com o polegar	Movimento com os demais dedos	Movimentos com o punho	Flexão e extensão do antebraço	Movimento de rotação com o antebraço	Movimentos com ombro e braço
Agogô 			×	×		
Alfaia 			×	×	×	×
Berimbau 	×	×	×	×		
Bongô 	×	×	×	×		
Caixa 		×	×	×	×	×
Caxixi 			×	×	×	
Cajon 		×	×	×	×	×
Castanholas 		×	×			
Congas 		×	×	×		
Cuíca 		×	×	×		
Derbake 		×	×	×		
Djembê 		×	×	×		
Ganzá 			×	×		
Maracá 			×	×		
Moringa 		×	×	×		
Pandeiro 	×	×	×	×	×	
Pratos 			×	×		×
Reco-reco 			×	×		
Repique 	×	×	×	×	×	
Sinos 			×	×		
Surdo 			×	×		×
Tabla 	×	×	×	×	×	
Tamborim 			×	×	×	
Timpano 			×	×		
Triângulo 	×	×	×			
Xilofone 			×	×		
Zabumba 		×	×	×		×

5.2 Análise de Captação de Gestos

Abaixo visualizamos uma tabela com a descrição de gestos primitivos humanos e duas interfaces em análise. Foram, assim, mapeados os gestos possíveis de reconhecimento pelas interfaces escolhidas nesse trabalho.

Tabela 2 – Tabela de interfaces

Gestos	Sensor Kinect	Wii remote
Movimentos com o polegar		x
Movimentos com os demais dedos		x
Movimentos com o punho	x	x
Flexão e extensão do antebraço	x	x
Movimentos de rotação com o antebraço		x
Movimentos com ombro e antebraço	x	

Percebemos que há uma demanda de gestos que não são captados pelas interfaces gestuais. Esses gestos são usados na prática de diversos instrumentos de percussão e sem eles os instrumentos perdem características técnicas importantes. Sendo assim, com os resultados descritos de quais gestos mapeáveis são tecnicamente úteis, captáveis pelas interfaces, iremos descrever um experimento onde foi avaliado esse subconjunto de gestos possíveis, nas duas interfaces gestuais.

6. Método

Pretendendo avaliar a interação das interfaces do ponto de vista do usuário, propomos, com base no método de Barbosa (et al, 2011), realizarmos um experimento, onde vários participantes com perfis musicais distintos fossem orientados a usar as interfaces pela primeira vez.

Foram convidados músicos percussionistas profissionais que atuam na área e possuíssem diferentes perfis artísticos, visando obter um resultado mais satisfatório. Os perfis escolhidos foram: estudantes universitários da área erudita, estudantes universitários da área popular e percussionistas autodidatas. Passando pela mesma série de tarefas, cada experimento levou cerca de quarenta e cinco minutos até sua conclusão.

6.1 Coleta de Dados

O recolhimento dos dados foi concluído a partir de três fases:

- Exploração livre: O usuário é orientado a usar a interface por algum tempo sem restrições, à sua própria vontade.
- Exploração guiada: O usuário escuta algumas amostras de áudio com intuito de experimentar possibilidades, inspirando-se na amostra.
- Entrevista semi-estruturada: Nesta fase o usuário é incentivado a discutir sobre a sua experiência realizada com as interfaces gestuais nos momentos de exploração livre e exploração guiada.

6.2 Análise de Dados

Na análise de dados, as informações obtidas no experimento foram examinadas em três eixos: controle temporal – o qual diz respeito à latência do som; variação timbrística – diz respeito às mudanças de coloração sonora oferecidas pelos instrumentos; e controle de dinâmica – que diz respeito ao controle de intensidade do som, mais forte e mais fraco. Nesta análise, utilizamos o método conhecido como análise do discurso (AD).

6.2.1. Análise do Discurso

A análise do discurso possui 5 etapas:

1. **Transcrição** – Neste passo todo material falado é transformado em texto.
2. **Livre associação** – etapa em que as impressões são anotadas para posterior análise.

3. **Discriminação** – Etapa em que todo texto é quebrado em vários pedaços e os elementos mais comuns são reconhecidos, fazendo menção das terminologias usadas na entrevista.
4. **Reconstrução do mundo do usuário** – Etapa em que o analisador constrói uma representação de cada experiência, na perspectiva do usuário com base na lista de elementos mais comuns utilizados e suas relações.
5. **Exame do contexto** – Por final a representação é comparada e analisada.

7. Experimento

O experimento iniciou-se com a apresentação da primeira interface, o *Kinect*, Figura 3. Nele foram mapeados gestos para dois instrumentos percussivos: um Ganzá (tipo de chocalho); e uma Conga (tipo de tambor). O Percussionista recebeu a instrução de que teria um momento livre (*free exploration*) onde ele poderia testar ritmos e reconhecer que gestos poderiam ser utilizados no instrumento virtual. O momento livre tinha duração média de dois minutos. Em seguida o percussionista testou o instrumento sendo acompanhado por um metrônomo em três andamentos diferentes (bpm - batidas por minutos): 90 bpm, 120 bpm e 160 bpm. Logo após o experimento continuou com acompanhamento de uma harmonia servindo como base em ritmos de Baião em 80, 100 e 120 bpm; Frevo em 100, 140 e 160 bpm; Maracatu em 80, 100 e 120 bpm; Sambinha em 80 bpm e um ritmo latino americano 120 e 140 bpm. Após às performances com cada instrumento virtual, era feita uma sessão de comentários pelo usuário, falando sua opinião sobre a interação do instrumento musical usando a interface. O mesmo ciclo se repetiu com todos os instrumentos experimentados no *Kinect*.

Em seguida foi apresentada a segunda interface, os controles Wii, Figura 4. Nele foram mapeados três instrumentos que utilizam baquetas, aproveitando a formato de bastão e os sensores usados na interface. Os instrumentos virtualizados pelo Wiimote foram: uma Caixa e uma Zabumba (Tambores) e um Agogô (Espécie de sino duplo oco). O experimento com o *Wiimote* seguiu o mesmo ciclo de tarefas usadas no *Kinect*.

Na web é possível visualizar a gravação de alguns trechos do experimento, mostrando os usuários usando os instrumentos virtuais nas interfaces gestuais².



Figura 3 – Perspectiva do usuário com o Kinect

² Disponível em: zabumba virtual <<https://www.youtube.com/watch?v=7PfWE-V8lmk>>, conga virtual <https://www.youtube.com/watch?v=pVkJ_5_OyW0>, caixa virtual <<https://www.youtube.com/watch?v=6IZRyFQ7q4M>>, agogô virtual <<https://www.youtube.com/watch?v=tfNfcmkXPA>> - Acesso em: 06 Ago. 2013.



Figura 4 – Usuário usando o Wii Remote

7.1. Análise do Discurso

7.1.1. Transcrição

Como citado anteriormente em uma das etapas do método utilizado, o material de áudio coletado no experimento foi totalmente transcrito e encontra-se disponível na web³.

7.1.2. Discriminação

Percebemos que os participantes não se sentiram a vontade diante das interfaces e mesmo tendo conhecimento em informática e experiência musical no seu instrumento, levaram um bom tempo para se adaptar a ideia de tocar em um modelo virtual de percussão. Ao experimentar pela primeira vez as interfaces, os movimentos comuns, utilizados na prática do instrumento acústico eram lembrados e automaticamente gesticulados, buscando atingir a mesma sonoridade no instrumento virtual. Porém, quando não conseguiam o retorno esperado, era notório o modo como se sentiam perdidos. Ao passar por essa experiência o usuário 2 comentou:

“Me senti como se nunca tivesse tocado na minha vida, parece que nunca toquei e não sei nada sobre música (...) É diferente do contato com o instrumento material, mas a sensação foi essa, assim, estou mais perdido do que tudo aqui num é?”

Isso acontece também, por causa da importância que os músicos dão a presença do corpo do instrumento, porque ao tocar um instrumento convencional, além de ter um retorno auditivo, ele tem retorno tátil, o que provavelmente deixa o usuário sem confiança diante de uma interface que se manipula apenas por gestos no ar. O limite em que o gesto deve parar é o ponto de ataque onde o músico espera atingir. Contudo, para ele, este ponto precisa ser previsível, pois enquanto o músico toca, ele calcula o tempo em que deve soar à próxima nota e o gesto que será utilizado deve fazer um percurso que atinja no tempo previsto, o ponto de ataque. O usuário 4 chama de instrumento imaginário e ainda salienta:

“Porque é aquela questão, aqui tá a cadeira, eu calculo, automaticamente minha vista calcula onde eu vou atacar e no imaginário isso dificulta.”

Algo que é bastante relevante para o músico é a variação de sonoridades. Em todos os experimentos os percussionistas procuraram variações que segundo eles eram importante para execução de algum ritmo, como por exemplo: a caixa no frevo, que necessita de acentuações o tempo inteiro e sem elas, o ritmo fica sem sentido ou a

³ Disponível em: <http://www.4shared.com/rar/0WDAkDV4/Experimento_Percusso_interativ.html?> - Acesso em: 06 Ago. 2013.

zabumba no baião, que precisa da variação timbrística da pele superior para fazer o swing.

A percepção que o músico tem ao tocar o seu instrumento é a mesma que ele usa para tocar em um instrumento digital, por tanto ele procura usar os mesmo gestos para obter as variações sonoras de timbre, altura e intensidade.

Dentre os problemas citados pelos usuários, os mais relevantes foram: a falta de variação de intensidade das notas, dada a imprecisão encontrada em alguns movimentos; variação de timbres no mesmo instrumento e atrasos em movimentos rápidos.

7.4.3. Reconstrução do Mundo do Usuário

Abaixo, encontra-se um resumo da entrevista semi-estruturada, onde alguns usuários descrevem suas experiências com as interfaces em suas perspectivas, salientamos que uma versão completa desta seção está disponível. Essa fase é uma representação do mundo do usuário, resultado da etapa AD e pode nos ajudar a entender limitações na interação entre usuário e interface.

Usuário 1 - Foi um músico de perfil autodidata. Toda sua experiência musical se deu participando de grupos musicais que tocam na noite e percorrem pelo Brasil. Não teve experiência com instrumentos virtuais, mas tem simpatia por novidades, desde que traga contribuições ao seu trabalho. Apesar de achar que não há condições de tocar com os atrasos encontrados, gostou da experiência e viu nas interfaces pontos positivos referentes à performance. O usuário achou a novidade muito interessante, principalmente por causa do problema de logística (transporte) que sofre o percussionista e acredita que no mercado, as interfaces iriam ajudar bastante. Ele percebeu que havia diferença na captação entre gestos mais lentos e mais rápidos na maioria dos instrumentos, menos na zabumba. No início teve dificuldades, mas salientou a necessidade de conhecer primeiro o equipamento, pra assimilar os gestos e no dia a dia domina-lo.

Usuário 2 - Foi um músico de perfil erudito. Aprendeu em escola de música e estuda percussão erudita na universidade. Possui muita experiência na área de percussão e participa de orquestra. Ele foi o usuário que mais se sentiu indiferente ao usar as interfaces, pois não acredita que instrumentos virtuais possam ser comparados e ter as mesmas qualidades encontradas nos instrumentos convencionais. Apresentou bastantes queixas, mas achou interessante para o uso na iniciação musical, como instrumento didático. Comentou que quando acompanhou o frevo no agogô, tentou fazer semicolcheias em 140bpm, mas o *Wii remote* não conseguiu acompanhar a articulação. Segundo ele, no instrumento acústico, quando é necessário fazer ritmos mais rápidos, os gestos são mais curtos e na interface, gestos curtos não são bem captados, sendo necessário usar movimentos mais longos, gerando atrasos no ritmo. Em seu ponto de vista, a vantagem em usa-las é impressionar o público com a tecnologia, não tecnicamente, nem musicalmente, mas visualmente.

Usuário 3 - Foi um músico de perfil popular. Estudou em conservatório de música e estuda percussão erudita e popular. Trabalha profissionalmente como professor de música e participa de grupos como: *bigband*, orquestra e grupos instrumentais. Nunca teve experiência com instrumentos virtuais, mas possui conhecimento em tecnologia computacional e gosta de novidades. Teve dificuldade em aplicar algumas técnicas, porque de acordo com sua descrição, no instrumento acústico ele percebe o corpo e calcula o golpe, já no virtual é diferente. Percebeu atrasos e apontou como um dos motivos, a falta de visualização do corpo como *feedback* visual,

pois algumas vezes levou a mão para baixo e esperava o som do golpe na pele, mas vinha um pouco depois e outras vezes o som vinha quando estava voltando a mão. O usuário falou que as interfaces apresentam dificuldades de precisão quando são usadas subdivisões, por isso que funcionam melhor em andamentos lentos. O instrumento que ele mais sentiu falta de variação de dinâmica foi a caixa, por ser um instrumento que usa muito esse tipo de técnica, para fazer rudimentos e nuances. Usaria em apresentações por ser uma novidade e pelo aspecto futurista, mas o usuário ainda não se sentiu confiante diante das interfaces. Diz-se defensor dos instrumentos acústicos, mas não é contra o digital, desde que seja interessante e vá somar ao seu set.

Usuário 4 - Foi um músico de perfil popular. Estudou no conservatório e atualmente estuda Licenciatura em música na universidade. Participa de grupos populares e apesar de ter conhecimento informatizado e simpatia por tecnologia, não possui experiência em instrumentos virtuais. O usuário reclamou os atrasos e a falta de controle da dinâmica, mas se sentiu bem com a sensação de liberdade e a possibilidade da troca de timbres. Gostou muito da ideia de colocar diversos timbres sem ter a presença física dos instrumentos e vê como bastante importante para bateristas e percussionistas. O usuário sentiu interesse em possuir as interfaces e usaria em apresentações em público, pois achou que a sensação de liberdade que a interface traz ao músico é fantástica.

7.4.4. Examinando o Contexto

Conforme o contexto visualizado, todos os usuários sentiram bastante dificuldade com relação à **latência**, além da **sensibilidade** funcionar melhor com gestos largos, era perceptível a dificuldade relacionada a **controle de fluxo**, que acontece, quando certa quantidade de comandos é executada em série e o fluxo de dados faz atrasar o programa. Para percussionistas essa situação é intolerável, pois segundo eles, a **métrica rítmica** fica comprometida, não havendo condições de compensar o atraso, e tanto a bateria quanto a percussão são os principais responsáveis pelo andamento rítmico.

Nenhum dos usuários conseguiu sentir **variação de dinâmica** e a **variação timbrística** foi muito resumida, fazendo com que houvesse queixas sobre **acentuação** na *performance* rítmica. Percebemos que a **acentuação é responsável pelos rudimentos e nuances** nos instrumentos solistas do naipe de percussão, como a Caixa, o Agogô e a Zabumba. Sem a acentuação, a música não tem o balanço (*swing*⁴).

Mesmo com toda a atratividade existente nas interfaces gestuais, as limitações encontradas na interação, trazem para os músicos **insegurança**, criando uma espécie de **preconceito a dispositivos digitais**. Se no instrumento convencional, onde normalmente eles manipulam bem, já encontram recursos suficientes para a extração do som desejado, porque trocar por outro artefato que por si só já é problemático?

7.5. Análise de Dados

Na interação com a música é necessário que aquilo que se toca em um instrumento tenha uma resposta imediata. Dillon (2003, apud CALEGARIO, 2010) afirma que:

(...) a resposta imediata do sistema ao gesto de entrada aumenta bastante o significado que a música tem para o usuário que está

⁴ Swing – Característica encontrada na música que faz com que ela tenha uma dinâmica envolvente. Geralmente induz ao desejo da dança.

interagindo com o sistema, em outras palavras, com a resposta em “tempo real” existe uma maior atração natural e envolvimento com a música por parte do usuário.

Como proposto anteriormente, um dos resultados da análise do experimento foi realizada a partir de três critérios de avaliação:

7.5.1. Variação Timbrística

Percebemos dificuldades nas captações de nuances das interfaces *Kinect* e *Wii remote* interferindo nas variações timbrísticas vinculadas aos gestos usados no instrumento virtual. Os usuários testaram, mas acharam limitados porque o instrumento convencional demonstrava muito mais possibilidades timbrísticas. Isso acontece também porque o mapeamento de gestos, que é necessário ser feito, vai de gestos musicais usados no instrumento, até as suas localizações espaciais. Quer dizer, além de a interface nos dar condições de tratarmos como uma variável, o gesto da batida de uma baqueta na pele da caixa, deve também nos dar condições de mapearmos a posição em que o gesto acontece, se no centro da pele da caixa ou na borda ou até mesmo no aro da caixa, pois no instrumento convencional essa variação é comum de ser utilizada na prática do instrumento.

Esse tipo de problema que encontramos, diz respeito às limitações de algoritmo e de hardware. Primeiro porque os algoritmos criados não tratam esse problema e segundo porque muitos dos gestos utilizados pelos músicos nos instrumentos fazem movimentos curtos que a própria interface não consegue captar.

7.5.2 Variação de Dinâmica

Ela se refere à variação de intensidade de volumes dentro da música ou de ostinatos⁵ rítmicos, de algum instrumento, arranjo ou fraseado. Também é responsável pelas notas acentuadas e pela não linearidade em uma música. De acordo com o experimento, nenhuma das duas interfaces ofereceu variação de dinâmica. Sendo assim, os algoritmos utilizados não resolveram problemas de mapeamento de gestos fracos e fortes.

Para ter controle de dinâmica a interface tem que possuir uma boa sensibilidade, e um mapeamento de gestos que contenha algoritmos que tratem desse critério. Uma das razões pelo qual gostamos mais de músicas interpretadas por seres humanos do que por programas de computador, está nessas variações de dinâmica, que são naturais na interpretação humana.

7.5.3 Controle temporal

O tempo foi um dos requisitos que mais deixou a desejar. Sua importância para este tipo de instrumento é fundamental, porque a percussão é um dos responsáveis pelo andamento da música. Usando o *Kinect* a latência foi bem maior do que com o *Wii remote*, pois os sensores usados por ele possuem dificuldade em captar movimentos rápidos. Apesar do instrumento de percussão ser um dos instrumentos que possuem gestos mais largos, existem detalhes como, por exemplo, rudimentos que são feitos com

⁵ Ostinato – é um motivo ou frase musical que é persistentemente repetido num mesmo padrão rítmico.
<<http://pt.wikipedia.org/wiki/Ostinato>>

movimentos curtos e são muito importantes para a expressão musical. Quando foi usado o ganzá também ocorriam sons com latência, porém o tipo de latência que apareceu era um problema de *jitter*⁶, conhecido também como problema de fluxo de dados, acontecia quando o usuário utilizava os gestos do ganzá de maneira mais rápida, fazendo com que as informações ocorressem com maior intensidade.

8. Conclusão e Trabalhos Futuros

Esta pesquisa nos deu condições de visualizar pontos relevantes a respeito de o porquê dos músicos usarem tão pouco as novas tecnologias digitais para a execução musical. Além dos problemas técnicos detectados, ainda existem outros que não foram citados por não fazer parte do nosso escopo de pesquisa, como: falta de informação, preconceito com tecnologia digital, falta de recursos financeiros que possibilitem a aquisição de equipamentos, problemas relacionados à cultura e outros. Apesar das interfaces se proporem a uma determinada interação, muitos dos gestos utilizados pelos instrumentos não são captados com facilidade em tempo real. A princípio podemos dizer que há dois pontos-chaves nessa questão. O primeiro é que existem limitações nas interfaces, que comprometem uma série de atividades na interação musical; e o segundo é que os desenvolvedores precisam criar melhores algoritmos.

Um dos problemas que os algoritmos podem tratar é a questão de variação timbrística. Podemos ter excelentes amostras de sons, e inúmeros tipos de instrumentos, mas se o mapeamento não for feito, levando em consideração a localização das sonoridades, nunca vamos ter uma variação timbrística real. Não foi possível trabalhar variação de dinâmicas e a sensibilidade das interfaces ainda não está refinada. A dinâmica e a sensibilidade estão relacionadas, porém a dinâmica deve ser tratada com algoritmos que percebam diferenças de velocidade, pois um movimento mais rápido deve proporcionar uma dinâmica mais forte e vice-versa, em quanto a questão da sensibilidade deve ser tratada no hardware, porque é ele quem capta os gestos, sejam eles longos ou curtos. No momento em que as interfaces forem mais sensíveis a pequenos gestos e os algoritmos melhorados, possivelmente a dinâmica irá funcionar com mais precisão.

Para o músico, independente dos recursos que o instrumento venha oferecer ou forma física, gesto novos ou antigos, visual, *expertise* necessário, maior liberdade, novas técnicas, múltiplos timbres, etc... a coisa mais importante a ser oferecida é uma interação que promova a sensação de total controle da interface através de gestos naturais.

Nessa área é interessante pesquisarmos:

- *Feedback* tátil na ergonomia
- Idiomatismo dos instrumentos como referência para mapeamento de gestos
- Interface gestual como ferramenta didática
- Mudança de técnicas instrumentais no uso de interfaces gestuais (novas formas de controle)

O futuro da nossa pesquisa é refazer o experimento com novos questionários, agora incorporando novos gestos e comparando com outra interface. Por exemplo: O *PlayStation Move*, *Leap Motion*.

⁶ Jitter é uma variação estatística do atraso na entrega de dados. <<http://pt.wikipedia.org/wiki/Jitter>> Acesso em: 06 Ago. 2013.

9. Referências

ARAÚJO, C. R. *Uma Proposta de Ferramenta de Apoio à Educação Musical Via Web Usando Java e XML*. 2002. 118 p. Dissertação (Mestrado) – Departamento de Comunicações, Faculdade de Engenharia Elétrica e Computação, Campinas, 2002.

BARBOSA, J.; CALEGARIO, F.; MAGALHÃES, F.; CABRAL, G.; TEICHRIEB, V.; RAMALHO, G. Towards an evaluation methodology for digital musical instruments considering performer's view: a case study. In PROCEEDINGS OF THE 13TH BRAZILIAN SYMPOSIUM ON COMPUTER MUSIC (SBCM'11), Vitória, ES. Brazil, 2011.

BOULEZ, Pierre (1977) "Technology and the Composer", *Language of Electroacoustic Music*. Ed. Simon Emmerson, London: Macmillan Press, 1986, pg. 5-14.

CANDÉ, Roland de. *História universal da música* v.1. 2 ed. São Paulo: Martins Fontes, 2001. ISBN 8533615000

CALEGARIO, F. Estudo de geração de música baseada em gestos do usuário. Trabalho de Graduação. Centro de Informática – UFPE, 2010.

CORREA, A.G.D., NETO, O.B., e LOPES, R.D. MusandScene: uma Interface Gestual de Expressão e Criação Musical no Auxílio ao Aprendizado de Música em Educação Especial. In: XXV Congresso da Sociedade Brasileira de Computação, 25, 2005, Rio Grande do Sul. *Anais...Canelas e São Leopoldo*, 2005.

JORDÀ, S.. Digital Lutherie. Tese de Doutorado. Universitat Pompeu Fabra, 2005.

IAZZETTA, F. A Música, o Corpo e as Máquinas. *Opus: Revista da Associação Nacional de Pesquisa e Pós-Graduação em Música*, IV(4), 27-44, 1997.

MIRANDA, Eduardo Reck; WANDERLEY, Marcelo. *New Digital Instruments: control and interaction beyond keyboard*. Middletown: A-R Editions, 2006.

WANDERLEY, M. M. (2006). Instrumentos Musicais Digitais: gestos, sensores e interfaces. In: Beatriz Ilari (ed.) *Em Busca da Mente Musical*. Curitiba: Editora da Universidade Federal do Paraná. <http://idmil.org/_media/wiki/instrumentos_digitais_final.pdf>

WANDERLEY, M. M. Projeto e utilização de instrumentos musicais digitais. In: KELLER, D. e BUDASZ, R. *Criação musical e tecnologias: teoria e prática interdisciplinar*. Goiânia: ANPPOM, 2010. P. 70-96.

Organização Automática de Coleções de Partituras Musicais Utilizando Agrupamento Hierárquico

Lucas M. de Oliveira¹, André Y. Kashiwabara¹, Carlos N. Silla Jr.¹

¹Laboratório de Computação e Tecnologia Musical
Universidade Tecnológica Federal do Paraná
Av. Alberto Carazzai, 1.640 – 86300-000
Cornélio Procópio, PR, Brazil

Abstract. *This work proposes a system to automatically organize music sheets in MusicXML format using a music sequence comparison algorithm and the hierarchical clustering technique. Our experimental results show that this approach is often able to group together music scores from the same music genre.*

Resumo. *Neste trabalho foi proposto um sistema para organizar automaticamente partituras musicais em formato MusicXML utilizando um algoritmo de comparação de sequências musicais e a técnica de agrupamento hierárquico. Os resultados experimentais obtidos mostram que o sistema é capaz de agrupar partituras de um mesmo gênero musical.*

1. Introdução

De forma geral, a música pode ser entendida como a arte de organizar e combinar sons no espaço e no tempo [Priolli, 2007]. A cadência de acordes é um exemplo de organização espacial, enquanto que as notas de uma melodia configuram uma organização temporal [Miranda, 2002]. Uma vez que a música possui papéis importantes dentro dos contextos sociais e psicológicos, formas de representação eficazes devem permitir buscas por conteúdo que contemplem estes aspectos. Tais formas são tipicamente baseadas em quatro tipos de informações: gênero, emoção, estilo e similaridades [Huron, 2000].

Estas informações estabelecem as principais aplicações em pesquisas que envolvem recuperação da informação musical. Gêneros e estilos musicais são, portanto, importantes descritores, uma vez que sumarizam características (ou padrões) comuns entre peças. Estes descritores são usados há anos para organizar coleções e refletem interação entre culturas [Holt, 2007]. Apesar do amplo uso, gêneros musicais não formam um conceito claro e definido, e suas fronteiras são controversas. A distinção entre gênero e estilo musical é similarmente controversa e redundante. Enquanto que gênero é geralmente considerado um conceito mais amplo e subjetivo em termos de conteúdo musical [Fabbri, 1999], não há um acordo geral sobre o que um estilo deve abranger ou se é aceitável o uso indiscriminado desses dois conceitos [Moore, 2001]. Muitas peças musicais são semelhantes, algumas mais semelhantes que outras. Além de servir como uma fonte infinita de discussão, tais semelhanças também são cruciais para o desenvolvimento de sistemas de recuperação da informação musical (do inglês, *Music Information Retrieval* - MIR) eficientes [Cilibrasi et al., 2004].

O crescimento constante dos dados musicais na Internet tem encorajado diversos pesquisadores a desenvolver ferramentas adequadas para a análise e a classificação destes dados. O objetivo principal de tais ferramentas é extrair a informação de forma compacta e representativa ao conteúdo dos bancos de dados. [Corrêa, 2012]. Este projeto de pesquisa consiste em desenvolver um algoritmo capaz de organizar automaticamente as partituras

monofônicas, sabendo que a classificação não supervisionada ou agrupamento de dados (do inglês, clustering), é relativamente pouco explorada para tarefas de classificação de gêneros musicais [Scaringella et al., 2006].

2. Revisão de Conceitos musicais

2.1. Entendendo o som

O som possui 4 propriedades:

- **Duração:** É o tempo que o som perdura ao ser executado.
- **Intensidade:** É a força do som, ou seja, força maior ou menor do som, a qual aumenta com a amplitude das vibrações.
- **Altura:** São as notas musicais (Dó, Ré, Mi, Fá, Sol, Lá, Si, Dó) em várias oitavas desde o grave até o agudo. O conjunto de notas existentes entre uma nota e a sua primeira repetição é chamado de oitava.
- **Timbre:** É através dessa propriedade que podemos distinguir a diferença entre sons da mesma nota (altura), duração e intensidade. Ex: a voz da criança, a voz da mulher e a voz do homem. Se pedirmos que essas pessoas cantem a mesma música separadamente, facilmente poderemos identificar quem está cantando, se uma criança, uma mulher ou um homem. Essa distinção entre as vozes é possível graças ao timbre.

Todo som produzido possui essas quatro propriedades, exceto alguns instrumentos de percussão, como o tambor, o bumbo, dentre outros.

2.2. Representação musical

Definição: Conjunto de sinais gráficos que permitem fixar em notação escrita, uma ideia musical com todos os seus detalhes: notas, pausas, intensidade, duração e timbre.

- **Nota:** Monossílabo que designa um som regular.
Sinal gráfico que representa a altura e duração dos sons musicais.[Med, 1996]
- **Pauta ou Pentagrama:** Utilizado para representar a notação musical.
A pauta ou pentagrama é composta por 5 linhas horizontais, paralelas e equidistantes entre si, formando 4 espaços ou entrelinhas, contadas de baixo para cima.
- **Clave:** é um sinal, colocado na extremidade esquerda da pauta, recebem essa denominação em virtude da posição onde são escritas representarem as notas sol, fá e dó, respectivamente, e com isto, fixa também o nome das outras notas, pela relação com a clave, as posições das claves e as notas estão exemplificadas na Figura 1. Há três tipos de clave: clave de sol, clave de fá e clave de dó.

2.3. MusicXML

MusicXML [Good, 2006] é um formato de arquivo de notação musical aberto, baseado em XML, utilizado para integrar partituras digitais entre aplicações. Foi inventado por Michael Good na empresa Recordare em 2000 e vem sendo desenvolvido com a colaboração de uma comunidade de centenas de músicos e desenvolvedores de software durante 13 anos. Disponível sob uma licença aberta, isenta de royalties, sendo vantajosa tanto para softwares open-source e softwares proprietários. Suportada por mais de 160 aplicações em todo o mundo.

(a) Clave de Sol

(b) Clave de Fá

(c) Clave de Dó

Figura 1: Exemplo das posição das claves e as notas representadas

3. Visão Geral

Como pode ser visto na Figura 2, tem-se como entrada o arquivo em formato musicXML, a partir deste, é feita a criação de um vetor de tóplas. O vetor de tóplas é utilizado como entrada do algoritmo de alinhamento proposto por [Mongeau and Sankoff, 1990]. Tem-se como resultado uma matriz de distância e a partir desta é gerado o dendrograma através do algoritmo de agrupamento hierárquico.



Figura 2: Visão geral do processo

Tabela 1: Mapeamento das notas musicais

C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	1	2	3	4	5	6	7	8	9	10	11

3.1. Parser de musicXML

Para ser possível a comparação de sequências musicais a partir de partituras, foi desenvolvido um parser de musicXML, onde o software lê a partitura e gera um vetor de túplas, onde cada túpla tem a informação sobre a nota, sua oitava e a sua duração.

Formalmente, este procedimento é definido como dado uma música M vai ser composta de um vetor de túplas com 3 elementos $\langle A, B, C \rangle$ onde:

- A = Letra da nota musical ou @ caso seja uma pausa.
- B = Oitava da nota musical ou VAZIO \emptyset caso seja uma pausa.
- C = Duração de cada nota ou pausa. O elemento de duração é um inteiro que representa a duração de uma nota em termos de divisões por semínima.

Este procedimento esta exemplificado na Figura 3. Onde o primeiro item apresentado é parte de uma música em musicXML e depois é apresentado o resultado da geração do vetor de túplas, considerando apenas o trecho de código apresentado.

3.2. Transposição de Tons

Após a criação do vetor de túplas, é necessário transpor todas as músicas para o mesmo tom ser efetuada a comparação das mesmas. Esta etapa é importante, pois pode ocorrer de uma música ser similar a outra mas em uma tonalidade diferente, tendo a mesma melodia, porém com notas diferentes, e ao ser comparada, pode gerar resultados inconsistentes. Assim para efetuar a transposição em larga escala, foi criada a equação para transpor todas as partituras para o tom C (Dó), podendo assim gerar resultados consistentes. Para automatizar a transposição, cada nota foi associada a um conceito numérico, como apresentado na Tabela 1.

Então utilizando como exemplo a música infantil "Brilha, brilha estrelinha" no tom A (Lá), iremos transpo-la para o tom C (Dó), ou seja, iremos "descer" cinco tons e um semitom. Em notação escrita em valores tem-se a música no tom A (Lá) como: 9, 9, 4, 4, 6, 5, 4, 2, 2, 1, 0, 11, 11, 9. O tom de uma música, geralmente é a primeira nota tocada, portanto neste exemplo o tom é a nota A, ou escrita em valores 9.

A equação utilizada para a transposição é $NOTATRANSPOSTA = NOTA - TOM$. A nota A (Lá) escrita em valores equivale a 9, portanto para transpo-la utilizando a equação será $9 - 9 = 0$, onde 9 é a nota a ser transposta, ou seja, a nota A (Lá), o outro 9 é o tom da música, e o resultado 0 equivale a nota C (Dó).

Outro problema encontrado na transposição de tons foi quando uma nota está em uma oitava diferente da outra nota, não conseguindo transpor de forma correta algumas notas. Por exemplo a nota E (Mi), no tom A (Lá), transposta para o tom C (Dó) pela equação, seria a nota F (Fá), mas na realidade é a nota G (Sol). Para solucionar este problema, foi criada a equação $NOTAREAL = (OITAVA - 1) * 12 + NOTA$. Com isto encontra-se o real valor da nota a ser transposta, assim podendo transpor qualquer nota em qualquer oitava para o tom C (Dó), então transpondo a nota E (Mi) para o tom C (Dó), sabe-se que a tom está na quarta oitava, e a nota E (Mi) está uma oitava acima, ou seja, na quinta oitava. Então temos para a nota E (Mi) $(5 - 1) * 12 + 4 = 52$, e para a nota A (Lá) $(4 - 1) * 12 + 9 = 45$, então utilizando a equação da transposição temos $|52 - 45| = 7$, onde 7 equivale a nota G (Sol).

```

<note>
<pitch>
<step>D</step>
<octave>4</octave>
</pitch>
<notations>
<technical>
<fret>0</fret>
<string>4</string>
</technical>
</notations>
<voice>1</voice>
<duration>960</duration>
<type>quarter</type>
</note>
<note>
<pitch>
<step>C</step>
<octave>4</octave>
</pitch>
<notations>
<technical>
<fret>3</fret>
<string>5</string>
</technical>
</notations>
<voice>1</voice>
<duration>1920</duration>
<type>half</type>
</note>
</measure>
<measure number="5">
<note>
<rest/>
<voice>1</voice>
<duration>1920</duration>
<type>half</type>
</note>

```



...;D,4,960;C,4,1920;@,Ø,1920;...

Figura 3: Exemplo de como é criado vetor de túplas

3.3. O Algoritmo de Comparação

Na comparação de sequências musicais, existem alguns critérios a serem utilizados, estes são a dissimilaridade, as transformações típicas e os pesos.

A dissimilaridade entre duas sequências musicais, é dada a partir da menor quantidade de transformações típicas necessárias para uma sequência A transformar-se em uma sequência B. Foram utilizados três tipos de transformações: a Deleção, a Inserção e a Substituição. Para melhor exemplificar as transformações típicas e o conceito de dissimilaridade, segue o exemplo: Dado duas sequências A e B (Figura 5) para transformar A



(a) Música em tom A (Lá)



(b) Música em tom C (Dó)

Figura 4: Trecho da Música “Brilha, brilha estrelinha” em dois tons diferentes

em B será necessário efetuar a deleção da nota D (Ré) (Figura 6), a substituição da nota B (Si) pela nota C (Dó) (Figuras 7.a e 7b.) e a inserção da nota C# (Dó sustenido) (Figuras 7.c e 7.d).



Figura 5: Sequências A e B de notas musicais e representação das mesmas em partitura

Para transformar A em B, foram necessários 3 transformações, tendo uma dissimilaridade de valor 3.

O conceito de dissimilaridade entre duas notas, pode ser generalizado associando um peso a cada tipo de transformação, podendo assim, pré-definir valores relacionados para cada transformação (inserção, deleção, substituição) e os elementos envolvidos na transformação.

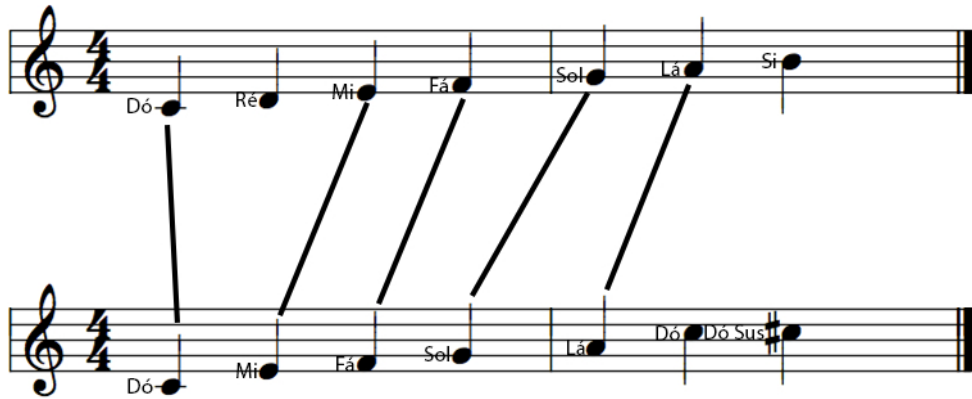
3.3.1. Pesos

A dissimilaridade vem da soma dos menores pesos de todas as possibilidades de transformação de A para B. Sendo $m_{i,j}$ a dissimilaridade entre as notas de A e B, pode ser visto na equação 1 os pesos associados a deleção, inserção e substituição [Mongeau and Sankoff, 1990].

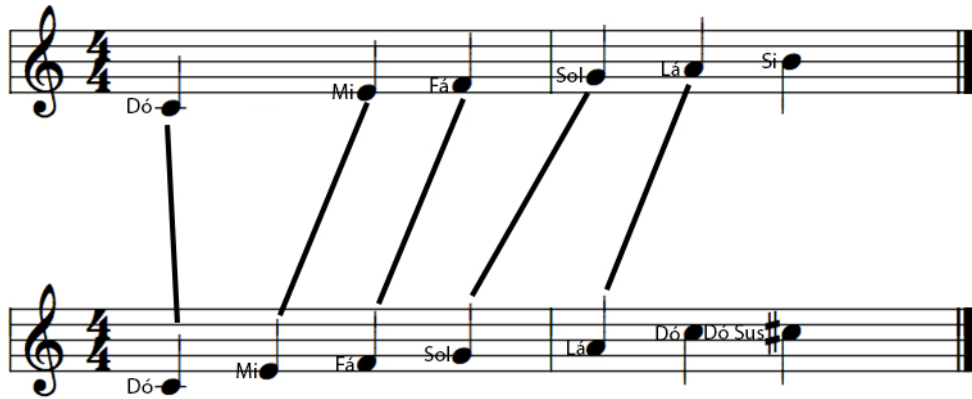
$$m_{i,j} = \min \begin{cases} m_{i-1,j} + w(a_i, \emptyset) & \text{(Deleção) ;} \\ m_{i,j-1} + w(\emptyset, b_j) & \text{(Inserção) ;} \\ m_{i-1,j-1} + w(a_i, b_j) & \text{(Substituição) ;} \end{cases} \quad (1)$$

O peso é definido pela soma de outros dois pesos, sua equação pode ser vista na equação 2.

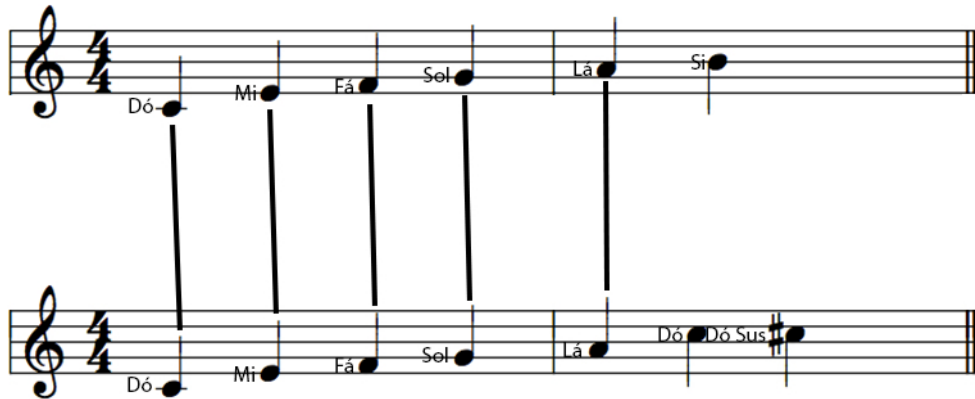
$$w(a_i, b_j) = w_{interval}(a_i, b_j) + k * w_{length}(a_i, b_j) \quad (2) \quad 171$$



(a) Exemplo da necessidade de deleção para o alinhamento.

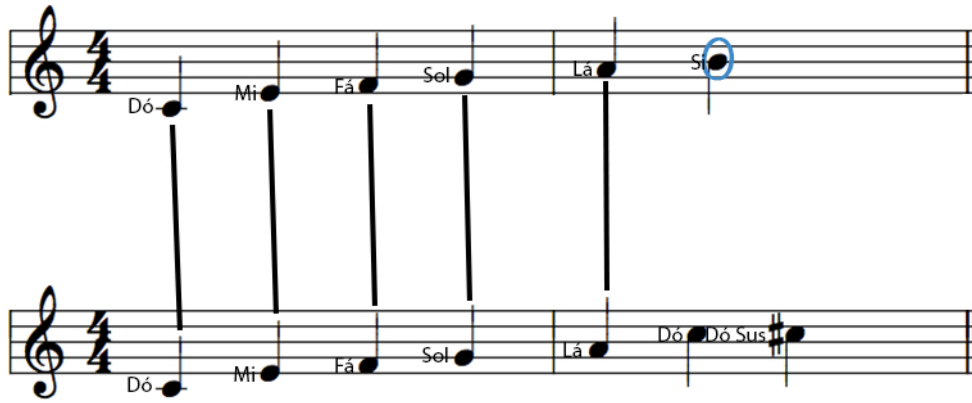


(b) Segunda nota do 1o compasso da 1a sequencia (Ré) removido.

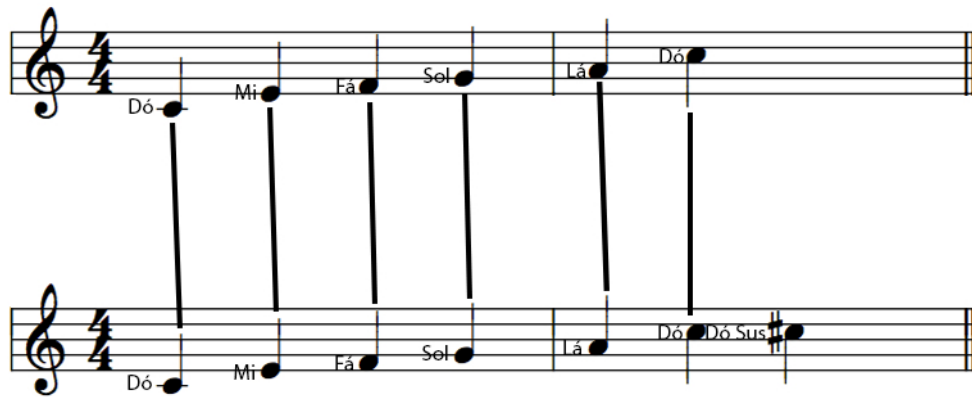


(c) Primeiro compasso de ambas as sequencias alinhados.

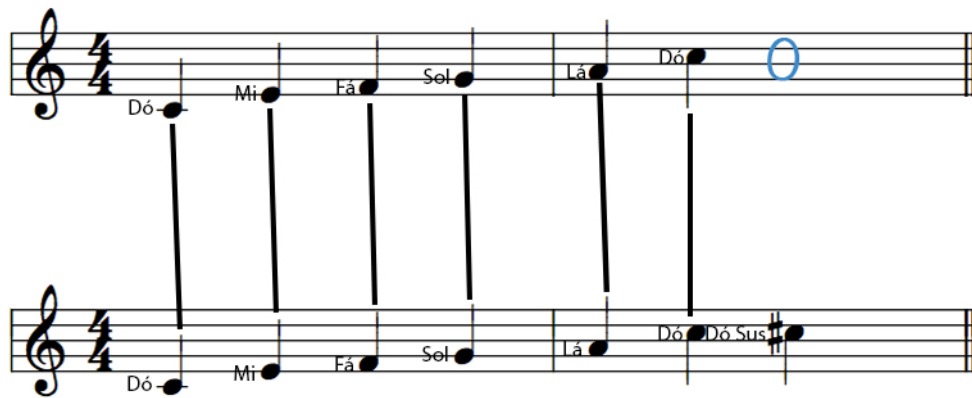
Figura 6: Exemplo de Deleção



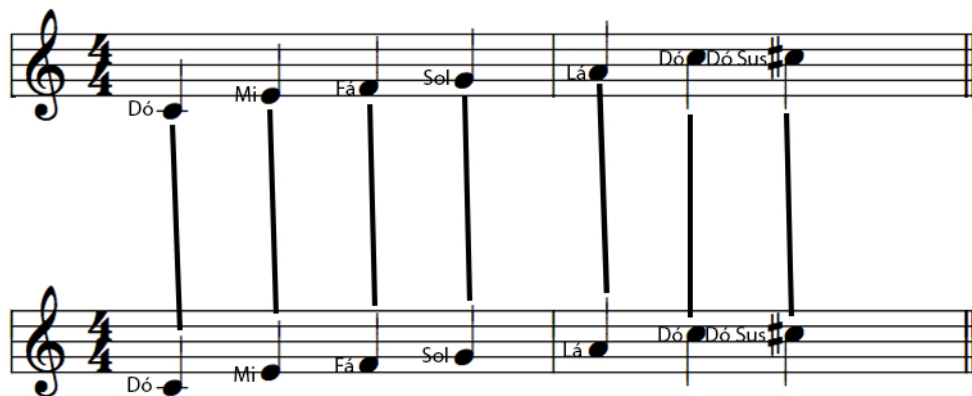
(a) Exemplo de Substituição: A 2a Nota do 2o Compasso (Si) vai ser substituída por um Dó pelo algoritmo de alinhamento.



(b) Resultado da substituição.



(c) Exemplo de Inserção: Para que as duas sequencias fiquem iguais é necessário incluir uma nota (Dó Sustenido) no 2o compasso.



(d) Resultado da inclusão.

Figura 7: Exemplo de Substituição e Inserção

Onde a_i e b_j são as notas das músicas A e B e k representa a contribuição relativa da diferença dos tamanhos pela diferença de notas e foi considerado com valor 3.

A função $w_{length}(a_i, b_j)$ é definida pela equação 3.

$$w_{length}(a_i, b_j) = ||a_i| - |b_j|| \quad (3)$$

A função $w_{interval}(a_i, b_j)$ é definida pelo grau de diferença entre a nota a_i e a nota b_j , onde cada grau é definido por:

$$\begin{aligned} deg(0) &= 0; \\ deg(4) &= 0.1; \\ deg(2) &= 0.2; \\ deg(5) &= 0.35; \\ deg(3) &= 0.5; \\ deg(6) &= 0.8; \\ deg(1) &= 0.9; \\ deg(> 6) &= 9999999999; \end{aligned}$$

4. Resultados e Discussões

Foram utilizadas vinte e cinco partituras no formato musicXML em cinco gêneros diferentes, apresentadas na Tabela 2.

Após a criação do vetor de túplas são feitas as transposições de todas as partituras do banco de dados, foram comparadas todas as músicas, par a par, utilizando o algoritmo proposto em [Mongeau and Sankoff, 1990]. Para efetuar a comparação, é necessário inicializar a matriz $m_{i,j}$, para $1 \leq i \leq m$, $1 \leq j \leq n$, definindo os pesos associados a deleção e inserção entre duas músicas a serem comparadas.

$$\begin{aligned} m_{i,0} &= m_{i-1,0} + w(a_i, \emptyset) \quad , i \geq 1; \\ m_{0,j} &= m_{0,j-1} + w(\emptyset, b_j) \quad , j \geq 1. \end{aligned}$$

Então a partir da aplicação da equação 1, é gerado um escore da comparação entre duas músicas. Após efetuar a comparação par a par entre todas as músicas, foi construído uma matriz de distância, e utilizando o método de aglomeração hierárquica, construiu-se um dendrograma. Uma matriz de distâncias é onde cada linha explicita a hierarquia de proximidade de cada observação em relação às restantes, para cada observação, a proximidade com as outras será maior, quanto maior for a similaridade que lhes corresponde.

O algoritmo de aglomeração hierárquica progressivamente agrupa as N amostras em C grupos, de acordo com um parâmetro definido, geralmente associado a uma medida de distância ou similaridade entre os vetores de atributos das amostras. No primeiro passo ($k = 1$), as N amostras são atribuídas como N grupos. No passo seguinte ($k = 2$), ocorre uma partição de $N - 1$ grupos, em seguida uma partição de $N - 2$ grupos e assim por diante. No último passo ($k = N$), todas as amostras fazem parte de um único grupo. Portanto, esta sequência progressivamente agrupa amostras que são mais similares entre si antes de considerar amostras menos similares. É possível dizer que, no k -ésimo passo, $C = N - k + 1$ (C é o número de grupos) [Corrêa, 2012].

O dendrograma é a representação gráfica do processo de constituição dos clusters, onde permite identificar os clusters agrupados ao longo de todo o processo e observar o incremento nos valores da distância entre os clusters.

Tabela 2: Código, Gênero, Artista e Título das partituras selecionadas

Código	Gênero	Artista	Título
BOL1755	Bolero	Luis Miguel	Inolvidable
BOL1758	Bolero	Francisco Xavier	Recuerdos de Ipacará
BOL1764.2	Bolero	Linda Ronstadt	Perfídia
BOL1770.2	Bolero	Francisco Xavier	El reloj
BOL1770	Bolero	Francisco Xavier	El reloj
BOS0001	Bossa Nova	Chico Buarque	Bye, Bye brasil
BOS0002	Bossa Nova	Tom Jobim	Águas de março
BOS0003	Bossa Nova	Noel Rosa	A.B. surdo
BOS0004	Bossa Nova	Chico Buarque	A Rita
BOS0005	Bossa Nova	Tom Jobim	Chega de saudade
PAG2961	Pagode	Arlindo Cruz	Alem do meu querer
PAG2964	Pagode	Arlindo Cruz	Brasil Moleque
PAG2970	Pagode	Arlindo Cruz	Será que é amor
PAG2972	Pagode	Arlindo Cruz	Um sim para que te gosta
PAG2973	Pagode	Arlindo Cruz	Malandro sou eu
SER2614	Sertanejo	Rio Negro & Solimões	Tô Mal
SER2635	Sertanejo	Rio Negro & Solimões	Bate o pé
SER2644	Sertanejo	Rio Negro & Solimões	Saudade pulou no peito
SER2696	Sertanejo	Bruno & Marrone	Vai dar namoro
SER2822	Sertanejo	Leandro & Leonardo	Entre tapas e beijos
TAN6.2	Tango	Horacio Salgán	Los Mareados
TAN6.3	Tango	Horacio Salgán	Los Mareados
TAN48.3	Tango	Carlos Cardel	Caminito
TAN56.5	Tango	Carlos Gardel	Adios Muchachos
TAN81.5	Tango	Astor Piazzolla	Libertango

Analisando a figura 8, pode-se perceber que algumas músicas tal como as músicas do gênero sertanejo e pagode, foram agrupadas em uma mesma aglomeração. Percebe-se também que as variações da música TAN6 (TAN6.2 e TAN6.3) foram agrupadas unicamente em uma mesma aglomeração.

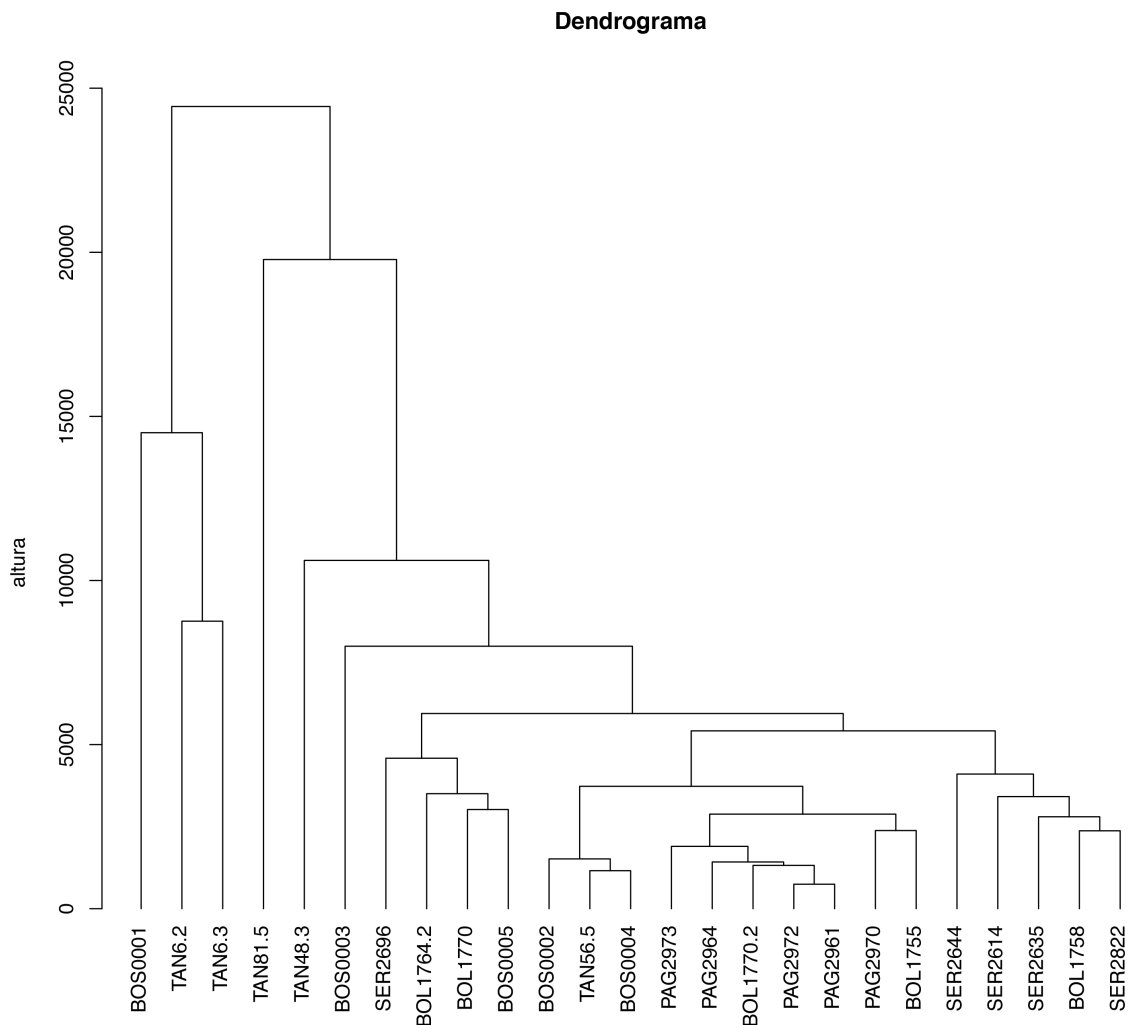


Figura 8: Dendrograma: Os prefixos BOS, TAN, SER, PAG, e BOL são, respectivamente, bossa nova, tango, sertanejo, pagode e bolero

5. Conclusões

Neste trabalho foi proposta uma abordagem para a organização automática de partituras utilizando um algoritmo de comparação de seqüências musicais e a técnica de agrupamento hierárquico. Foram realizados experimentos utilizando uma base com vinte e cinco partituras em formato musicXML de cinco gêneros musicais diferentes (bolero, bossa nova, pagode, sertanejo e tango). O método proposto se mostrou capaz de agrupar corretamente as partituras. Como trabalho futuro pretendemos fazer outros experimentos utilizando uma base de dados com um maior numero de partituras.

Referências

- Cilibrasi, R., Vitányi, P., and de Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal*, pages 49–67.
- Corrêa, D. C. (2012). *Inteligência artificial aplicada à análise de gêneros musicais*. PhD thesis, Universidade de São Paulo.

- Fabbri, F. (1999). Browsing music spaces: categories and the musical mind. *International Association For The Study of Popular Music*, pages 1–14.
- Good, M. (2006). Lessons from the adoption of musicxml as an interchange standard. In *Proceedings of the XML Conference*.
- Holt, F. (2007). *Genre in popular music*. University of Chicago Press.
- Huron, D. (2000). Perceptual and cognitive applications in music information retrieval. *International Society for Music Information Retrieval*, pages 1–6.
- Med, B. (1996). *Teoria da música*. Musimed.
- Miranda, E. R. (2002). *Composing music with computers*. Focal Press.
- Mongeau, M. and Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities*, 24(3):161–175.
- Moore, A. F. (2001). Categorical conventions in music discourse: style and genre. *Music & Letters*, 82:432–442.
- Priolli, M. L. d. M. (2007). *Princípios básicos da música para a juventude*, volume 1. Casa Oliveira de Músicas LTDA.
- Scaringella, N., Zoia, G., and Mlynek, D. (2006). Automatic genre classification of music content: a survey. *IEEE Signal Processing Magazine*, 23:133–141.

Classificação Automática de Gêneros Musicais Utilizando Cifras das Músicas

Ricardo Corassa¹, Carolina L. Dos Santos¹, Carlos N. Silla Jr.¹

¹Laboratório de Computação e Tecnologia Musical
Universidade Tecnológica Federal do Paraná
Av. Alberto Carazzai, 1.640 – 86300-000
Cornélio Procópio, PR, Brazil

Abstract. *For the task of automatic music genre classification the use of content-based features appears to have reached a glass-ceiling. For this reason, recent research in music genre classification has been working on developing new approaches for the task. In this paper we present five novel feature descriptors based on “cifras”(a document that contains the harmonic structure and progression of the songs). We have evaluated our approach on a subset of the Latin Music Database (LMD), which was manually created by the authors querying different websites for the songs presented in the LMD.*

Resumo. *Para a tarefa de classificação automática de gêneros musicais o uso de características extraídas diretamente do sinal de áudio das músicas parece ter encontrado um limite. Por esse motivo, pesquisas recentes tem focado em desenvolver novas abordagens para tarefa. Neste trabalho são apresentados cinco novos tipos de características baseados na informação presente nas cifras das músicas. Esses novos descritores foram avaliados em um subconjunto de músicas da Latin Music Database para as quais os autores deste trabalharam conseguiram obter as cifras de websites especializados.*

1. Introdução

A tarefa de classificação automática de gêneros musicais é um dos tópicos mais populares na comunidade de recuperação inteligente de informações musicais (Music Information Retrieval - MIR) [McKay and Fujinaga, 2006b]. A principal motivação desta tarefa é construir sistemas que possam organizar de forma automática coleções de músicas em formato digital. Além disso, existem estudos que mostram que o gênero musical é um dos termos frequentemente utilizados por usuários em sistemas de recuperação inteligente de informações musicais [Downie and Cunningham, 2002, Lee and Downie, 2004, Pampalk et al., 2002].

O trabalho mais importante da área foi o trabalho de [Tzanetakis and Cook, 2002] onde os autores propuseram a tarefa de classificação automática de gêneros musicais utilizando algoritmos de aprendizado de máquina. Neste trabalho os autores utilizaram 30 características que são extraídas diretamente do sinal de áudio.

Desde o trabalho de [Tzanetakis and Cook, 2002], diversos pesquisadores tem tentando resolver o problema utilizando: abordagens baseadas no conteúdo (isto é, usando apenas o sinal de áudio) [Tzanetakis and Cook, 2002]; abordagens baseadas na representação simbólica das músicas (isto é, utilizando informações sobre a melodia das músicas) [McKay and Fujinaga, 2006a]; abordagens utilizando outras representações como informações culturais [Whitman and Smaragdis, 2002], letras das músicas [Mayer et al., 2008], ou social tags [Chen et al., 2009].

Maior	Menor	Sustenido	Bemol	Quarta	Quinta	Sétima	Nona	Diminuta
C	Cm	C#	Cb	C4	C5	C7	C9	C°
D	Dm	D#	Db	D4	D5	D7	D9	D°
E	Em	E#	EB	E4	E5	E7	E9	E°
F	Fm	F#	Fb	F4	F5	F7	F9	F°
G	Gm	G#	Gb	G4	G5	G7	G9	G°
A	Am	A#	Ab	A4	A5	A7	A9	A°
B	Bm	B#	Bb	B4	B5	B7	B9	B°

Figura 1: Tipos de acordes utilizados neste trabalho.

A principal motivação para o uso de outras representações além do sinal de áudio foi o trabalho de [Lidy et al., 2007] que indicou que as abordagens baseadas apenas em conteúdo podem ter encontrado um limite. Por esse motivo neste trabalho investigamos o uso de descritores baseados na estrutura harmonica obtidas em cifras das músicas. Informações sobre a estrutura harmônica já foram utilizada nos trabalho de [Cheng et al., 2008] porém neste trabalho os autores utilizaram uma abordagem automática para estimar os acordes a partir do sinal de áudio das músicas para então fazer a classificação automática de gêneros musicais.

A principal contribuição deste trabalho é analisar o uso de cinco diferentes descritores obtidos a partir das cifras das músicas para fazer a classificação automática de gêneros musicais utilizando apenas a estrutura harmônica presente nas cifras das músicas. O restante deste trabalho está organizado da seguinte maneira: na seção 2 são apresentados alguns conceitos musicais; na seção 3 são apresentados os descritores obtidos a partir das cifras; na seção 4 são apresentados os experimentos realizados; e finalmente na seção 5 são apresentadas as conclusões deste trabalho.

2. Conceitos musicais

Segundo Souza [Souza, 20], iniciou-se na Grécia antiga a representação de notas musicais por letras do alfabeto grego e, com o passar do tempo, houve a procura para formular um modo de representar e divulgar objetos musicais. Foi no início da idade média que o monge italiano Guido d'Arezzo, regente de um coro da Catedral de Arezzo (Toscana), empregou um sistema silábico baseado em um hino para a representação das notas musicais, que é utilizado até hoje. A representação utilizada por d'Arezzo deu o nome às notas simbolizadas pelas sete primeiras letras do alfabeto (A = lá, B = si, C = dó, D = ré, E = mi, F = fá, G = sol).

A cifra é uma notação musical, utilizada pra representar as notas musicais por meio dos símbolos universais, conhecidos por qualquer músico, independente de seu instrumento. Em uma cifra usa-se a representação de acordes ou arpejos que são formados pela associação de 3 ou mais notas tocadas simultaneamente. Os acordes mais comuns são classificados como tríades, constituídos por 3 notas, organizadas para formarem intervalos de terças superpostas. Um intervalo é a distância entre duas notas, no caso do violão um intervalo é equivalente a um semi-tom e uma casa do braço do violão [Faria, 1999]. A Figura 1 apresenta os tipos de acordes utilizados neste trabalho e a Figura 2 apresenta o trecho de uma cifra.

A transposição de uma cifra consiste em aumentar ou diminuir a altura todas as notas presentes na cifra. A transposição de uma nota musical é a mudança da altura de uma nota. Cada nota tem sua escala com seus descentes. Uma escala de Tom/Semitom é uma escala a qual cumpre escala octatônica, a qual vai de oito em oito notas. C (dó), D (ré), E (mi), F

Estrela Primeira

Banda Beijo

Tom: D

Intro: Bm A G F#

Bm A
Reluz do meu ser a fonte mais pura do viver
Em F#
Será que é ilusão as batidas do meu coração
Bm A
Pairando no ar, toda sapiência de um ser

Figura 2: Exemplo de um trecho de uma das cifras utilizadas.

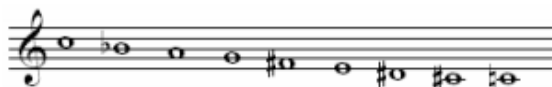


Figura 3: Exemplo de como as cifras foram processadas.

(fá), G (sol), A (lá), B (si) e C(dó), exemplificada na figura 3 [Kostka, 1999]. Cada escala tem um número limitado de possibilidades de transposições. Segundo [Zuben, 2005], as transposições são baseadas em um sistema cromático temperado, formando uma limitação de mudanças e formando grupos simétricos internos dentro de cada oitava, ou escala. Quando, no violão, passa-se de traste, as notas em cada corda se alteram. Por exemplo, na primeira corda E (mi) agudo, solta a corda tem a tonalidade E (mi), no primeiro traste, ao aperta-la, a corda emite o som da nota F (fá).

A Figura 4 exemplifica quantos tons e semitons tem em cada escala, como chamamos as notas que “combinam” com a nota principal.

Para realizar a transposição das cifras neste trabalho foi utilizada a codificação apresentada na Tabela 1, onde cada número corresponde a uma nota. Para fazer a transposição, analisa-se que o tom natural da música corresponde ao número '0', sendo assim, se subtrairmos a nota dela mesma obtemos o tom natural (C). Subtraindo o tom dos demais acordes, encontramos a quantos intervalos a nota está do seu tom, mostrando resultados positivos que estão no sentido crescente, logo para resultados negativos, estão no sentido decrescente. Desta forma percorremos uma lista circular, nos dois sentidos.

3. Descritores Obtidos a Partir das Cifras Processadas das Músicas

Nesta seção são apresentados novos descritores para classificação automática de gêneros musicais utilizando cifras.

3.1. Presença e Ausência de Acordes

O primeiro descritor proposto neste trabalho tenta caracterizar os diferentes gêneros musicais com base na presença e ausência dos acordes na cifra de cada música. Dessa forma, é criado um vetor de características com 63 atributos para cada música. Cada posição desse vetor corresponde a um acorde, e para cada cifra analisada os acordes encontrados são marcados como presentes (utilizando o número 1) ou ausentes (utilizando o número 0).

B	C	Db	D	Eb	E	F	F#	G	Ab	A	Bb	5 e 1/2
Bb	B	C	Db	D	Eb	E	F	F#	G	Ab	A	5
A	Bb	B	C	Db	D	Eb	E	F	F#	G	Ab	4 e 1/2
Ab	A	Bb	B	C	Db	D	Eb	E	F	F#	G	4
G	Ab	A	Bb	B	C	Db	D	Eb	E	F	F#	3 e 1/2
F#	G	Ab	A	Bb	B	C	Db	D	Eb	E	F	3
F	F#	G	Ab	A	Bb	B	C	Db	D	Eb	E	2 e 1/2
E	F	F#	G	Ab	A	Bb	B	C	Db	D	Eb	2
Eb	E	F	F#	G	Ab	A	Bb	B	C	Db	D	1 e 1/2
D	Eb	E	F	F#	G	Ab	A	Bb	B	C	Db	1
Db	D	Eb	E	F	F#	G	Ab	A	Bb	B	C	1/2
C	Db	D	Eb	E	F	F#	G	Ab	A	Bb	B	
B	C	Db	D	Eb	E	F	F#	G	Ab	A	Bb	1/2
Bb	B	C	Db	D	Eb	E	F	F#	G	Ab	A	1
A	Bb	B	C	Db	D	Eb	E	F	F#	G	Ab	1 e 1/2
Ab	A	Bb	B	C	Db	D	Eb	E	F	F#	G	2
G	Ab	A	Bb	B	C	Db	D	Eb	E	F	F#	2 e 1/2
F#	G	Ab	A	Bb	B	C	Db	D	Eb	E	F	3
F	F#	G	Ab	A	Bb	B	C	Db	D	Eb	E	3 e 1/2
E	F	F#	G	Ab	A	Bb	B	C	Db	D	Eb	4
Eb	E	F	F#	G	Ab	A	Bb	B	C	Db	D	4 e 1/2
D	Eb	E	F	F#	G	Ab	A	Bb	B	C	Db	5
Db	D	Eb	E	F	F#	G	Ab	A	Bb	B	C	5 e 1/2

Figura 4: Exemplo de como as cifras foram processadas.

Acorde	Valor
C	0
C#	1
D	2
D#	3
E	4
F	5
F#	6
G	7
G#	8
A	9
A#	10
B	11

Tabela 1: Acordes e seus correspondentes numéricos

Por exemplo, utilizando como base o trecho da cifra apresentada na Figure 11 será criado o descritor apresentado na Figura 5 .

Música	C	D	E	F	G	A	B	Cm	Dm	Em	Fm	Gm	Am	Bm	...
Estrela Primeira	0	1	0	0	0	1	0	0	0	1	0	0	0	1	...

Figura 5: Exemplo da presença / ausência dos acordes.

3.2. Tom

O Tom é quem determina qual será o grupo de acordes que será utilizado na música, logo influenciará diretamente nas demais características a serem estudadas. Durante o pré-processamento das cifras manteve-se esta informação, afim de analisar a sua presença. Para as cifras que não continham tom, gerou-se o tom baseado no acorde com o maior número de repetições na cifra. Por exemplo, utilizando como base o trecho da cifra apresentada na Figure 11 será criado o descritor apresentado na Figura 6.

Música	Tom
Estrela Primeira	D

Figura 6: Exemplo do uso do tom.

3.3. Quantidade de repetições dos acordes

A execução de uma música gera constantes repetições do mesmo grupo de acordes, muitas vezes repetidos em partes diferentes da música. Sendo assim, para computar esse descritor as cifras foram analisadas e o número de vezes que cada acorde aparece na cifra foi contabilizado. No caso dos acordes que não aparecem na cifra, estes tem o valor 0. Por exemplo, utilizando como base o trecho da cifra apresentada na Figure 11 será criado o descritor apresentado na Figura 7.

Música	C	D	E	F	G	A	B	Cm	Dm	Em	Fm	Gm	Am	Bm	...
Estrela Primeira	0	1	0	0	0	2	0	0	0	1	0	0	0	2	...

Figura 7: Exemplo da quantidade de repetições dos acordes.

3.4. Frequência dos Acordes

Sabendo que um acorde pode ser usado inúmeras vezes na mesma música, uma cifra muito longa pode apresentar muitas vezes o mesmo acorde, em comparação com uma cifra mais curta. Com o intuito de estabelecer uma relação, gerou-se a proporção com que o acorde se repete. Logo, podemos estabelecer a frequência pela equação 1. Um exemplo desse descritor, utilizando como base o trecho da cifra apresentada na Figura 11, é apresentado na Figura 8.

Música	C	D	E	F	G	A	B	Cm	Dm	Em	Fm	Gm	Am	Bm	...
Estrela Primeira	0	0.16	0	0	0	0.33	0	0	0	0.16	0	0	0	0.33	...

Figura 8: Exemplo da frequência dos acordes.

$$S_n = \frac{\sum_i^n Acorde}{Total_{acordes}}. \quad (1)$$

3.5. Transição de Acordes

A transição de um acorde é o avanço ou regressão em uma escala de semi-tons e também é conhecida como progressão harmônica. Para esse descritor a cifra é analisada e todas as transições de acordes são obtidas. Para as transições existentes na cifra é utilizado o valor 1, e 0 para as demais transições. Um exemplo desse descritor, utilizando como base o trecho da cifra apresentada na Figura 11, é apresentado na Figura 9.

Música	C → D	C → E	C → F	...	A → Em	...	Bm → A#	...
Estrela Primeira	0	0	0	...	1	...	1	...

Figura 9: Exemplo das transições dos acordes.

3.6. Junção de Todos os Descritores

Além do uso isolado de cada uma das representações propostas, neste trabalho também verificamos o uso de todas essas representações combinadas de forma simples, isto é, concatenando todos os vetores de características. Um exemplo desse procedimento, utilizando como base o trecho da cifra apresentada na Figura 11, é apresentado na Figura 10.

	Presença			Tom	Repetições				Frequência				Transições		
Música	C	D	...		C	D	E	...	C	D	E	...	C→D	C→E	...
Estrela Primeira	0	1	0	D	0	0	0	0	0	0	0	0	0	0	...

Figura 10: Exemplo da junção de todos os descritores.

Estrela Primeira

Banda Beijo

Tom: C

Am G
 Reluz do meu ser a fonte mais pura do viver
 Dm E
 Será que é ilusão as batidas do meu coração
 Am G
 Pairando no ar, toda sapiência de um ser

Figura 11: Cifra após o pré-processamento.

4. Detalhes Experimentais

4.1. Criação da Base de Dados

Para realizar os experimentos deste trabalho foram obtidas cifras para as músicas da Latin Music Database (LMD) [Silla Jr. et al., 2008]. O processo de obtenção das cifras foi manual e foi realizado pelos autores deste trabalho. Uma das dificuldades encontradas neste trabalho foi o de encontrar cifras para todas as músicas presentes na LMD. A Tabela 2 apresenta a quantidade de cifras encontradas por gênero musical.

Tabela 2: Número de cifras presentes na base criada em relação a base original por gênero musical. Onde: A = Axé; Ba = Bachata; Bo = Bolero; F = Forró; G = Gaúcha; M = Merengue; P = Pagode; Sa = Salsa; Se = Sertaneja; T = Tango

Base	A	Ba	Bo	F	G	M	P	Sa	Se	T	Total
Cifras	139	61	209	128	136	46	109	205	177	28	1.238
Original	304	308	303	315	306	307	301	303	310	404	3.161

4.2. Pré-Processamento dos Dados

Para a construção da base de cifras utilizadas nos experimentos, foram encontradas cifras em diferentes websites. É importante ressaltar que cada website possui o seu padrão específico na forma de apresentar uma cifra. Dessa forma, antes de realizar a extração dos descritores apresentados na Seção 3 todas as cifras passaram por uma etapa de pré-processamento.

Nessa etapa foram eliminadas todas as informações que não fossem referentes a estrutura harmônica das cifras. Além destas, também foram eliminadas informações como: “3”, “intro”, “refrão”, etc., pois estas informações não estavam presentes em todas as cifras e poderiam acarretar problemas durante a extração dos descritores.

4.3. Classificadores Utilizados

Neste trabalho foram utilizados dois algoritmos de aprendizado de máquina conhecimento como k Vizinhos mais próximos (IBk), onde os exemplos são classificados utilizando a

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	33,8%	35,3%	38,1%	39,6%	29,5%	32,4%	31,7%	28,8%
Bachata	19,7%	19,7%	24,6%	27,9%	6,6%	19,7%	14,8%	13,1%
Bolero	49,3%	29,2%	33,5%	37,8%	38,3%	34,4%	30,6%	33,5%
Forró	19,5%	31,3%	32,0%	29,7%	7,0%	28,1%	18,0%	15,6%
Gaúcha	41,2%	35,3%	35,3%	34,6%	29,4%	30,1%	26,5%	22,8%
Merengue	8,7%	6,5%	6,5%	8,7%	2,2%	2,2%	0	0
Pagode	40,4%	25,7%	28,4%	31,2%	27,5%	16,5%	11,0%	10,1%
Salsa	14,3%	19,0%	15,5%	15,5%	33,3%	29,8%	28,6%	28,6%
Sertanejo	44,1%	26,6%	31,1%	36,2%	46,3%	29,4%	32,2%	39,0%
Tango	0	25,0%	10,7%	7,1%	0	14,3%	0	0
Média	34,2%	27,9%	29,7%	31,6%	28,2%	27,4%	24,0%	24,4%

Tabela 3: Taxa de acerto para os descritores utilizando a presença/ausência dos acordes.

distância euclidiana e como parâmetros 1, 3 e 5 vizinhos mais próximos. Também foi utilizado o classificador denominado de máquinas de suporte vetoriais (Support Vector Machines - SVM) que utiliza hiperplanos para realizar a classificação dos exemplos. Ambos os classificadores são comumente utilizados nos trabalhos de área e por esse motivo foram utilizados neste trabalho. Para realizar as etapas de treinamento e teste, inerente a tarefa de aprendizado de máquina, foi utilizado o método de validação cruzada fator-10 para dividir os conjuntos de treinamento e teste. Para a avaliação dos resultados foi utilizada a taxa de acerto dos classificadores.

5. Experimentos

Nesta seção queremos responder as seguintes perguntas utilizando experimentos controlados: Qual a taxa de acerto obtida pelos descritores propostos? A junção destes descritores produz melhores resultados? Qual o impacto de transpor todas as cifras para o mesmo tom na taxa de acerto dos classificadores?

As Tabelas 3 a 7 apresentam os resultados obtidos para cada um dos descritores individuais propostos neste trabalho. O melhor resultado obtido foi utilizando o classificador SVM sem transposição e com a representação baseada na frequência dos acordes (taxa de acerto = 34.2%).

A Tabela 8 apresenta os resultados obtidos utilizando a junção de todos os descritores propostos. A análise da Tabela 8 mostra que ao utilizar todas as características juntas o classificador SVM obteve uma taxa de acerto de 39% sem utilizar a transposição das cifras.

Em relação ao impacto da transposição nos descritores propostos, a análise das Tabelas 3 a 8 mostra que de modo geral a transposição afeta de forma negativa a taxa de classificação, sendo os resultados sem transposição, normalmente superiores aos resultados com transposição.

A Tabela 2 apresenta um sumário dos resultados obtidos por todos os classificadores em todas as bases. A análise da Tabela 2 mostra que para a classificação automática de gêneros musicais utilizando cifras, o classificador SVM sempre obtém melhores resultados do que o classificador IBk.

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	9,5%	7,8%	10,2%	10,2%	18,7%	0	26,6%	26,6%
Bachata	0	1,7%	2,5%	2,5%	0	53,4%	0	0
Bolero	27,1%	17,7%	24,9%	24,9%	54,0%	8,0%	53,4%	52,9%
Forró	1,6%	6,2%	1,6%	1,6%	0	6,6%	8,0%	0
Gaúcha	7,4%	5,3%	6,6%	6,6%	1,5%	0	6,6%	6,6%
Merengue	0	3,0%	0	0	0	7,0%	0	0
Pagode	2,9%	2,1%	1,0%	1,0%	7,0%	35,8%	7,0%	7,0%
Salsa	0	1,1%	0	0	0	0	0	38,6%
Sertanejo	76,7%	52,5%	76,7%	76,7%	45,5%	21,2%	36,9%	0
Tango	0	0	0	0	0	0	0	0
Média	21,1%	20,1%	20,6%	20,6%	21,0%	21,1%	21,3%	21,5%

Tabela 4: Taxa de acerto para os descritores utilizando o tom das músicas.

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	11,5%	23,7%	30,2%	27,3%	9,4%	25,2%	38,1%	30,9%
Bachata	0	8,2%	11,5%	6,6%	3,3%	8,2%	6,6%	11,5%
Bolero	84,2%	40,2%	39,2%	42,6%	84,2%	43,1%	50,7%	48,8%
Forró	0	30,5%	25,0%	19,5%	16,0%	19,5%	12,5%	12,5%
Gaúcha	30,9%	36,8%	35,3%	51,5%	5,9%	25,0%	25,7%	31,6%
Merengue	0	65,0%	0	4,3%	0	4,3%	4,3%	2,2%
Pagode	0	30,3%	17,4%	20,2%	0	26,6%	11,9%	9,2%
Salsa	0	16,7%	11,9%	13,1%	0	7,1%	1,2%	1,2%
Sertanejo	53,7%	33,9%	26,6%	36,2%	64,4%	34,5%	32,8%	37,9%
Tango	0	14,3%	0	0	0	14,3%	0	0
Média	29,4%	29,0%	25,7%	29,0%	28,2%	26,0%	25,8%	25,8%

Tabela 5: Taxa de acerto para os descritores utilizando a repetição dos acordes.

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	23,0%	25,2%	38,1%	36,0%	19,4%	23,7%	46,8%	33,1%
Bachata	49,0%	14,8%	9,8%	9,8%	0	9,8%	13,1%	9,8%
Bolero	67,5%	38,3%	34,9%	39,7%	67,5%	43,1%	47,4%	45,9%
Forró	55,0%	29,7%	27,7%	18,0%	3,9%	24,2%	18,0%	15,6%
Gaúcha	49,3%	47,1%	45,6%	51,5%	22,1%	29,4%	25,7%	31,6%
Merengue	0	13,0%	2,2%	6,5%	0	2,2%	0	2,2%
Pagode	83,0%	32,1%	27,5%	22,9%	7,3%	29,4%	10,1%	9,2%
Salsa	0	15,5%	9,5%	95,0%	0	7,1%	1,2%	2,4%
Sertanejo	69,5%	37,9%	33,3%	38,4%	70,6%	43,5%	40,1%	46,3%
Tango	34,2%	25,0%	0	0	30,1%	25,0%	0	0
Média	34,2%	31,7%	28,7%	31,0%	31,0%	19,0%	28,0%	27,4%

Tabela 6: Taxa de acerto para os descritores utilizando a frequência dos acordes.

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	23,7%	26,6%	31,7%	29,5%	30,9%	31,7%	26,6%	19,4%
Bachata	6,6%	11,5%	8,2%	1,6%	4,9%	9,8%	4,9%	3,3%
Bolero	61,7%	57,9%	62,7%	68,9%	66,0%	64,1%	74,2%	80,4%
Forró	14,1%	19,5%	7,0%	3,9%	10,2%	23,4%	9,4%	3,1%
Gaúcha	54,4%	44,1%	46,3%	39,7%	33,1%	32,4%	26,5%	18,4%
Merengue	6,5%	30,2%	0	0	2,2%	4,3%	4,3%	0
Pagode	17,4%	15,6%	3,7%	2,8%	11,9%	10,1%	1,8%	0
Salsa	29,9%	7,1%	0	0	83,0%	6,0%	0	0
Sertanejo	10,7%	23,7%	18,6%	14,7%	39,5%	29,9%	32,2%	32,8%
Tango	30,6%	14,3%	0	0	3,6%	14,3%	0	0
Média	31,0%	28,7%	25,9%	24,5%	29,9%	29,8%	27,2%	25,0%

Tabela 7: Taxa de acerto para os descritores utilizando as transições dos acordes.

Gênero	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Axé	33,1%	31,6%	42,6%	36,0%	30,7%	26,3%	39,4%	33,6%
Bachata	21,3%	31,1%	27,9%	21,3%	18,0%	21,3%	27,9%	21,3%
Bolero	45,5%	36,8%	39,2%	35,4%	34,4%	28,2%	33,5%	34,9%
Forró	33,1%	29,9%	28,3%	31,5%	19,5%	24,2%	24,2%	21,9%
Gaúcha	41,9%	39,7%	38,2%	44,1%	30,1%	39,0%	28,7%	38,2%
Merengue	10,9%	17,4%	17,4%	15,2%	4,3%	6,5%	6,5%	4,3%
Pagode	40,4%	36,7%	22,9%	27,5%	37,6%	32,1%	19,3%	21,1%
Salsa	35,7%	16,7%	9,5%	13,1%	41,7%	41,7%	35,7%	41,7%
Sertanejo	53,7%	42,4%	35,0%	38,4%	53,7%	44,6%	33,3%	36,2%
Tango	26,1%	21,7%	0	0	0	0	0	0
Média	39,0%	33,7%	31,4%	31,8%	32,7%	30,9%	29,1%	30,2%

Tabela 8: Taxa de acerto para os descritores utilizando a junção de todos os descritores individuais.

Descritor	Sem transposição				Com transposição			
	SMO	IBk-1	IBk-3	IBk-5	SMO	IBk-1	IBk-3	IBk-5
Presença e Ausência dos Acordes	34,2%	27,9%	29,7%	31,6%	28,2%	27,4%	24,0%	24,4%
Tom da Música	21,1%	20,1%	20,6%	20,6%	21,0%	21,1%	21,3%	21,5%
Quantidade de Repetições dos Acordes	29,4%	29,0%	25,7%	29,0%	28,2%	26,0%	25,8%	25,8%
Frequência dos acordes	34,2%	31,7%	28,7%	31,0%	31,0%	19,0%	28,0%	27,4%
Transições de Acordes	31,0%	28,7%	25,9%	24,5%	29,9%	29,8%	27,2%	25,0%
Junção dos Descritores	39,0%	33,7%	31,4%	31,8%	32,7%	30,9%	29,1%	30,2%

Tabela 9: Sumário dos resultados obtidos por cada descritor com os diferentes classificadores sem e com transposição.

6. Conclusões

Neste trabalho foram apresentados novos descritores para a classificação automática de gêneros musicais utilizando cifras. Ao todo foram propostos cinco novos descritores baseados na presença e ausência de acordes, no tom da música, na quantidade de ocorrência de acordes, na frequência dos acordes e na transição dos acordes. Foram avaliadas também a junção destes cinco descritores e o impacto da transposição das cifras para um mesmo tom.

Para realizar os experimentos foi criada uma base de cifras obtidas para as músicas presentes na Latin Music Database, porém só foram encontradas cifras para 1.238 das 3.160 músicas da base. O melhor resultado individual obtido foi utilizando o descritor baseado em frequência dos acordes com o classificador SVM. O melhor resultado geral obtido foi utilizando a junção de todos os descritores propostos, sem transposição, com o classificador SVM.

Como trabalho futuro pretendemos investigar formas de combinar as informações harmônicas presentes nas cifras com o conteúdo do áudio e também das letras das músicas.

Referências

- Chen, L., Wright, P., and Nejdil, W. (2009). Improving music genre classification using collaborative tagging data. In *Proc. of the 2nd ACM Int. Conf. on Web Search and Data Mining*, pages 84–93.
- Cheng, H.-T., Yang, Y.-H., Lin, Y.-C., Liao, I.-B., and Chen, H. H. (2008). Automatic chord recognition for music classification and retrieval. In *Proc. of the IEEE Int. Conf. on Multimedia and Expo*, pages 1505–1508.
- Downie, J. S. and Cunningham, S. J. (2002). Toward a theory of music information retrieval queries: System design implications. In *Proc. of the 3rd Int. Conf. on Music Information Retrieval*, pages 299–300.
- Faria, N. (1999). *Acordes, Arpejos e Escalas para violão e guitarra*. Lumyar, Rio de Janeiro - RJ.
- Kostka, S. (1999). *Materials and techniques of twentieth century music*. Upper Sadle River.
- Lee, J. H. and Downie, J. S. (2004). Survey of music information needs, uses, and seeking behaviours: preliminary findings. In *Proc. of the 5th Int. Conf. on Music Information Retrieval*, pages 441–446.
- Lidy, T., Rauber, A., Pertusa, A., and Inesta, J. M. (2007). Improving genre classification by combination of audio and symbolic descriptors using a transcription system. In *Proc. of the 8th Int. Conf. on Music Information Retrieval*, pages 23–27.
- Mayer, R., Neumayer, R., and Rauber, A. (2008). Rhyme and style features for musical genre classification by song lyrics. In *In Proc. of the 9th Int. Conf. on Music Information Retrieval*, pages 337–342.
- McKay, C. and Fujinaga, I. (2006a). jSymbolic: A feature extractor for MIDI files. In *Proc. of the Int. Computer Music Conference*, pages 302–305.
- McKay, C. and Fujinaga, I. (2006b). Musical genre classification: Is it worth pursuing and how can it be improved? In *Proc. of the 7th Int. Conf. on Music Information Retrieval*, pages 101–106.

- Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proc. of the 10th ACM Multimedia Conf.*, pages 570–579.
- Silla Jr., C. N., Koerich, A. L., and Kaestner, C. A. A. (2008). The latin music database. In *Proc. of the 9th Int. Conf. on Music Information Retrieval*, pages 451–456.
- Sousa, R. (20–). A origem das notas musicais. Brasil Escola.
- Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302.
- Whitman, B. and Smaragdis, P. (2002). Combining musical and cultural features for intelligent style detection. In *Proc. of the 3rd Int. Conf. on Music Information Retrieval*, pages 47–52.
- Zuben, P. (2005). Ouvir o som: Aspectos de organização na música do século xx. *Ateliê Editorial*.

Implementação de estrutura de reverberação de áudio digital em FPGA

Diego V. C. do Nascimento¹, Valentin Obac Roda¹, José A. N. de Oliveira¹

¹Departamento de Engenharia Elétrica – Universidade Federal do Rio Grande do Norte
Natal, Rio Grande do Norte, 59072-970, Brasil

diego.cirilo@ifrn.edu.br, valentin@ct.ufrn.br, nicolau@ufrnet.br

Abstract. *This paper presents a novel implementation in FPGA (Field Programmable Gate Array) of a standard architecture for digital audio reverberation, using allpass filters, as proposed by Manfred Schroeder, presenting the use of reconfigurable hardware as a platform for development and implementation of digital audio effects, focusing on the characteristics of modularity and reuse.*

Resumo. *Neste artigo é apresentada uma implementação inédita em FPGA (Field Programmable Gate Array) de uma estrutura de reverberação digital de áudio clássica, usando filtros allpass, baseada na proposta de Manfred Schroeder, demonstrando a utilização de hardware reconfigurável como plataforma de desenvolvimento e implementação de efeitos digitais de áudio, com foco nas características de modularidade e de reuso.*

1. Introdução

O efeito de reverberação ocorre pela reflexão do som nas superfícies adjacentes à fonte sonora durante sua propagação até o ouvinte e a resposta ao impulso de um ambiente representa suas características de reverberação. Um exemplo pode ser observado na Figura 1. Por ser dependente do ambiente, a reverberação leva ao ouvinte características do espaço onde o som está sendo reproduzido e comumente sua ausência não soa como “natural”. Como nem sempre é possível ter características desejáveis de reverberação em gravações, métodos para reverberação artificial vêm sendo desenvolvidos há décadas [Valimaki et al., 2012].

O menor custo, maior portabilidade e confiabilidade dos equipamentos digitais levou parte do processamento de áudio para tais plataformas [Owsinski, 2006], e um dos primeiros trabalhos nesse sentido foi o de Manfred Schroeder que propôs uma estrutura de reverberação de áudio digital em seu artigo “*Colorless artificial reverberation*” [Schroeder and Logan, 1961] [Zölzer, 1997]. Uma das possibilidades para processamento digital de áudio é o FPGA (*Field Programmable Gate Array*), um dispositivo

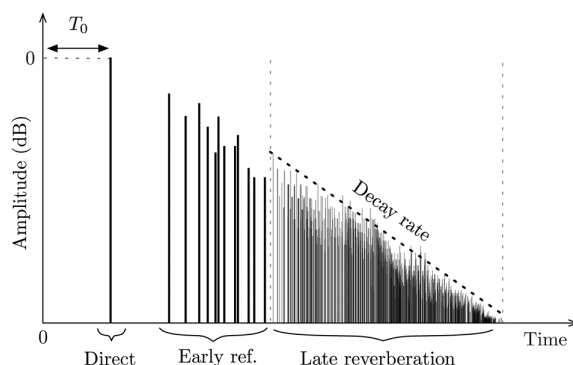


Figura 1: Resposta ao impulso de um ambiente genérico [Valimaki et al., 2012].

reconfigurável que sintetiza arquiteturas digitais descritas por meio de uma linguagem de descrição de *hardware*. Dentre as vantagens da utilização do FPGA, convém destacar a velocidade na implementação e teste de sistemas, modularidade e possibilidade de reuso [Rodriguez-Andina et al., 2007].

Considerando o exposto, neste artigo é apresentada a implementação em FPGA da estrutura de reverberação proposta por Schroeder.

2. Estruturas de reverberação de Schroeder

Em seu artigo, Schroeder apresenta a utilização de filtros *allpass* em cascata como alternativa para obter uma reverberação mais natural, se comparada ao método anterior, o uso de filtros *comb*. O grande problema de tais filtros é a sua resposta de frequência em formato de “pente”, que apresenta picos de ganho em algumas frequências resultando em um som dito “metálico”.

Os filtros *allpass* apresentam uma resposta em magnitude plana para todas as frequências, resultando em um áudio sem “coloração”. O filtro em questão pode ser implementado pela seguinte função de transferência, onde M é o atraso e g o ganho de realimentação:

$$H(z) = \frac{z^{-M} - g}{1 - gz^{-M}} \quad (1)$$

Para se aproximar da reverberação real o último passo foi a utilização de vários filtros em cascata, de forma a aumentar a densidade e reduzir a periodicidade das reflexões. [Schroeder and Logan, 1961]. Essa arquitetura, uma sequência de *allpass-filters* em cascata, é a arquitetura implementada em FPGA neste artigo.

3. Implementação

Para a implementação dos filtros apresentados em FPGA foi utilizado o *kit* de desenvolvimento *Digilent Atlys*, que entre vários dispositivos conta com um CODEC de áudio *National Semiconductor LM4550*, que segue o padrão *Intel AC’97*, trabalhando com 18 bits de resolução em até 48 kHz de amostragem [National Semiconductor, 2002]. O FPGA disponível no *kit* é o *Xilinx Spartan-6 LX45* [Digilent Incorporated, 2012]. O sistema foi descrito em VHDL (*VHSIC (Very High Speed Integrated Circuit) Hardware Description Language*) no ambiente proprietário da *Xilinx*, fabricante do FPGA. A Figura 2 mostra um diagrama de blocos do sistema em questão.

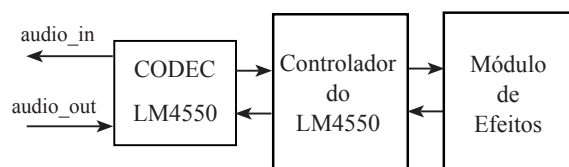


Figura 2: Diagrama de Blocos

O bloco de controle do LM4550 é responsável pelo protocolo de comunicação e configuração do CODEC de áudio, baseado na especificação do padrão AC’97.

O módulo de efeitos é responsável por implementar os filtros *allpass* necessários. O ponto mais importante do bloco é a estrutura para os longos atrasos nos sinais necessários para o efeito. Para os atrasos foi utilizado um *buffer* circular, implementado por

um bloco DPRAM com 64 KB de capacidade, o suficiente para um atraso de cerca de 500 ms a uma taxa de amostragem de 48 kHz e uma resolução de 18 bits. Para atrasos maiores seria necessário a redução da taxa de amostragem.

Seu funcionamento consiste na escrita de dados em posições incrementais, implementada por um contador simples, e a leitura com um *offset* que é dado pelo tamanho do atraso necessário. Como é possível ter atrasos diferentes apenas lendo posições diferentes, é possível replicar as estruturas usando um mesmo *buffer* circular, o que reduz o uso de blocos de memória do FPGA. Os sinais passam ainda por estágios de ganho e somadores, sendo uma implementação direta na linguagem VHDL da Equação 2, onde $x(n)$ é o sinal de entrada, $y(n)$ o sinal de saída, g o ganho e $d(n)$ o próprio sinal $y(n)$ atrasado pelo *buffer* circular na amostra anterior.

$$y(n) = g * (x(n) - g * d(n)) + d(n) \quad (2)$$

4. Resultados

Os testes foram realizados aplicando-se um sinal de áudio muito curto, simulando um impulso, na entrada do sistema e lendo-se a saída de áudio. O sinal de entrada foi gerado pelo *software Audacity*, assim como o mesmo *software* foi responsável pela gravação e corte do áudio de saída. O *software MATLAB* foi utilizado para importar os arquivos .WAV gerados pelo *Audacity* e gerar os gráficos aqui apresentados.

A Figura 3 apresenta os sinais de saída do sistema utilizando apenas um *allpass filter* e com um atraso de 200 ms, escolhido para melhor visualização. Utilizando o a possibilidade de obter a leitura de sinais em atrasos distintos devido o uso de um *buffer* circular, foram adicionados mais blocos *allpass* em série, para que a densidade de reflexões aumentasse e a característica periódica dessas reflexões fosse reduzida. O resultado para quatro blocos, apresentado na Figura 4 respectivamente com atrasos de 50, 30, 200 e 10 ms. Por fim, oito blocos *allpass* foram utilizados, e na Figura 5 é possível notar a diferença na densidade de reflexões em relação aos dois exemplos anteriores.

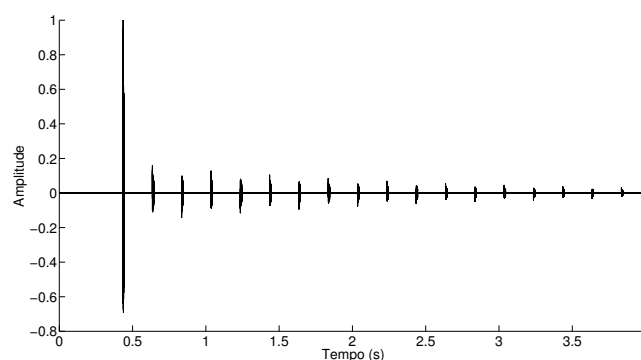


Figura 3: Resposta ao impulso, $g = 0,9$, atraso de 200 ms

5. Conclusão

Nesse artigo foi apresentada a implementação de uma estrutura de reverberação em *hardware* e algumas variações, utilizando um dispositivo FPGA. Foi possível observar o funcionamento da implementação por meio dos experimentos e medições realizados, sendo apresentados neste artigo na forma de gráficos. Observou-se a conformidade entre os resultados obtidos e os resultados esperados, de acordo com Schroeder e Logan (1961).

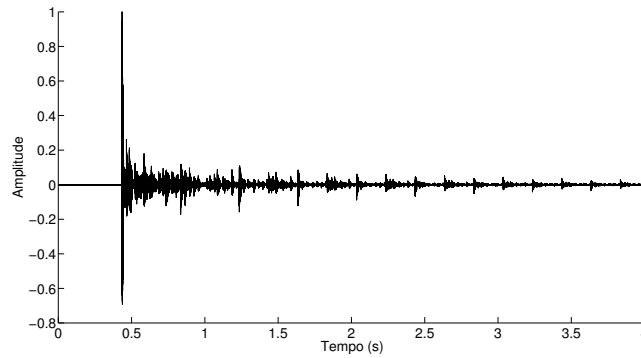


Figura 4: Resposta ao impulso, $g = 0,9$, atrasos de 50, 30, 200 e 10 ms

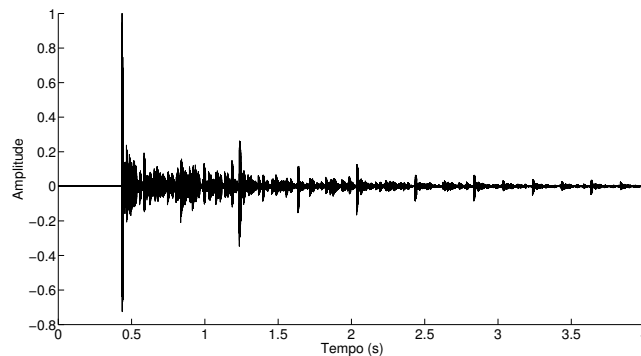


Figura 5: Resposta ao impulso, $g = 0,9$, atrasos de 50, 30, 200, 10, 80, 1, 5, 100 ms

O FPGA se mostra como uma ferramenta de prototipagem rápida para *hardware* de efeitos de áudio digital, além de ser mais barata e menos complexa de utilizar que o fluxo tradicional com componentes discretos. As características de reuso da plataforma podem ser exploradas em trabalhos futuros, onde os blocos de reverberação aqui descritos podem ser combinados com outros efeitos em uma unidade multi-efeitos e mesmo o *buffer* circular, elemento básico de atraso utilizado neste trabalho, pode ser utilizado para outros efeitos baseados em atraso, como *flanger*, *vibrato*, entre outros.

Referências

- Digilent Incorporated (2012). *Atlys Board Reference Manual*.
- National Semiconductor (2002). *LM4550 Data Sheet*. National Semiconductor.
- Owsinski, B. (2006). *The Mixing Engineer's Handbook*. Course Technology Ptr.
- Rodriguez-Andina, J., Moure, M., and Valdes, M. (2007). Features, design tools, and application domains of FPGAs. *IEEE Transactions on Industrial Electronics*, 54(4):1810–1823.
- Schroeder, M. and Logan, B. (1961). Colorless artificial reverberation. *IRE Transactions on Audio*, AU-9(6):209 – 214.
- Valimaki, V., Parker, J., Savioja, L., Smith, J., and Abel, J. (2012). Fifty years of artificial reverberation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20(5):1421–1448.
- Zölzer, U. (1997). *Digital Audio Signal Processing*. John Wiley & Sons.

FFT benchmark on Android devices: Java versus JNI

Antonio D. de Carvalho Jr¹, Max Rosan¹, André Bianchi¹, Marcelo Queiroz¹

¹Computer Science Department – University of São Paulo

{dj,maxrosan,ajb,mqz}@ime.usp.br

***Abstract.** This work presents a comparison of running times for Java and C/C++ implementations of the FFT algorithm on Android devices. We compare a pure Java implementation with the widely used FFTW library in C/C++, considering also the possibility of multi-threading. 35 different devices were benchmarked, and results on specific combinations of device model and operating system version are presented and discussed. We also discuss similarities between single- and multi-thread versions of FFTW on multicore devices and consider when developers can take advantage of each approach.*

1. Introduction

The Fast Fourier Transform (FFT) is an important algorithm for signal processing applications that can be used in many scenarios, for example, for creating tactile feedback by analyzing audio data (Lim et al., 2013), reducing noise on mobile voice communication (Jonathan and Leahy, 2010), performing face recognition (Cheng and Wang, 2011; Wang et al., 2010), and also for image processing on medical applications (Jonathan and Leahy, 2010). Since the FFT running time is $O(n \log n)$ (where n is the FFT block length in samples) (Cooley and Tukey, 1965), application development on devices with low performance may require fine-tuning some of the FFT internal details in order to ensure realtime operation.

FFTW (<http://www.fftw.org/>) is one of the fastest and most used FFT libraries (Lin et al., 2011), and its use on Android devices would appear to be a step forward for realtime signal processing compared to pure Java implementations on the Application level. Nowadays, multicore devices are becoming cheaper, thus turning multi-thread into a good approach both for developers and for users. On the other hand, since saving battery in mobile devices is also a primordial concern, using more than one processor just for the FFT algorithm is not an easy task. Another important consideration concerns devices' specific scheduling policies that might decide when to split processing into two or more cores or not. Multi-thread methods can give strange results depending on each device model internal peculiarities.

2. Methodology

This work presents the results of a benchmark of the realtime FFT on blocks of varying sizes using a Java implementation and the FFTW library, which is written in C/C++ and is called through the Java Native Interface (JNI). We have set up an environment to run arbitrary DSP algorithms over an audio stream segmented into blocks of N samples, allowing for the variation of algorithm parameters during execution. The software used is the Android DSP benchmarking application (Bianchi and Queiroz, 2012), an open source project available at <https://github.com/andrejb/DspBenchmarking/>, with some modifications to include the FFTW via JNI. To compare the performance of different implementations of the FFT algorithm, we considered a pure Java implementation, the single-thread

FFTW and the multi-thread FFTW. As the FFTW implementation is written in C, JNI was used to include the code into the benchmark. All performance measurements are made by the application started by the user. User interactive assistance is kept at a bare minimum, by starting the experiment and pressing a button to e-mail the results to the authors. To obtain as many results as possible, we launched an open call for participation through e-mail, and got responses comprising 35 different devices. Instructions were sent to stop all applications and turn off communication (tests could only be started after the user enabled flight mode) to impose an "idle" scenario on every device. The result of imposing these constraints is an overall experiment that automatically cycles through all benchmarking algorithms, and then sends an e-mail report with results back to the authors.

3. Results and discussion

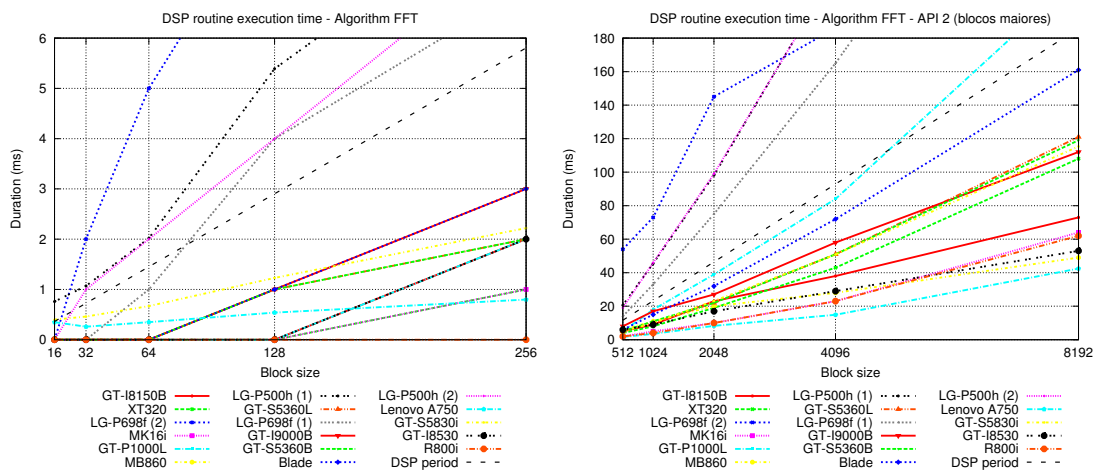


Figure 1: Benchmark of FFT implemented in pure JAVA in devices with API version 2.X for smaller (above) and larger (below) block sizes.

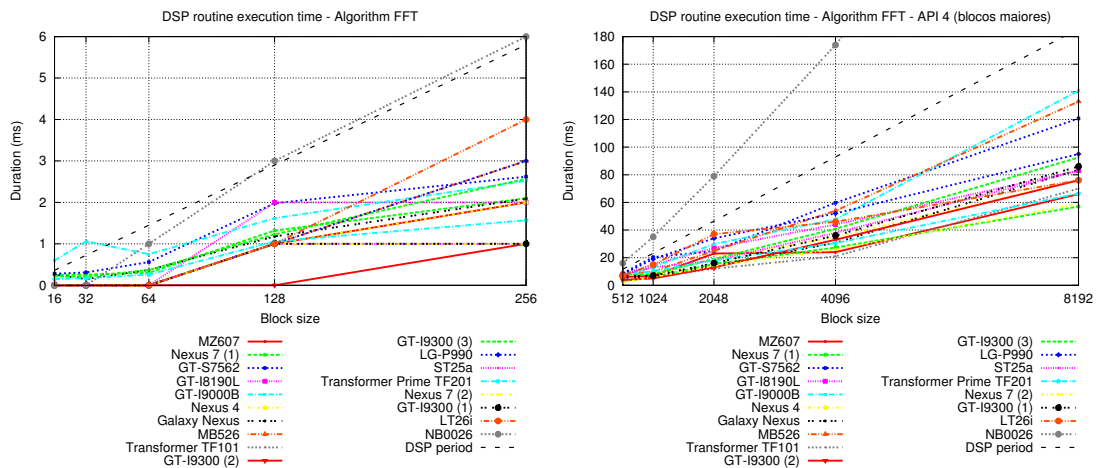


Figure 2: Benchmark of FFT implemented in pure JAVA in devices with API version 4.X for smaller (above) and larger (below) block sizes.

Figures 1 through 6 show the time taken by each of the FFT algorithms to be executed on each device as a function of block size. By comparing the Java FFT and single-thread FFTW results, on Figures 1, 2, 3, and 4 respectively, it is possible to notice that single-thread FFTW is faster than Java FFT on blocks larger than 16 samples. For $N=16$ the overhead of loading a dynamic library is not compensated by the efficiency of

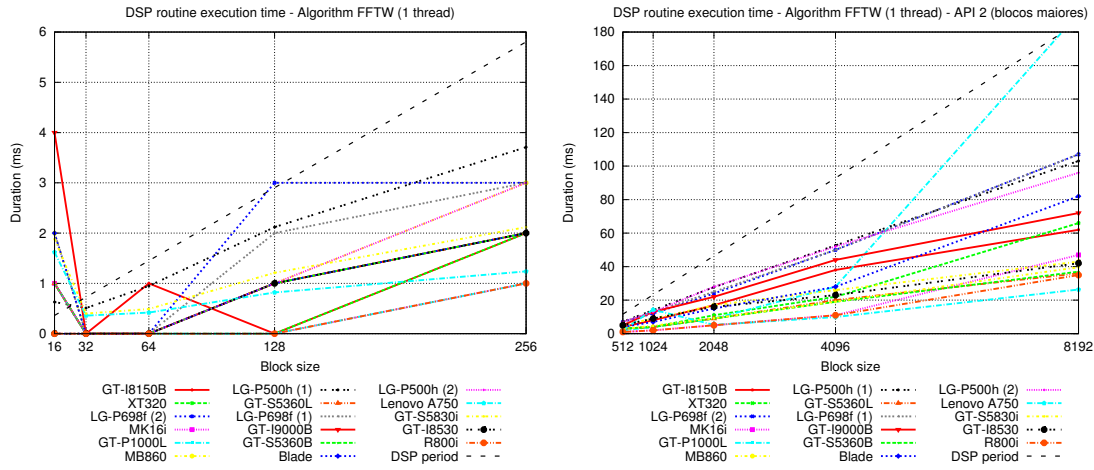


Figure 3: Benchmark of FFT using single-thread FFTW in native code in devices with API version 2.X smaller (above) and larger (below) block sizes.

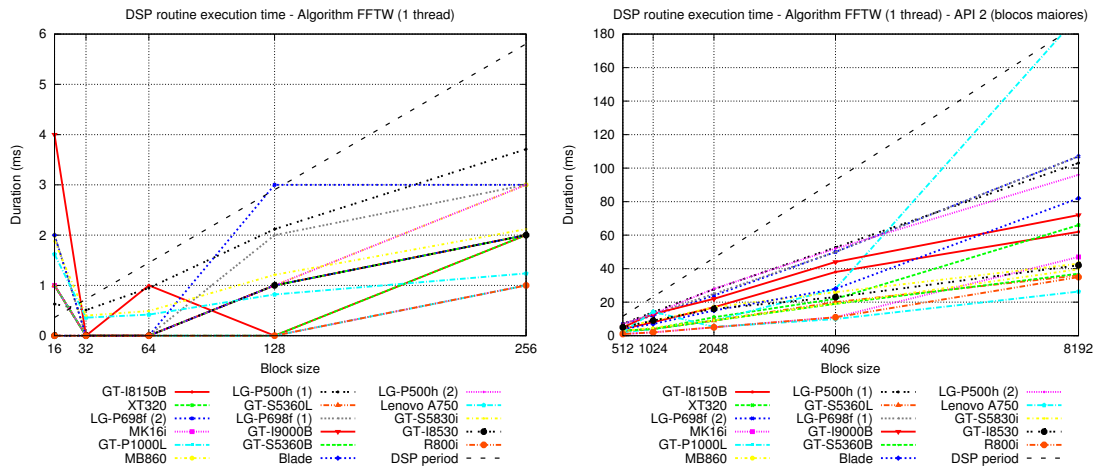


Figure 4: Benchmark of FFT using single-thread FFTW in native code in devices with API version 4.X smaller (above) and larger (below) block sizes.

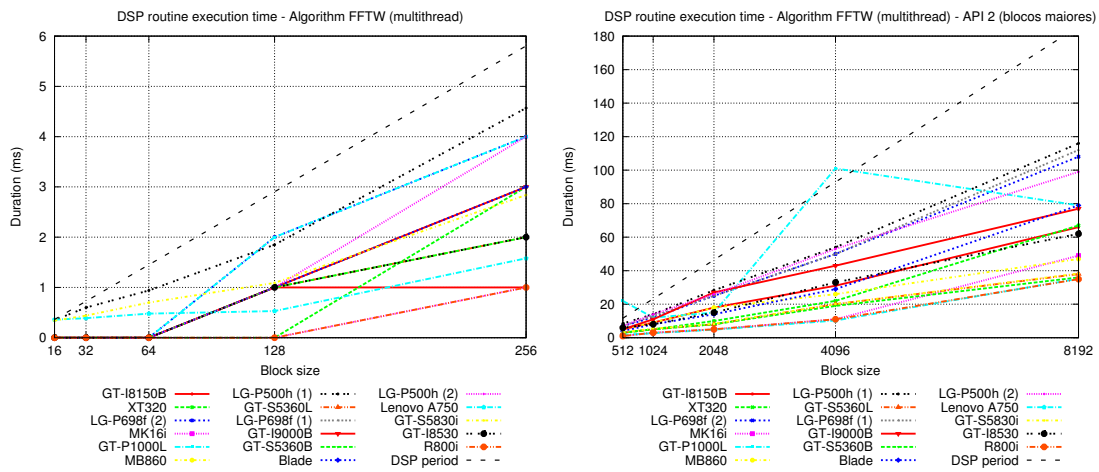


Figure 5: Benchmark of FFT using multi-thread FFTW in native code in devices with API version 2.X for smaller (above) larger (below) block sizes.

the C code, so that the results for the single-thread FFTW are actually worse than for the Java FFT, but for larger block sizes the library is already loaded. The same is observed with the multi-thread FFTW: it is executed after the single-thread FFTW and therefore

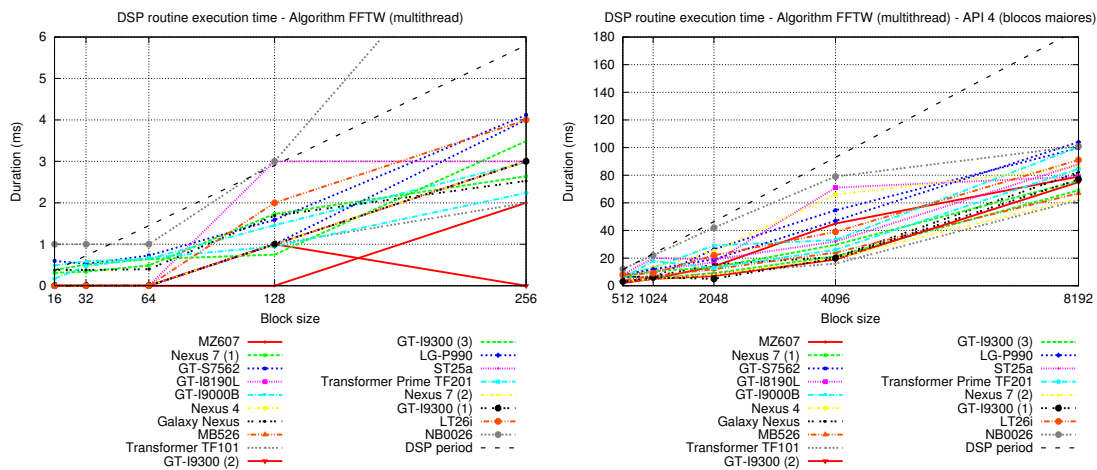


Figure 6: Benchmark of FFT using multi-thread FFTW in native code in devices with API version 4.X for smaller (above) larger (below) block sizes.

there is no overhead for the initialization of the library.

It is also possible to notice on Figures 5 and 6 that FFTW has worse performance when it is used with multiple threads. As it turned out, the Android kernel has different policies in different devices, and some devices move threads to different cores only when they are CPU-intensive and have been running for a sustained period. The use of native code does not automatically imply better performance on every situation. It had been noticed that using JNI increases application complexity and also has a cost associated with calls to non-Java code, which makes it indeed worse for some applications (Lin et al., 2011). Nevertheless, the implementation and comparison with native code for realtime signal processing evaluated in our work show that there are more subtleties that should be kept in mind, like the amount of data processed and the device specification about starting cores and moving threads around.

References

- Bianchi, A. J. and Queiroz, M. (2012). On the performance of real-time dsp on android devices. *Proceedings of the 9th Sound and Music Computing Conference*, pages 113–120.
- Cheng, K.-T. and Wang, Y.-C. (2011). Using mobile gpu for general-purpose computing: a case study of face recognition on smartphones. In *VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on*, pages 1–4.
- Cooley, J. and Tukey, J. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301.
- Jonathan, E. and Leahy, M. (2010). Investigating a smartphone imaging unit for photoplethysmography. *Physiological Measurement*, 31(11):N79.
- Lim, J.-M., Lee, J.-U., Kyung, K.-U., and Ryou, J.-C. (2013). An audio-haptic feedbacks for enhancing user experience in mobile devices. In *Consumer Electronics (ICCE), 2013 IEEE International Conference on*, pages 49–50.
- Lin, C.-M., Lin, J.-H., Dow, C.-R., and Wen, C.-M. (2011). Benchmark dalvik and native code for android system. In *Innovations in Bio-inspired Computing and Applications (IBICA), 2011 Second International Conference on*, pages 320–323.
- Wang, Y.-C., Pang, S., and Cheng, K.-T. (2010). A gpu-accelerated face annotation system for smartphones. In *Proceedings of the international conference on Multimedia, MM '10*, pages 1667–1668, New York, NY, USA. ACM.

Sound processors for live performance

Eduardo da Silva Afonso¹, Ruy Borges da Cunha Junior¹, Regis Rossi A. Faria¹

¹Laboratório de Acústica e Tecnologia Musical – (LATM),
Departamento de Música, Universidade de São Paulo (USP)
Ribeirão Preto –SP – Brazil

eduardo.afonso@usp.br; ruy.cunha@usp.br; regis@usp.br

Abstract. *We introduce a set of sound processing tools under development for live music applications using the Pure Data (Pd) patch-oriented framework. The development addresses aspects such as ease of interaction with the user and ease of tool integration in real-time applications. Three tools are presented on the current stage: a progressive filter, combining stop-band and band-pass types with time-varying partial tuning capability; a time-stretcher and pitch transposition tool based on granular synthesis; and an augmented interface to integrate a MIDI control surface with the patches in Pd. Several differential features are addressed in the current design, such as an individual tagging scheme to easy address control metadata all over the patches.*

1. Introduction

Filters and time-frequency modifiers are not novelties in the computer music scene, but means to explore their creation, their application to a sound, and its operation are always a typical arena for innovations.

Live composition and performance employing sound processors are often limited by exhausting procedures to install software, to apply and integrate them into a patch of processes, and to operate all the parameters so to adjust the behavior as desired. It is not rare to find difficulties in achieving the correct expected reaction following a calibration action. The need of adjustments on levels and ranges and the lack of resources to monitor the result of partial operations add more complexity and insecurity.

Aiming to approach interaction and integration issues of applying sound processors in live applications, we are exploring new avenues to design and to use sound processors modules on a patch-oriented framework. A first generation of tools produced in the current stage of this research project is presented in the following sections.

2. The sound processors

Three tools – two sound processors and an augmented control interface – were developed in the form of abstractions in the Pd (Pure Data) patch-oriented framework [1].

The first tool designed was so called a *progressive filter*, i.e., an implementation of band-pass and stop-band filters that can be connected in parallel or in series and can have their central frequency (f_c) and tuning parameter (Q) vary in time and tuned to specific harmonic partials of a tone. What makes this implementation differ from others lies in two features: first, it is possible to immediately tune the filter over the fundamental frequency of a harmonic tone, access its first 10 partials on-the-fly and

then sweep it, progressively transforming its spectral density; second, it is possible to address the filter parameters using a tagged message to it.

The progressive filter module combines in one block three types of filters which can be selected one at a time per instantiated object: a 1st order band-pass filter; a 2nd order band-pass filter; or a 2nd order stop-band filter. A slider controls the central (or stop) frequency (f_c) and another controls the quality factor level (Q). It also includes a fundamental (f_0) and partial frequency selector (up to 10 partials) to alternatively determine a frequency around which the filter shall operate, a RMS level indicator (VU) and a gain controller, plus a toggle button to turn on/off. These auxiliary resources add monitoring capability to the module, extremely useful for live situations.

Figure 1 shows the processor band-pass-filtering a white noise signal centered on the 4th partial (800 Hz) of a tone of fundamental frequency 200 Hz with a Q=10. A log-spectrum window at right shows the filtering profile resulting.

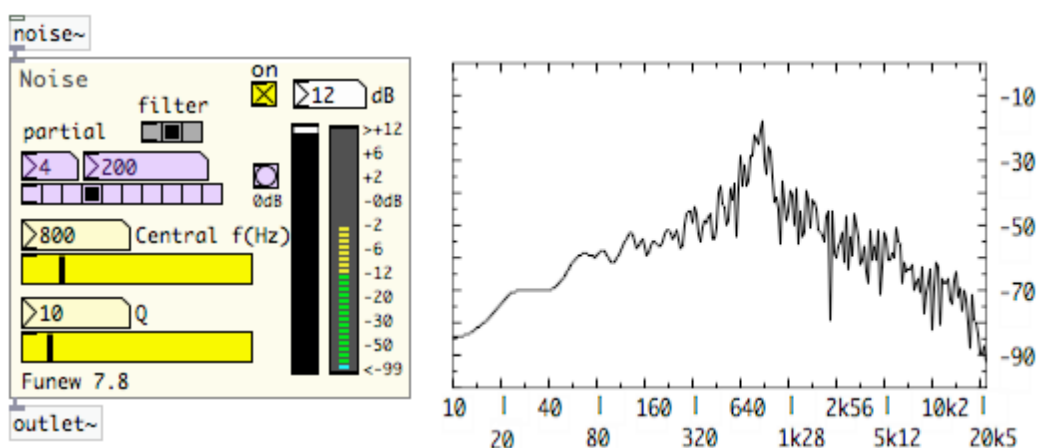


Figure 1: Progressive filter over a white noise, an a log-spectrum example

A differential feature in this tool is the possibility of having metadata generators control the filter behavior by sending commands to modify its parameters, either directly seeding its inlet ports or by sending messages to one of its parameter. For example one can have a pre-programmed table driving the fundamental frequency f_0 and another changing the Q factor of an individual instantiation filtering a specific sound.

The second tool developed is a *time-pitch processor* for changing the duration and/or the pitch of sound objects, a case of time-frequency stretching tool. This type of processor is already available in various programs and devices for sound processing [3] [4], but particular operating requirements and features are very frequent, demanding customized implementations.

The original concept can be extended in many ways and there are several forms to implement the effect. In this research, the theory of granular synthesis was the theoretical basis for its construction [5]. According to this theory, the signal can be enveloped in grains, which define portions of the signal to be sequenced in time in a determined speed, without changing its pitch.

The current abstraction was built inspired by reference implementations found in the literature and on the web [6]. Major additions and modifications were done however,

including operations in the grain formation algorithm, modularity, controllability and pitch transposition capability.

Figure 2 shows the time-pitch processor interface. One can stretch or shorten the duration of the signal and independently alter its pitch. Two sliders, number boxes and preset bars are provided to define levels for stretching and for pitch transposition in standard tempered scale intervals. The sound is loaded into an internal array, shown on left up. The grain size, shape and generative parameters can be customized, including through manipulation of the grain envelope, by direct editing its curve over the array.

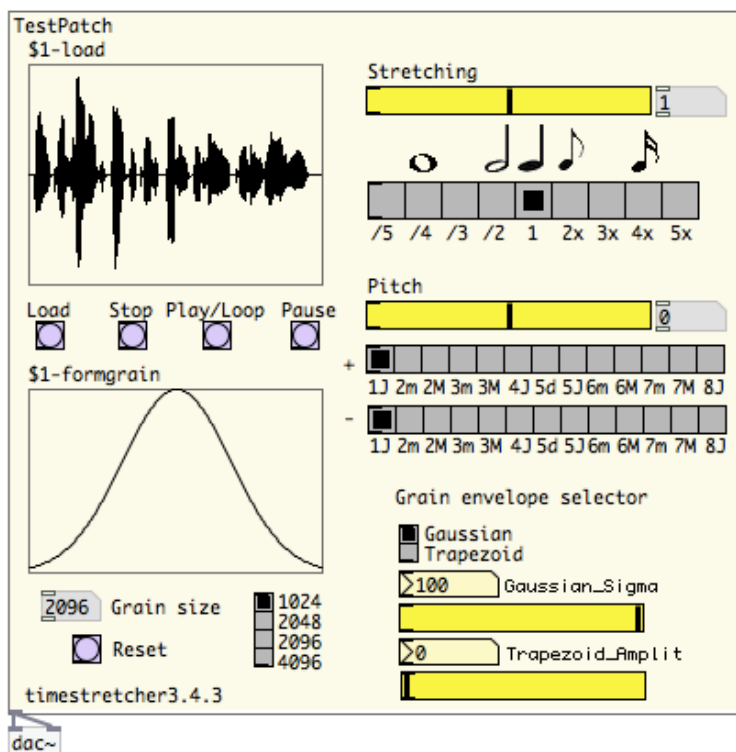


Figure 2: The time-stretcher and pitch transposition tool

The third tool developed aims to augment the addressing capabilities of an existing MIDI controller: a KORG nanoKontrol2 [2]. Its interface, shown on figure 3, permits not only to monitor and alter the values of buttons and sliders but it also extends the functions available in the original hardware, by permitting to assign each column (or channel) buttons and sliders to control addressable parameters of any processor module existing in the patches.

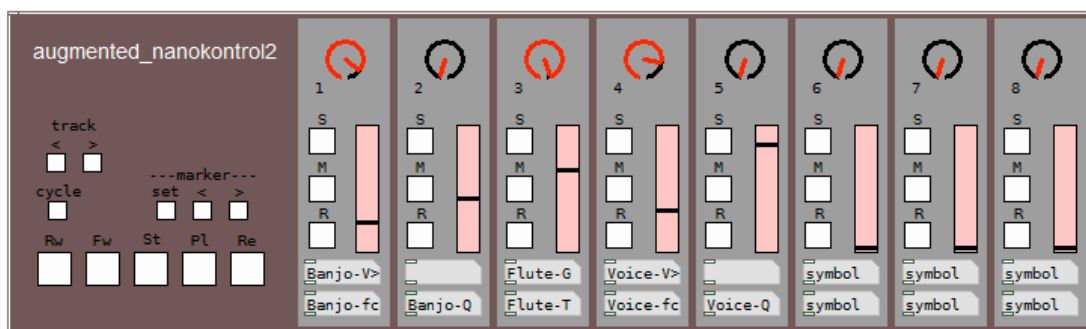


Figure 3: Augmented nanoKontrol2 GUI

Knobs and sliders are easily addressable through *tags* named by the user. For example, one can assign the gain control of a progressive filter called *Banjo* ("Banjo-Vol") to the column 1 knob, and its central frequency ("Banjo-fc") to the column 1 slider by just typing the tags in appropriate symbol fields in the GUI.

3. Conclusions

The spectral and time-pitch processors plus the augmented controller interface are the initial building blocks towards a more comprehensive library of tools for live music processing under development. We are particularly interested in investigating the processes of producing and adapting software to specific demands, mostly related to the creation and shaping of musical sounds in real time interactive applications, with a focus on usability and controllability.

Several features presented throughout this article make this development flexible enough to meet the needs of a composer building a real-time generative music application, and controlling it using a patch-oriented framework. These include a standard design to permit individual processors instantiation, integration, and a unique tag assignment scheme for addressing their individual parameters all over the patches.

In our tests we also found that many issues can affect the quality of the time-stretcher, such as the grain shape, size, inter-grain distance, and phase synchrony, which can for example introduce graininess in extreme stretching conditions. These issues are under investigation now and are to drive improvements for a second generation of tools.

4. Acknowledgments

The authors thank CNPq Brazilian National Research Council (process no. 134887/2012-3) and Dean's Office for Undergraduate Studies of the University of São Paulo for supporting this research through their respective institutional programs of Undergraduate Research (PIBIC) and Teaching with Research.

5. References

- [1] Puckette, M. Pd documentation, 2012. In: <http://crca.ucsd.edu/~msp/>. Access on 31 Dec. 2012.
- [2] KORG NanoKontrol2. In: <http://www.korg.com/nanoseries2/>. Access on 31 Dec. 2012.
- [3] Jaroszewicz, M. Spectral Tools. In: <http://dl.dropbox.com/u/25820520/SpectralToolsSource.zip>. Access on 31 Dec. 2012.
- [4] Charles, Jean-François. A tutorial on Spectral Sound Processing using MAX/MSP and Jitter. *Computer Music Journal*, 32:3, p.87-102, Fall 2008.
- [5] Roads, C. Introduction to Granular Synthesis. *Computer Music Journal*, Vol.12, No.2, p.11-13, 1988.
- [6] PURE DATA forum~ / Time Stretching. "timestretch_mmb_bd". In: <http://puredata.hurlleur.com/sujet-2779-1.html>. Access on 31 Dec. 2012.

Uma roupa computacional para apoiar a educação musical no contexto da sala de aula.

Thiago Marcondes Santos¹, Denise Filippo², Mariano Pimentel³, Monica Duarte⁴

^{1,3} PPGI-Programa de pós Graduação em Informática – Universidade Federal do Estado do Rio de Janeiro (UniRio) Av. Pasteur 458, Térreo, Urca, 22290-240, Rio de Janeiro, RJ, Brasil.

⁴ Instituto Villa Lobos – (UniRio)

² Universidade do Estado do Rio de Janeiro (UERJ) Escola Superior de Desenho Industrial (ESDI)

Rua Evaristo da Veiga 95, Lapa, Rio de Janeiro, RJ, Brazil

thiago.marcondes@uniriotec.br, dfilippo@esdi.uerj.br
pimentel@unriotec.br, monicaduarte33@gmail.com

Abstract. *Brazilian elementary schools are behind in the adoption of computer technologies used to aid teaching in a classroom environment. This article describes the phases of construction of a prototype that intends to support music lessons in the context of a classroom through the use of ubiquitous computing allied with teaching theories. The article also proposes a debate on the future of music education and its evolution along with the evolution of new computer technologies.*

Resumo. *As escolas brasileiras de ensino fundamental estão defasadas em relação à absorção de tecnologias computacionais para o ensino em sala de aula. Este artigo descreve as fases de construção de um protótipo que visa dar suporte às aulas de música no contexto da sala de aula através da utilização de computação ubíqua conjugada com teorias de aprendizagem. O artigo também apresenta um debate sobre os rumos da educação musical e sua evolução junto às novas tecnologias computacionais.*

1 Introdução

Os avanços em computação fornecem uma gama enorme de sistemas e equipamentos com potencial para serem adaptados às mais variadas aplicações e demandas. No entanto esses avanços ainda não estão presentes no cotidiano das escolas de ensino fundamental do Brasil.

Neste trabalho foi projetada uma roupa computacional para ser conjugada com a ação do professor visando possibilitar aos estudantes o acesso à exploração do discurso musical e sonoro. A motivação para essa pesquisa parte da constatação de que são poucas as propostas de mudança no paradigma da educação musical em salas de aula. Existem inovações ligadas à música nas áreas de produção e comercialização, mas ainda não são encontradas muitas pesquisas na educação musical e as possibilidades de interação desta com a informática. Boas alternativas para a construção de novas formas de se relacionar com a música e aprender a natureza do seu discurso, terão grande impacto nas escolas uma vez que a lei 11.769 de 2008 também vem reforçar a importância do aprendizado e da expressão musical pelas crianças durante sua formação em um ambiente cultural diversificado e rápido em inovações.

1.1 Abordagem: pesquisa-design, um processo iterativo teoria ↔ artefato

Nesse trabalho adotou-se a abordagem denominada “Design Science Research” (Hevner et al., 2004; Hevner e Chateerjee, 2010) que, assim como a pesquisa-ação, também tem um duplo objetivo: desenvolver tanto o design de um artefato quando realizar uma pesquisa científica – conforme esquematizado na Figura 1.

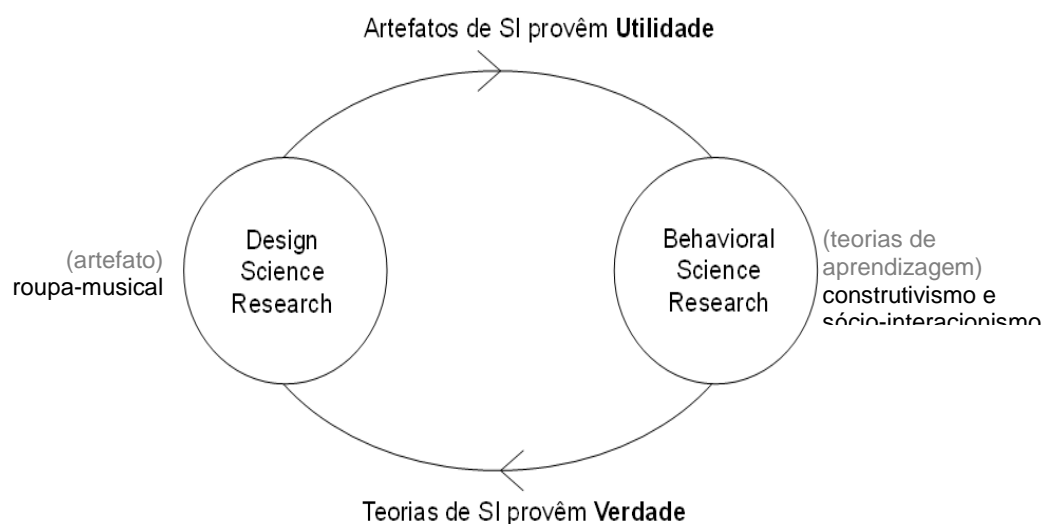


Figura 1. Pesquisa-design

No presente trabalho, a roupa-musical é o artefato projetado que encapsula um conjunto de suposições sobre as teorias de aprendizagem, isto é, projetamos o artefato em função de nossas suposições baseadas em teorias sobre como os sujeitos aprendem (discutido na seção 2) e como interagem com a tecnologia (discutido nas seções 2 e 3). Enquanto o artefato é uma concretude do referencial teórico, o uso do artefato possibilita teorizar sobre o produto. O uso possibilita avaliar se as suposições encapsuladas no artefato se verificam na prática, e a partir dos dados é possível realizar generalizações (estatísticas ou analíticas) e concluir sobre a teoria. A empiria (o uso do artefato) retroalimenta o design com indicações sobre o que mudar no artefato, e retroalimenta o referencial teórico com a incorporação, refinamento, confirmação ou refutação de conjecturas teóricas.

É muito usual que essa atividade oscilatória em busca de uma solução de design se realize apoiada num processo iterativo de prototipação, pois um protótipo (atividade prática) possibilita a reflexão (atividade conceitual), que por sua vez leva a um novo protótipo e assim sucessivamente até que a solução para o problema seja considerada satisfatória.

Na seção 2, foi apresentado o panorama da educação musical segundo diferentes teorias enquanto na seção 3 foram levantados trabalhos correlatos. Na seção 4 é descrito o processo de prototipagem e a seção 5 conclui o artigo.

2 Cibercultura e educação para a sociedade contemporânea

2.1 Implicações da cibercultura para a educação

Levy (1999) apresentou a definição de cibercultura que se tornou clássica: “*um conjunto de técnicas, práticas, atitudes, modos de pensamento e valores que se desenvolvem dentro do “ciberespaço”, meio de comunicação que surge da interconexão mundial de computadores*”.

As mudanças tecnológicas possibilitam a alteração de procedimentos e a reconfiguração de um ambiente onde os estudantes interagem entre si e com o conhecimento de uma forma bem característica. Com a facilidade de acesso a

informação, o ciberespaço vem impondo certos desdobramentos e a interdisciplinaridade e a ligação entre as diversas áreas do conhecimento são uma consequência direta do processo, segundo Malaggi e Marcon (2011).

2.1.2 A realidade da escola hoje

No caso específico da educação musical existem muitos obstáculos à sua implementação no cotidiano das escolas, como por exemplo, a falta de instrumentos para todos os estudantes, tratamento acústico ou soluções que permitam a exploração sonora sem prejudicar as outras classes, atividades estimulantes para o aprendizado musical que se integrem ao contexto social e cultural /tecnológico dos estudantes. Gohn, (2007) também destaca problemas de resistência por parte de profissionais da educação sobre a entrada de ferramentas computacionais no ensino musical.

O mundo evolui em direção a uma sociedade do conhecimento e a escola que não acompanhar esse processo afasta-se de seu público alvo, os estudantes e a comunidade na qual é inserida.

O computador possibilita mais opções de interação entre os estudantes e o conteúdo abordado. O professor deve direcionar as atividades de maneira a usar o computador como um suporte para trabalhar as áreas do conhecimento em questão.

Existe uma grande defasagem tecnológica entre as ferramentas educacionais e as ferramentas de lazer dos estudantes no seu dia a dia. A realidade tecnológica da escola está atrasada em relação a realidade dos alunos em seu cotidiano fora da escola. Eles escutam música pela internet, mandam clipes e vídeos para os amigos em suas páginas pessoais. Ensinam, muitas vezes, os pais e avós a navegarem pela web e abrirem uma conta de email. A informática já faz parte dessa geração desde cedo, concretizando o que foi previsto por Mc Luhan (1969 *apud* Lima, 71, p.8).

“Haverá um dia - talvez este já seja uma realidade - em que as crianças aprenderão muito mais, e mais rapidamente, quando em contato com mundo exterior do que no recinto da escola”.

Os índices de evasão nas escolas são grandes em todas as faixas etárias em nosso país. A escola não atrai seu público alvo como as *lan houses*, cinemas, bate-papos virtuais, entre outros. Esses ainda geralmente são exteriores ao espaço pedagógico da sala de aula tradicional.

2.2 Teorias de aprendizagem

Nessa seção serão vistos modelos e teorias que apoiam o professor de música.

2.2.1 Modelo Espiral e Modelo TECLA (Swanwick)

Keith Swanwick, educador musical inglês, criou o Modelo CLASP (1979/2003), traduzido do inglês para TECLA, em que apresenta os seguintes parâmetros da Educação Musical:

Técnica, Execução, Composição, Literatura e Apreciação.

Este autor enfatizou em seu livro “Teaching music musically”(2007) que o processo de aprendizado musical está calcado na conjugação sem hierarquia de nenhum dos elementos chave do modelo TECLA. Esses elementos devem ser bem dosados pelo educador musical para que os estudantes tenham diferentes visões do mesmo fenômeno.

Swanwick também defende uma teoria baseada na abordagem de Piaget de que o desenvolvimento cognitivo se dá por etapas complementares. Pesquisou com crianças de escolas inglesas e organizou um modelo de espiral crescente de desenvolvimento cognitivo em que existem fases que se sobrepõem ao longo do amadurecimento e do entendimento do discurso musical da infância até a adolescência. Elaborou o que denominou Modelo Espiral – figura 2. Ao lado esquerdo da espiral se colocam elementos mais sensoriais e ao lado direito ficam os elementos lógicos do processo. A passagem pelos dois lados da espiral é obrigatória onde a um momento sensorial segue-se a acomodação do que foi sentido e se eleva mais um pouco de fase continuando o mesmo movimento espiral ascendente em direção ao valor da experiência musical vivida pelo estudante.

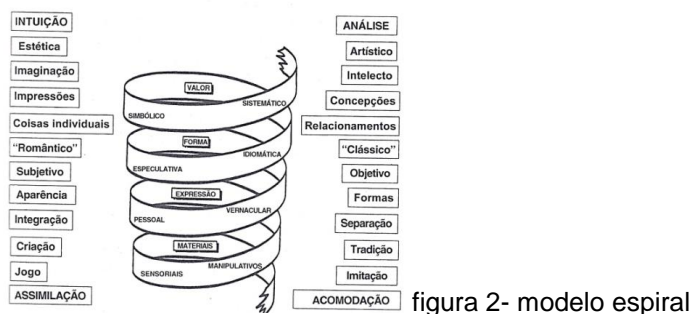


figura 2- modelo espiral

O professor deve monitorar o andamento da aula de forma a equalizar esses elementos e integrá-los com as ações dos estudantes, possibilitando ao estudante ascender na espiral e passar por todas as quatro fases e oito níveis.

2.2.2 Eúritmia (Dalcroze)

Emile Dalcroze, educador suíço que em seu trabalho *Music Rythm and Education*, publicado em 1920, já criticava as velhas formas de se educar. Para Dalcroze, os conservatórios de música se prendiam em demasia às formulas de execução e de repetição de exercícios técnicos para os dedos deixando de lado boa parte do fenômeno musical, como sua expressão, dinâmica e colorido tonal. Existia naquela época uma maior preocupação com a parte racional e cerebral do que com o som em si. O ouvido não entra na equação e não há atividades para o desenvolvimento das habilidades aurais. O adestramento muscular não implica em crescimento musical para ele. A ênfase no adestramento e repetição estafante apenas torna o estudante mais distante da música e mais robótico em sua execução. O uso dos movimentos para a metáfora sonora é valorizado e incentivado na abordagem de Dalcroze que traz os elementos do discurso musical para a vivência corporal do sujeito na aprendizagem.

3. Revisão de trabalhos na área

Ao se analisar os trabalhos onde ocorriam interações da tecnologia com a música foi feita uma divisão em duas categorias básicas:

1. Informática auxiliando o fazer musical/sonoro
2. Informática auxiliando especificamente a aprendizagem musical no contexto da sala de aula

Muitos são os trabalhos que utilizam a tecnologia para a produção de som seja ele parte de um discurso musical ou apenas forma de expressão sonora como Yang (2011), Zhang (2011) , Mann et AL (2011), Yang (2011) e Gelineck (2012). Estes trabalhos propõem novas formas de se produzir som mas não focam em aspectos e ou teorias de aprendizagem. Por outro lado os trabalhos de Volpe et al (2012), Zhou et al (2011), indicam uma direção a se olhar em termos de ferramentas construídas especificamente para o contexto de sala de aula e aprendizado musical. O controle sobre o som gerado e as notas por parte dos estudantes ainda pode ser aprimorado, mas

interessantes soluções como a de Zhou et al (2011) estabelecem considerações em relação ao espaço da sala de aula e o fazer musical, oferecendo a alunos e professores recursos que superam as adversidades da educação musical em ambientes físicos não estruturados e preparados para tal atividade.

Nesse caso o uso de tecnologias mobile permitiu ao professor separar a classe em grupos que através dos fones de seus aparelhos mobile pudessem se ouvir em um modo de apresentação coletiva e também dividida por canais onde cada grupo só escuta seus elementos facilitando a concentração nas tarefas. Outra observação é que o número de trabalhos com interfaces gestuais e música está aumentando como pode ser visto em Gelineck(2012), Ferguson (2006), e Horace(2005) assim como os trabalhos que incorporam música, tecnologia e vestuário como os de Yang (2011), e Volpe et al (2012) onde é possível explorar os conceitos de Euritmia de Dalcroze como suporte ao aprendizado musical trazendo o aprendizado para o corpo através de vestes computacionais e sensores dos mais variados.

O fato de se facilitar o acesso ao som fabricando-o com gestos corporais e não mais com o desenvolvimento técnico e virtuosístico característico dos instrumentos musicais tradicionais reforça as ideias de Swanwick de gestos/sentidos sonoros e sua espiral de desenvolvimento musical que passa pelo modelo TECLA Swanwick (2007) além das considerações de Dalcroze sobre a mediação do corpo no processo de aprendizagem do discurso sonoro.

4-Protótipo

4.1 Projeto inicial: uma roupa-musical colaborativa.

Definiu-se que seria desenvolvida uma roupa-musical e foi levantada a proposta de investigação da colaboração na aprendizagem em música. A escolha da roupa como meio de interação deve-se a necessidade de transportar para o corpo e seus gestos o fazer musical. O passo seguinte foi elaborar um protótipo conceitual que induzisse os estudantes a interagirem uns com os outros.

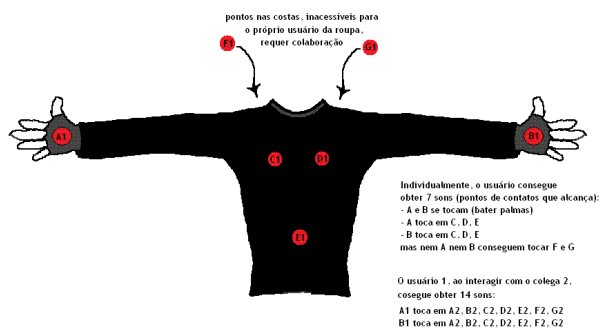


Figura 2. Roupas-musical para induzir a colaboração

A ideia inicial da roupa era possibilitar produzir um som ao bater palmas, individualmente ou com o colega como se estivessem brincando de “adoleta”. Depois o projeto evoluiu para que a roupa possibilitasse o usuário tocar as sete notas padrão da escala maior de dó (dó, ré, mi, fá, sol, lá, si). Assim, considerando que cada mão produziria um som diferente, com três pontos de acionamento na frente da roupa seria possível produzir seis notas, e ao bater palmas seria produzida uma sétima nota.

A roupa deve possibilitar liberdade de movimentos, o que implica em tecnologia wireless para não termos crianças pulando e se mexendo pela sala com fios pendurados em equipamentos sensíveis. A roupa deve ter contatos que, uma vez acionados, emitem comandos para um computador central que processa os comandos de acordo com a dinâmica da atividade educacional em questão. O processador gera os sons e os sobrepõem de modo a permitir aos estudantes interagirem uns com os outros.

Para produzir som com boa qualidade não se deve utilizar um speaker, mas sim emitir o som de um computador central que então vai processar os sinais de acionamento recebidos das roupas como feito em Yang (2011). Primeiramente optou-se por colocar os contatos em pontos móveis pelo tórax utilizando velcro.

Devido ao fato das linhas condutivas serem suscetíveis a problemas de contato optou-se por utilizar fios conectados entre o Arduino e os pontos de contato para gerar som. Nesta primeira fase do projeto optou-se por descartar a colaboração entre os processadores Arduinos das roupas através dos transmissores-receptores wireless XBee já que sua configuração é mais complexa e demanda mais tempo de implementação.

4.2 Simplificação do projeto: uma roupa para produzir música individualmente.

O objetivo foi redefinido para se construir uma roupa-musical que, ainda que tocada individualmente, servisse como “prova de conceito” para ilustrar que é possível desenvolver uma roupa-computacional para ser usada para a aprendizagem de música. A partir das lições aprendidas na etapa anterior, foi elaborado o protótipo conceitual da roupa esquematizada na Figura 3. O uso de fios no lugar das linhas condutivas trouxe problemas para o acesso e configuração dos pontos de forma móvel. Optou-se por fixar os pontos no tórax e fazer a troca de sons dos contatos pelo software.

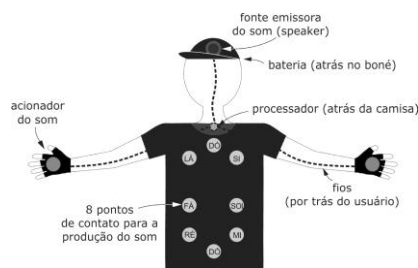


Figura 3. Roupa-musical para servir de “prova de conceito”

Uma decisão neste projeto “prova de conceito” foi produzir o som na própria roupa do usuário, sem assim precisar da transmissão sem fio. Foi assim projetado um boné para dar suporte ao speaker.

4.3 Protótipo descartável da roupa: ergonomia e customização



Figura 4. Protótipo descartável

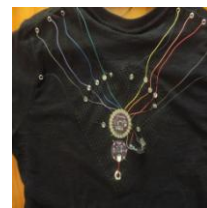
Após a construção do circuito e da programação do Arduino, o passo seguinte foi criar um protótipo descartável para ajustar a posição dos pontos de acionamento no corpo do usuário. Para esse tipo de protótipo, o circuito foi montado sobre a camisa usando fita Durex, conforme pode ser visto na Figura 4, em que a roupa foi fotografada

em sua configuração final do circuito após alguns ajustes decorrentes de testes. Esse protótipo foi montado em 5 de outubro de 2012.

4.4 A roupa apresentada



ã) frente, com fios embutidos



b) costas, com exibição dos fios e dos circuitos, mas com a bateria embutida



c) placas para o acionamento do som



d) speaker para a produção do som

Figura 5. Roupa apresentada

Com relação ao protótipo anterior, foi retirado o boné e fones de ouvido pequenos foram costurados à própria roupa. Também foram adotadas placas de contato maiores e de um mesmo tamanho; uma tela foi costurada nas costas da camisa para fixar o circuito evitando-se costurá-lo diretamente na roupa. Desta forma é possível retirá-lo da roupa quando for preciso lavar a camisa. O speaker se fixa na camisa por meio de velcro e 19 Ilhoses também foram colocados para ora embutir ora exhibir os fios, os circuitos e a bateria.



Figura 6. Exibição da roupa durante a apresentação do TCC para graduação em Educação musical pela Uni rio .

5. Conclusão

O presente artigo permitiu a reflexão e um debate sobre as novas estratégias de aprendizagem musical dentro do contexto da sala de aula, apoiadas pelas tecnologias computacionais mais recentes.

O levantamento de trabalhos correlatos indica que esta é uma área de interesse crescente em outros países e com diferentes linhas de pesquisa a serem exploradas. Foi também observado que ainda é encontrado um maior número de trabalhos que intencionam a manipulação e geração sonora sem pensar o aspecto educacional na sala de aula sendo essa também uma área promissora de pesquisa.

A associação de gestos com diferentes tipos de entrada de dados e o próprio fazer musical está se desenvolvendo. As atividades propostas pelo professor devem estar amparadas por teorias educacionais bem fundadas para que se possa oferecer um ambiente estimulante e fértil ao ensino musical. Professores e estudantes devem se conscientizar de que a tecnologia não resolverá os desafios educacionais, mas pode apoiar muito o cotidiano da sala de aula se houver real interesse dos professores em modernizar suas estratégias.

Propostas de pesquisas incluem a implementação do mesmo protótipo com tecnologia wireless para atender às necessidades de liberdade de movimento da sala de aula e de seu público alvo. Outra proposta é a realização de estudos sobre como usar um computador central para o processamento das informações e feedback de áudio, ou dividir essas tarefas de processamento entre aparelhos mobile usados pelos estudantes. Outra proposta é a melhoria dos sons gerados através de integração do microprocessador e seus dados com bancos de som, midi, síntese sonora e amostras de áudio.

Referências

- Dalcroze, E.J. Rhythm music and Education. G.P. Putnam's Sons New York 1921.
- Ferguson, S. (2006) Learning Musical Instrument Skills Through Interactive Sonification. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06), Paris, France.

- Gelineck, S., Böttcher, N. (2012) 6to6Mappr An educational tool for fast and easy mapping of input devices to musical Parameters. Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound
- Gohn, D. (2007) Tecnofobia na música e na educação: origens e justificativas. Revista Opus, 13(2), 161-174.
- Hevner, A.R., Hatterjee, S. Design Research in Information Systems, Integrated Series in Information Systems, v.22, Springer, 2010.
- Horace H I., Belton Kwong.,and Ken C K Law (2005) BodyMusic: A Novel Framework Design for Body-driven Music Composition. Image Computing Group, Department of Computer Science.
- Mann,S., Janzen, R., Hobson ,Tom.(2011) Multisensor Broadband High Dynamic Range Sensing for a Highly Expressive Step-based Musical Instrument. University of Toronto. Proceedings of the fifth international conference on tangible, embedded and embodied interaction.pages 21-24.
- Swanwick, K .Teaching music musically. Mar 14, 2007 Kindle edition.
- Swanwick,K; TILLMAN, J. The Sequence of Music Development: a study of children.s composition. *British Journal of Music Education*, 3(3): 305-339, Nov, 1986.
- Volpe,G., Varni, G., Mazzarino, B., Addessi, Anna.(2012) BeSound: Embodied Reflexion for Music Education in Childhood. IDC 2012 SHORT PAPERS 12th-15th June, Bremen, Germany
- Yang,Y. (2011) A musical wearable : Integrating electronics into clothing. MIT press
- Zhang, C., Shen, Li., Wang D., Tian, F., Wang, H.(2011) CoolMag: A Tangible Interaction Tool to Customize Instruments for Children in Music Education.UbiComp'11 / Beijing, China
- Zhou, Y., Percival, G, Wang,X., Wang., Zhao, S. (2011) - MOGCLASS: Evaluation of a Collaborative System of Mobile Devices for Classroom Music Education of Young Children School of Computing (SoC), National University of Singapore – CHI-2011.

Videoaula no ensino coletivo de instrumentos musicais: um relato de experiência

Juliana Rocha de Faria Silva¹

¹Instituto Federal de Educação, Ciência e Tecnologia de Brasília (IFB)

70.860-100 – Brasília – DF – Brazil

juliana.silva@ifb.edu.br

***Abstract.** This article describes an experience report in Music Education Computer-aided. The tool is the video that contains step by step to learning a musical instrument in a song. The story tells how the proposal was held pedagogical and musical education was aimed collective musical instrument in a group with high school students. From experience we emphasize aspects of the peer pedagogy, in which the student is replaced by the role of the teacher and goes in search of its autonomy in learning and also of methodological approaches to the music educator who works in the school context.*

***Resumo.** Este artigo descreve um relato de experiência na Educação Musical auxiliada por computador. A ferramenta é o vídeo que contém o passo a passo para o aprendizado de um instrumento musical em uma música. O relato conta como foi realizada a proposta pedagógico-musical cujo objetivo foi o ensino coletivo de instrumento musical em grupo com alunos do Ensino Médio. Da experiência são ressaltados aspectos da pedagogia de parceria, na qual o aluno passa a ter o papel do professor, vai em busca da sua autonomia no aprendizado, e também dos caminhos metodológicos para o educador musical que atua no contexto escolar.*

1. Introdução

O ensino e aprendizagem da Música é um desafio para os professores que atuam na educação básica. Ensinar música musicalmente, como é proposto pelo educador musical Keith Swanwick (2003), é algo que requer do professor, seja ele músico ou não, reconhecer a música integrada ao contexto social e cultural em que a sua teoria e a prática devem responsabilizar-se por essa contextualização da música e do fazer musical. A atualidade pressupõe o avanço tecnológico com as mudanças sociais, culturais, educacionais, econômicas, políticas e as relações entre seres vivos e computadores (SANTAELLA, 2003). Em se tratando dos contextos educacionais, há a necessidade de integrá-los às mídias atuais que abrem espaço a novas possibilidades pedagógicas. Para isso, tornar a aprendizagem e a atuação dos alunos mais interessantes em contextos educacionais, deve-se favorecer um trabalho baseado nas ideias que eles desenvolvem no e com o uso dos recursos tecnológicos (MARTINS, 2003).

Este artigo é um relato de experiência de aulas de Música no Ensino Médio em uma instituição de ensino técnico e tecnológico. Esse relato baseou-se em um plano de

ensino que tem como objetivo principal compreender as características musicais e extramusicais dos estilos/gêneros de músicas preferenciais dos alunos. O plano de curso de Música foi desenvolvido e elaborado com base nas reflexões trazidas pelos resultados de um projeto de iniciação científica finalizado em agosto de 2013. A ementa desse plano é conhecer a história de grupos musicais, bandas e cantores e compreender o contexto de produção, disseminação e recepção dos estilos/gêneros musicais. Além disso, vivenciar os elementos da música presentes nos estilos/gêneros musicais por meio da apreciação e *performance* musicais.

2. Referencial teórico

Neste trabalho, tecnologia é entendida como metodologias, competências, capacidades e conhecimentos necessários para realizar tarefas produtivas e não inclui apenas os produtos artificiais fabricados pela humanidade. Além disso, envolve a interação de diferentes agentes e processos sociais pertencendo e atuando em um meio o qual sofre influências e é modificado (VERASZTO et al, 2008).

A proposta para as aulas de Música é sustentada por dois tripés: da aprendizagem musical informal e da pedagogia de parceria (*peer pedagogy*). Para compreender melhor e diferenciar as práticas de aprendizagens informais e o ensino de música formal, a tabela 1 destaca cinco características de aprendizagem que fazem parte do cotidiano dos músicos populares que também são percebidas nos alunos do Ensino Médio, traçando um paralelo que os distinguem daquelas práticas que são encontradas na educação musical formal que são exercidas nos contextos escolares:

Tabela 1. Quadro comparativo das práticas informais e ensino formal de música

Práticas informais de aprendizagem musical de músicos populares	Os processos de aprendizagem da Educação Musical formal
Aprender a música que é escolhida de acordo com a preferência pessoal, que é familiar e com a qual apreciam e se identificam fortemente;	Introduzir o aluno em músicas que frequentemente são novas e desconhecidas e normalmente escolhidas pelo professor;
Aprender ouvindo gravações e copiando as músicas de ouvido;	Aprender por meio de notação musical ou de outros comandos escritos ou verbais;
Aprender ao lado de amigos por meio de conversas sobre música; estímulo dos pares/ colegas; ouvindo, assistindo ou imitando o outro e geralmente sem a supervisão de um adulto;	Aprender por meio da instrução de um especialista ou a supervisão de um adulto;
Assimilar habilidades e competências pessoais muitas vezes de maneira casual ou de acordo com as preferências musicais começando com peças de músicas de todo o “mundo real”;	Aprender a sequência de uma progressão de processos do mais simples ao mais complexo, envolvendo especialmente a execução de músicas de outros compositores, um currículo ou um programa gradual (de níveis iniciais, intermediários e avançados);
Manter uma estreita integração entre ouvir, tocar, improvisar e compor em todo o seu processo de aprendizagem.	Especializar gradualmente nas habilidades de ouvir, tocar, improvisar e compor; muitas vezes tendendo a enfatizar a reprodução mais que as habilidades criativas.

Fonte: GREEN; WALMSLEY, 2006 (tradução nossa).

As mudanças trazidas pelas tecnologias na escola geraram transformações no processo de ensino e aprendizagem na educação formal. Introduzir novas tecnologias na

escola não melhora o aprendizado automaticamente, porque a tecnologia dá suporte à pedagogia e não o contrário. A tecnologia modifica os papéis de professores e alunos e não dá apenas suporte às aulas expositivas. Os alunos começam a ensinar a si mesmos com a orientação dos professores. Antes, porém, se limitavam apenas a ouvir e fazer anotações. O papel do aluno passa a ser de pesquisador, de usuário especializado em tecnologia. O professor passa a ter papel de guia e de treinador. Ele estabelece metas para os alunos e os questiona, garantindo o rigor e a qualidade da produção da classe e abre mão do papel de controlador para assumir o de guia dos alunos. Isso é o que pode-se chamar de pedagogia de parceria (PRENSKY, 2001; 2008).

3. Material e Métodos

Neste artigo relata-se a experiência do ensino e aprendizagem de instrumentos musicais por meio de vídeos. A escolha do repertório se deu pela divisão dos grupos de alunos e dos seus interesses na pesquisa de algum estilo/gênero musical. O repertório também teve como ponto de partida a análise das práticas musicais informais percebidas nas primeiras aulas por meio do diálogo e da observação. Como havia muitos alunos que não tinham qualquer contato com um instrumento musical, buscou-se vídeos com versões simplificadas dos arranjos das músicas. Foram então escolhidas as videoaulas das seguintes músicas: 1) *Beat it*, Michael Jackson, aula de bateria; 2) *Twist and shout*, Phil Medley e Bert Russell, versão do Beatles aula de teclado; e 3) *Te esperando*, Luan Santana, aula de violão.

Todos os vídeos foram escolhidos da página de videoaulas do Cifra Club por apresentarem *riffs* fáceis de executar por alguém que ainda não experimentou tocar um instrumento. Uma canção é construída com múltiplos *riffs* curtos muitos dos quais são fáceis de lembrar e tocar dependendo em qual instrumento é tocado (GREEN, 2008). A metodologia utilizada contemplou três aspectos: i) visualização da videoaula a partir de um projetor multimídia; ii) execução nos instrumentos musicais desde a primeira vez, mas na segunda, o vídeo era parado e reprisado quantas vezes fossem necessárias; iii) interferência do professor depois que os alunos tivessem a primeira experiência de tocar o *riff*. Nenhum dos vídeos apresenta ao aluno a partitura com os símbolos notacionais tradicionais, o método é baseado no ouvir e na visualização da *performance* do professor. Em todas as músicas, a complexidade da escrita dos ritmos seria um fator limitante para o aprendizado das músicas se partisse da leitura da notação convencional.

4. Resultados

Deste relato foi possível perceber a aprendizagem que ocorre com os pares. Muitos alunos possuem conhecimentos em música provenientes das práticas informais de aprendizagem que têm origem na enculturação musical, ou seja, aprendem em casa com algum parente ou amigo. Em geral, esse aprendizado é através do ouvido, da visão e da imitação (GREEN, 2001). Elegem como modelo o aluno “tocador” que se dispõe a demonstrar como executar, corrige erros e sugere caminhos para a execução do instrumento musical. Esse aluno geralmente é autodidata que aprende sozinho ou com o auxílio de materiais disponíveis, principalmente, na Internet. A aprendizagem por meio de vídeos do *YouTube* é um fenômeno que cresce e ganha popularidade entre os adolescentes e jovens do Ensino Médio. Ressalta-se também o encorajamento oferecido por aqueles que relatam suas tentativas, com erros e acertos, de aprendizagem do

instrumento musical dentro ou fora da aula. Os alunos veem no depoimento do outro aquilo que podem realizar e se sentem capazes para experimentar.

O que foi observado a cada aula foi o desenvolvimento dos alunos nas suas *performances* musicais. A cada semana, mesmo não tendo o instrumento musical, nota-se o rapidez da memorização dos acordes, das posições no teclado e no violão e do desenvolvimento da coordenação motora. Alunos comentaram que procuravam os alunos que já tocavam há mais tempo para lhes ensinar a música trabalhada na sala de aula. Além disso, outros alunos levaram para casa, os vídeos gravados em seu *pendrive* ou no celular no intuito de “tirarem” a música.

Discussão e contribuições

Dois aspectos podem ser discutidos a partir deste relato: i) os papéis do aluno e do professor nesta era da informação gerada pela Internet; e ii) a formação inicial e continuada do professor de música. No primeiro aspecto destaca-se a inversão de papéis, o professor que antes era o detentor de todo o conhecimento, torna-se aquele que gerencia o ensino e aprendizagem podendo em algum momento atuar como o mestre e em outro, como aprendiz. Uma das grandes preocupações geradas pela inversão dos papéis é o medo do professor em perder a autoridade na sala de aula. Alguns desses questionamentos são discutidos por Prensky (2011) ao abordar o ensino na era das tecnologias. Saber quando é a hora de “dar a voz” ao aluno e retomá-la.

O segundo aspecto que este relato contribui é na reflexão da formação do professor de música. Esse professor precisa tocar todos os instrumentos musicais? Será que os recursos tecnológicos podem ser suficientes para suprir as lacunas da sua formação musical? Alguns questionamentos são mais amplos e requerem outras pesquisas, no entanto, o que se pode afirmar a partir da experiência deste relato, é que por meio de videoaulas da Internet é possível planejar aulas para o aprendizado coletivo de instrumento musical mesmo o professor não tendo o domínio de todos os instrumentos.

Referências

- GREEN, Lucy, and WALMSLEY, Abigail (2006) “Classroom Resources for Informal Music Learning at Key Stage 3”. www.musicalfutures.org/PractionersResources.html.
- GREEN, Lucy (2001) “How popular musicians learn”. England, Aschgate.
- PRENSKY, M. (2008) “The role of technology in teaching and the classroom”. In: Educational Technology, v. 48, n. 1. www.marcprensky.com/writing/Prensky-The_Role_of_Technology-ET-11-12-08.pdf
- _____ (2001) “Digital Natives, Digital Immigrants part 1”. In: On the Horizon. NCB University Press, n. 5, v. 9, p. 1-6.
- SANTAELLA, Lucia (2003) “Culturas e artes do pós-humano: da cultura das mídias à cibercultura. São Paulo, Paulus.
- SWANWICK, Keith. (2003) “Ensinando Música Musicalmente”. Tradução de Alda Oliveira e Cristina Tourinho. São Paulo: Moderna.
- VERASZTO, E. et al. (2008) “Tecnologia: buscando uma definição para o conceito”. Revista Prisma.com. Ed. n. 7, p. 60-85.

Estudo interdisciplinar de obras para flauta solo

Marcos S. Sampaio¹, Guilherme Bertissolo¹, Lucas Robatto¹,
Alisson G. Silva¹, José Rodriguez¹

¹Grupo de Pesquisa Genos – Escola de Música da Universidade Federal da Bahia
Av. Araújo Pinho, 58 – 40110-913 Salvador, BA

marcos@sampaio.me, guilhermibertissolo@gmail.com, robattolucas@gmail.com

Abstract. *Accordingly to Volks et al., the Computational Musicology requires an interdisciplinary effort to reach its full potential. This paper describes an interdisciplinary research on a collection of solo flute pieces available at the International Music Score Library Project (IMSLP), the connections and challenges among the areas involved, the current development status of the MusiAnalysis—the computer-assisted analysis system—, the research methodology, and the expected results.*

Resumo. *De acordo com Volks et al., a Musicologia Computacional requer um esforço interdisciplinar para alcançar todo o seu potencial. Este artigo descreve uma pesquisa interdisciplinar sobre uma coleção de obras para flauta solo disponível no International Music Score Library Project (IMSLP), as conexões e desafios entre as áreas envolvidas, o estado atual do desenvolvimento do MusiAnalysis—o sistema de análise assistida por computador—, a metodologia da pesquisa e os resultados esperados.*

1. Introdução

O presente artigo apresenta o estado atual da pesquisa interdisciplinar realizada pelo Grupo de Pesquisa Genos sobre o repertório de cerca de 200 obras para flauta solo compostas entre 1710 e 1810 e disponíveis no *International Music Score Library Project (IMSLP)*¹. O objetivo desta pesquisa² é delimitar um corpus de obras para flauta solo compostas no período mencionado, definir fontes aceitáveis para transcrição, analisar estatisticamente características musicais dessas obras com o auxílio do computador, e usar os resultados destas análises para criar ferramentas composicionais de interação em tempo real.

O potencial da área de Musicologia Computacional ainda não foi plenamente alcançado e depende tanto de uma intensificação da cooperação entre a Musicologia e a Ciência da Computação, quanto de iniciativas da Musicologia em levantar questões que possam ser modeladas computacionalmente [Volk et al., 2011]. Este trabalho pretende contribuir com a área de Musicologia Computacional por meio de um esforço interdisciplinar realizado por uma equipe das áreas de Interpretação Musical, Musicologia Sistemática, Composição, Estatística e Computação.

Neste artigo apresentamos as conexões entre os subprojetos destas cinco áreas e seus desafios, o estado atual de desenvolvimento do *MusiAnalysis*, sistema para análise das obras com auxílio do computador, a metodologia do trabalho e os resultados esperados.

¹Disponível em <http://imslp.org/>.

²Disponível em <http://genosmus.com/pesquisa/flauta-solo/>.

2. Natureza interdisciplinar do projeto

A articulação entre os subprojetos das áreas componentes desta pesquisa ocorre mediante o desenvolvimento de ferramentas compartilhadas, como o *MusiAnalysis*, o sistema de informações e análise de obras. Nas subáreas musicais, o estabelecimento de um tema geral comum—repertório para flauta solo entre 1710 e 1810—e o desenvolvimento de ferramentas compartilhadas permitem que cada subprojeto parta dos questionamentos tradicionais que caracterizam as suas áreas de pesquisa em música.

A utilização de tais ferramentas permite a elaboração de questões advindas da ampliação do escopo dos dados analisados. Por exemplo, a área da interpretação musical trata frequentemente do estabelecimento de traços identificáveis em um determinado repertório, classificando-os enquanto padrão estilístico ou desvio deste (por exemplo, a utilização de ritmos pontuados em música de caráter francês). Esta classificação possibilita ao intérprete realçar determinadas características de uma obra, conferindo-lhes então significado estético (realçar os ritmos pontuados em obra de compositor alemão, por ex.). Tradicionalmente, esta definição de padrões é baseada na experiência empírica acumulada por indivíduos, que, a partir do seu conhecimento do repertório específico, avaliam determinados traços, classificando-os. A musicologia estatística permite que uma grande quantidade de dados seja analisada em detalhe, destacando traços específicos, avaliando-os com rigor estatístico, e com isto estabelecendo mais claramente padrões e seus desvios (estabelecendo outros traços característicos à maioria do repertório francês, antes não percebidos enquanto tais, por exemplo). Estas questões acabam ultrapassando os limites tradicionais das pesquisas em cada área específica.

Estes exemplos demonstram as inter-relações entre a interpretação musical e a musicologia sistemática. Contudo, estas mesmas questões são extrapoladas ao serem compartilhadas com a área de composição, que pode partir da delimitação de traços estilísticos, e de características técnico instrumentais (assim como de padrões estéticos ou seus desvios) para criar novas situações musicais, em contextos estéticos bastante diversos (como a aplicação de padrões estatísticos de resposta a contornos melódicos, estruturas de frase, e dados significativos retornados na análise estatística).

Na área da Estatística, este estudo permite a verificação de quais métodos são mais adequados à otimização dos dados e ao estabelecimento de hipóteses, como os padrões melódicos característicos de um determinado compositor. Na área da Computação, este estudo permite o desenvolvimento, implementação e aplicação prática de algoritmos para a busca de padrões e para o estabelecimento de similaridade entre estruturas musicais, como motivos e contornos.

Por exemplo, o ACMEMB [Marvin and Laprade, 1987] é um algoritmo de medida de similaridade de contornos ineficiente, pois consome processamento de máquina de forma exponencial ao comparar todos os subconjuntos possíveis dos dois contornos comparados³. Com este projeto é possível desenvolver, implementar um algoritmo alternativo e aplicá-lo à análise de um grande corpus de composições. Portanto, este projeto estimula a busca por soluções para este tipo de problema computacional.

3. O sistema *MusiAnalysis*



A principal ferramenta em desenvolvimento neste projeto é o *MusiAnalysis*, um sistema de armazenamento e processamento das informações musicais e musicológicas do repertório trabalhado. Este sistema aproveita as capacidades do Python e do conjunto de

³Para mais informações sobre os problemas da Teoria de Relações de Contornos Musicais, consultar [Sampaio, 2012].

ferramentas para musicologia assistida por computador, Music21⁴. O Music21 dispõe de funcionalidades avançadas úteis para analisar, transformar e visualizar dados musicais.

O *MusiAnalysis* está sendo construído com um módulo musicológico e outro estrutural. Ambos os módulos dispõem de classes de objetos com atributos e métodos para armazenagem e processamento de informações. O módulo musicológico tem classes de objetos para processamento e registro de informações sobre os compositores, composições, editores e sobre as fontes documentais; e o módulo estrutural contém classes para processamento e registro do conteúdo musical, com informações sobre as obras, como intervalos, contorno melódico, duração das notas, padrões rítmicos dos tempos métricos, âmbito e registro. O módulo estrutural está sendo construído de modo a permitir a aplicação de algoritmos como os mencionados na seção 2.

As fontes musicais em formato PDF disponíveis no IMSLP são codificadas em formato XML com softwares como o Finale e parseadas com o Music21. O Music21 organiza as estruturas musicais como compassos, notas e alturas em classes e subclasses de objetos e fornece métodos como plotagem de contornos e visualização da partitura. O *MusiAnalysis* extrai destas classes as informações necessárias às questões musicológicas formuladas pela equipe do projeto e as guarda em um banco de dados. Conforme a seção 2, a formulação dessas questões ocorrem em um processo simbiótico com o próprio levantamento das informações. Finalmente, estas questões estimulam a implementação de scripts para o levantamento de dados e formulação de hipóteses.

Em caráter ilustrativo, o script abaixo retorna os dois padrões rítmicos de tempo mais recorrentes no segundo movimento da primeira fantasia de Kuhlau (Op. 38). A função `makeMovement` parseia o arquivo xml de nome `IT32089_01b`⁵ e gera o objeto `Movement`. Durante este processo, diversas informações musicais, como os padrões rítmicos, são calculados e armazenadas como atributos. O método `countBeatRhythmicPattern` conta os padrões rítmicos usando a classe `Counter`, do Python, e finalmente, o método `most_common` retorna os ritmos mais comuns. Neste caso, os padrões rítmicos de tempo mais comuns são  e .

```
movement = makeMovement('IT32089_01b')
countObject = movement.countBeatRhythmicPattern()
print countObject.most_common(2)
```

4. Metodologia

Esta pesquisa está sendo realizada em dois ramos interconectados. O primeiro compreende o desenvolvimento do *MusiAnalysis* paralelamente à definição e transcrição manual das fontes documentais das obras. Estas transcrições (em formato XML) servem como dados de entrada para testar o software, bem como obter dados musicológicos que levam aos questionamentos simbióticos mencionados na seção 3. Este processo resulta em hipóteses e ideias para a aplicação de métodos estatísticos e desenvolvimento de algoritmos.

O segundo ramo compreende o desenvolvimento de ferramentas para a composição de obras para flauta solo e eletrônica em tempo real. Estas ferramentas estão sendo construídas com o software PureData com implementação de cadeias de Markov e redes neurais artificiais para processamento dos dados gerados pelo *MusiAnalysis*, como intervalos, contorno melódico, duração das notas, âmbito e registro.

⁴Disponível em <http://mit.edu/music21>.

⁵Este código, elaborado pela equipe do projeto, se refere à transcrição (T) da fonte 32089 do IMSLP, música 01, segundo movimento (b).

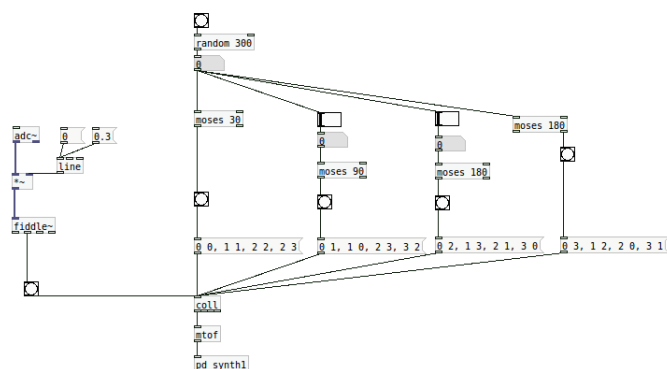


Figura 1: Cadeia de Markov no PureData

As cadeias de Markov processam os dados gerados pelo *MusiAnalysis* a partir da interação entre os objetos *Moses* e *Random*, controlando coleções de listas acionadas pelo objeto *Coll*. Por exemplo, a figura 1 contém uma cadeia de Markov que seleciona contornos melódicos de quatro notas, a partir dos percentuais de 10%, 20%, 30% e 40%, respectivamente. Esses percentuais de aparição de cada contorno são operados pelos dados gerados pelo *MusiAnalysis*.

As redes neurais possibilitam a criação de ferramentas de inteligência artificial para a composição musical. A abordagem prévia do uso de redes neurais em obras para eletrônica em tempo real [Toffolo, 2010] oferece um importante referencial e ponto de partida para este aspecto da pesquisa.

As estratégias e ferramentas oriundas deste projeto podem oferecer contextos interessantes para a composição assistida por computador, pois poderão fornecer dados concretos de um escopo relevante de obras musicais.

5. Resultados esperados

Os resultados esperados neste projeto são a classificação das obras da coleção trabalhada, a formulação e verificação de hipóteses sobre características destas obras, a definição de um corpo de métodos estatísticos adequados a este estudo, a elaboração e implementação de algoritmos para busca de padrões e estabelecimento de similaridade de estruturas, a composição de obras musicais para flauta e eletrônica em tempo real, a gravação e apresentação pública de tais obras, com disponibilização eletrônica das partituras, patches e áudios, bem como das ferramentas composicionais geradas durante a pesquisa.

Referências

- Marvin, E. W. and Laprade, P. A. (1987). Relating musical contours: Extensions of a Theory for Contour. *Journal of Music Theory*, 31(2):225–267.
- Sampaio, M. d. S. (2012). *A Teoria de Relações de Contornos Musicais: inconsistências, soluções e ferramentas*. Tese de doutorado, Universidade Federal da Bahia.
- Toffolo, R. B. G. (2010). Desenvolvimento do processos composicionais eletroacústicos a partir da relação entre live-electronics e redes neurais artificiais. In *Anais do 6º Simpósio de Cognição e Artes Musicais*, Rio de Janeiro. Programa de Pós-Graduação em Música/UFRJ.
- Volk, A., Wiering, F., and Kranenburg, P. V. (2011). Unfolding the potential of computational musicology. *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations (ICISO)*.

***Quattro Pezzi per Orchestra* de Scelsi sob a ótica da Análise Musical Assistida por Computador**

Ivan Eiji Simurra¹⁻², Igor Leão Maia², Jônatas Manzolli¹⁻²

¹Núcleo Interdisciplinar de Comunicação Sonora – NICS/UNICAMP
Rua da Reitoria, 165 – Cidade Universitária “Zeferino Vaz” – CEP: 13 083 – 872 –
Campinas – São Paulo - Brasil

²Instituto de Artes – UNICAMP
Rua Elis Regina, 50 – Cidade Universitária “Zeferino Vaz” – CEP: 13 083 – 854 –
Campinas – São Paulo – Brasil

{ieysimurra,igorleaomaia,jotamanzo}@gmail.com

Abstract. *We present a collaborative work in music analysis between music score and computational support with the use of audio descriptors. This methodology was used to analyze the first movement of Giacinto Scelsi's Quattro Pezzi per Orchestra. We aim to identify and to evaluate the brightness level by the relevance of the choice of musical instruments and music dynamics.*

Resumo. *Apresentamos o trabalho colaborativo de análise musical entre a partitura e o suporte computacional, com a utilização de descritores de áudio. Esta metodologia foi utilizada para analisar o primeiro movimento das Quattro Pezzi per Orchestra, de Giacinto Scelsi. Nosso objetivo foi identificar e avaliar o índice de percepção de “brilho sonoro” pela relevância da escolha de determinados instrumentos musicais e intensidades sonoras.*

1. Introdução

O trabalho de análise realizado diretamente pela partitura musical disponibiliza resultados consideráveis por ressaltar aspectos estéticos e conceituais do planejamento particular do compositor. Contudo, muitas vezes, a análise realizada somente pela partitura pode não ser suficiente para a compreensão macro e microestrutural do universo timbrístico. Uma das possibilidades de expansão das ferramentas de análise musical são os modelos e os suportes computacionais os quais analisam e descrevem as características do comportamento do espectro de frequência de um sinal sonoro. Essas ferramentas são denominadas como descritores de áudio.

O compositor italiano Giacinto Scelsi desenvolveu seu projeto composicional com o objetivo de extrapolar as particularidades da notação simbólica, de uma partitura, face a complexidade timbrística de uma única nota musical. Scelsi explora as mudanças de instrumento ou mesmo os diversos modos de ataque e técnicas de execução instrumental de forma a ressaltarem diferentes aspectos do comportamento do espectro de frequência de uma única nota musical. A sua obra *Quattro Pezzi per Orchestra* desenvolve-se, predominantemente, por intermédio deste projeto composicional. A quantidade de instrumentos de metais e a intensidade sonora, resultando em sonoridades mais “brilhantes”, estabelece pilares estruturantes da forma musical do primeiro movimento das *Quattro Pezzi per Orchestra*. Para enfatizar ainda mais esses momentos,

o compositor altera a percepção do timbre desses pilares ao empregar técnicas instrumentais específicas de cada instrumento, como *tremolo* nas cordas ou com a utilização de *sordinas*, nos metais. A motivação principal da análise do primeiro movimento das *Quattro Pezzi per Orchestra* é realizar um trabalho colaborativo entre a partitura e o suporte computacional, com o uso dos descritores de áudio, para identificar aspectos os quais manipulam a percepção de brilho timbrístico.

As próximas duas seções deste artigo apresentam, respectivamente, os descritores de áudio utilizados para a realização das análises musicais (seção 2) e a metodologia do trabalho de pesquisa (seção 3). As seções subsequentes descrevem os resultados obtidos (seção 4); as análises e considerações a partir dos resultados (seção 5); as discussões e conclusões (seção 6) e os futuros trabalhos (seção 7).

2. Descritores de Áudio

Os recentes métodos de análise e composição com suporte computacional, possibilitam a extração de vários parâmetros ou curvas os quais descrevem características sonoras a partir do espectro de frequências. Muitos desses modelos computacionais focam a “recuperação da informação musical” ou simplesmente MIR (“*Music Information Retrieval*”) [BYRD; FINGERHUT, 2009]. Utilizou-se dois descritores os quais relacionam-se com a região do espectro de frequência na qual concentra-se a maior quantidade de magnitude, denominado centroide espectral [Theodoridis & Koutrumbas, 2009]. Utilizou-se, ainda, uma medida psicoacústica relativa à percepção de intensidade sonora, denominada loudness, como definida em Thiede (1999).

3. Metodologia

A primeira etapa do processo de análise se concentrou em segmentar a gravação sonora do primeiro movimento das *Quattro Pezzi per Orchestra* em curtos e distintos arquivos de áudio. A gravação utilizada foi a comercialmente disponibilizada pela gravadora ACCORD¹ e possui os seguintes parâmetros: 16 Bits, 44.110 Hz. Para realizar a segmentação sonora, utilizou-se o software Audacity². Como suporte para a segmentação do áudio, utilizou-se a partitura original da obra³. Desta forma, foi possível parear os elementos simbólicos encontrados na partitura com os segmentos de áudio. Uma análise inicial da partitura demonstra que há distintos agrupamentos de alturas, articulações e inflexões microtonais. A demarcação desses pontos guiou a segmentação sonora como por exemplo: o “ataque” ou a entrada de um instrumento musical específico, o comportamento dinâmico das intensidades musicais e os diversos modos de articulação e técnicas instrumentais. Cada segmento sonoro do áudio foi relacionado aos parâmetros e as alterações supracitadas. No total, definiu-se 36 segmentos apresentados a seguir:

segmentação	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
partitura	cps.: 01 - 02	Cps.: 2 - 5	Cps.: 5 - 7	Cps.: 7 - 8	Cps.: 8 - 10	Cps.: 10 - 11	Cps.: 11 - 12	Cps.: 12 - 12	Cps.: 12 - 12	Cps.: 12 - 12	Cps.: 12 - 13	Cps.: 13 - 13	Cps.: 13 - 13	Cps.: 13 - 13	Cps.: 13 - 14	Cps.: 14 - 17	Cps.: 17 - 18	Cps.: 18 - 19
segmentação	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
partitura	Cps.: 19 - 19	Cps.: 19 - 19	Cps.: 20 - 20	Cps.: 20 - 20	Cps.: 21 - 21	Cps.: 21 - 22	Cps.: 22 - 25	Cps.: 25 - 27	Cps.: 27 - 28	Cps.: 28 - 30	Cps.: 30 - 31	Cps.: 31 - 31	Cps.: 31 - 32	Cps.: 32 - 33	Cps.: 34 - 35	Cps.: 36 - 37	Cps.: 37 - 38	Cps.: 38 - 39

Figura 1: Tabela com os trinta e seis segmentos sonoros.

¹ Para mais informações, ver <http://www.discogs.com/Giacinto-Scelsi-Quattro-Pezzi-Per-Orchestra-Anahit-Uaxuctum/release/997365>. Data de Acesso: 18/08/2013

² AUDACITY – Disponível em: <http://audacity.sourceforge.net/>. Data de Acesso: 18/08/2013.

³ SCELSEI, G. – *Quattro Pezzi per Orchestra* – Editions Salabert, Paris.

Definida a segmentação, considerou-se a análise a partir da partitura, concentrando-se no efetivo instrumental e nas dinâmicas musicais simbólicas, em cada segmento sonoro. A etapa conseguinte concentrou-se em extrair as informações estatísticas adquiridas por intermédio dos descritores de áudio. Utilizou-se o software Sonic Visualizer⁴ para a extração das informações de centroide espectral. Para a extração das informações em Loudness, utilizou-se a biblioteca de funções PDescriptors⁵, dentro do ambiente virtual PureData (PD)⁶. O nosso objetivo é verificar se há elementos de similaridade, continuidade ou contrastes sonoros entre os segmentos sonoros, a partir das relações entre as análises realizadas pela partitura e pela gravação sonora.

4. Análise dos Resultados

A Figura, a seguir, apresenta a quantidade de instrumentos musicais, eixo “y”, de cada naipe orquestral (representados pelas seguintes marcações: “losango” para madeiras; “quadrado” para metais; “triângulo” para a percussão e; “círculo” para as cordas):

	01 - 02	2 - 5	5 - 7	7 - 8	8 - 10	10 - 11	11 - 12	12 - 12	12 - 12i	12 - 12ii	12 - 13	13 - 13i	13 - 13ii	13 - 14	14 - 17	17 - 18	18 - 19	
Madeiras	1	2	5	2	4	3	2	2	2	2	2	2	2	2	0	0	4	
Metais	4	4	4	2	3	3	7	1	7	1	7	1	7	1	7	2	4	
Percussão	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	
Cordas	3	3	3	3	4	4	3	3	3	3	3	3	3	3	1	2	3	
	19 - 19i	19 - 19ii	20 - 20i	20 - 20ii	21 - 21i	21 - 21ii	22 - 22i	22 - 22ii	27 - 28	28 - 30	30 - 31	31 - 31i	31 - 31ii	32 - 33	34 - 35	36 - 37	37 - 38	38 - 39
Madeiras	1	5	6	1	6	1	2	0	2	6	4	6	3	0	0	0	0	
Metais	3	3	4	4	7	6	1	2	2	8	0	8	0	1	2	2	1	
Percussão	0	1	1	1	1	1	0	0	0	1	0	1	0	1	0	1	1	
Cordas	3	3	3	0	3	3	2	4	4	4	1	0	0	1	3	2	2	

Figura 2: Quantidade de instrumentos musicais de cada segmento sonoro.

Os próximos resultados foram coletados a partir dos descritores de áudio. A Figura, a seguir, apresenta os dados, normalizados, de centroide espectral e de Loudness da segmentação sonora:

segmentos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
centroide (Hz)	0.34785712	0.44136515	0.36047192	0.31836814	0.45767115	0.34944385	0.6268835	0.30726179	0.54086188	0.36730159	0.58313855	0.34609559	0.55876356	0.29104819	0.36750641	0.3764105	0.47268559	0.52187897
segmentos	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
centroide (Hz)	0.36097053	0.4463619	0.27384556	0.48370948	0.29113298	0.5517754	0.46391169	0.53254867	0.71804034	0.79422375	0.49778165	1	0.33460863	0.2634158	0.38694772	0.30984691	0.45319374	0.9210829
segmentos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
loudness	0.00313684	0.01190119	0.07179823	0.02563249	0.01952623	0.05840593	0.21843591	0.18668157	0.49080343	0.28365203	1	0.29216705	0.55290746	0.2809107	0.07312896	0.0010294	0.00400929	0.04318734
segmentos	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
loudness	0.19051251	0.41980759	0.67831384	0.9619547	0.37110079	0.0341872	0.00357582	0.00059206	0.20452122	0.27276019	0.46759304	0.42997902	0.02302404	0.00158393	0.00123186	0.00106637	0.00018943	4.35546-05

Figura 3: Tabelas do comportamento da Centroide Espectral (acima) e Loudness (abaixo) de cada segmento sonoro.

5. Análise

Pelos resultados obtidos a partir da metodologia desenvolvida neste trabalho de pesquisa, há uma proeminente relação dos dados extraídos a partir da gravação sonora com algumas características simbólicas, a partir da análise da partitura. É possível definir três agrupamentos distintos de segmentos sonoros. Estes agrupamentos distinguem os principais comportamentos das medidas e parâmetros utilizados nesse trabalho de análise, tanto no domínio das análise via descritores de áudio quanto nas análises simbólicas, pela partitura musical: Grupo I (segmentos 07 a 15, Cps.: 11 – 12 até Cps.: 13 – 14); Grupo II (segmentos 18 a 24, Cps.: 18 – 19 até Cps.: 21 – 22) e Grupo III (segmentos 27 a 30, Cps.: 27 – 28 até Cps.: 31 – 31).

⁴ SONIC VISUALIZER. Disponível em: www.sonicvisualiser.org. Data de acesso: 16 de agosto de 2013.

⁵ MONTEIRO, A. PDescriptors. Disponível em <https://sites.google.com/site/pdescriptors/>. Data de acesso: 16 de agosto de 2013.

⁶ PureData. Disponível em: www.puredata.org. Data de acesso: 16 de agosto de 2013.

6. Discussão e Conclusão

Apresentamos um estudo sobre o uso de descritores de áudio na análise de obras musicais, via gravação sonora, à qual possa ser relacionada também com o trabalho de análise musical a partir da sua representação simbólica, a partitura musical.

O principal aspecto desenvolvido neste trabalho concentrou-se na análise timbrística, à qual relaciona-se com a percepção do “brilho” de um som. Tomando por base toda trajetória do trabalho de análise, podemos comprovar duas características marcantes acerca da análise da percepção de brilho ou Centroide Espectral do som: a) a presença de instrumentos de metais e b) a percepção de intensidade sonora ou Loudness. Na obra de Scelsi, percebe-se momentos de alto índice de “brilho sonoro” mesmo com o âmbito restrito à uma altura musical. Os três agrupamentos mais proeminentes, a partir dos descritores utilizados, compartilham as mesmas características musicais: as dinâmicas musicais com maior intensidade e a técnica instrumental de *tremolo* ou *frullato*. Tais modos de ataque instrumental adicionam uma maior quantidade de ruído, aumentando os índices de centroide espectral. O projeto composicional de Scelsi ressalta aspectos os quais não estão evidenciados somente no âmbito da partitura. A análise, via descritores de áudio, demonstrou a complexidade informacional de uma única uma nota musical.

7. Projeções

Realizamos outras medidas com novos descritores de áudio para extrair aspectos relacionados com distribuição estatística da energia do espectro a partir de bandas de frequência relacionadas com as classes de alturas musicais. Além disso, realizamos outras análises concentrando-se em outros dados adquiridos por intermédio da partitura musical. Estes resultados serão abordados em publicações futuras.

Agradecimentos

Os autores gostariam de agradecer o apoio da FAPESP à bolsa de doutorado, processo: 2011/23972-2, à bolsa de mestrado, processo: 2011/16715-3 e ao CNPq aos projetos no. 477343/2010-4 e no. 304064/2010-6.

Referências Bibliográficas

- Byrd, D.; Fingerhut, M. (2002) “The History of ISMIR – A Short Happy Tale”. D-lib Magazine, Vol. 8, No. 11, 2002.
- Elezovic, I. (2007) “Scelsi’s Approach to The Third Dimension in “Quattro Pezzi (su Una nota Sola). Dissertation. University of Illinois.”
- Peeters, G. (2004) “A Large Set of Audio Features for Sound Description (Similarity and Classification)” in the CUIDADO Project, CUIDADO I.S.T. Project Report.
- Reish, G. N. (2006) “Una nota sola: Giacinto Scelsi and the genesis of music on a single note”. Journal of Musicological Research, 25: 149 – 186.
- Theodoridis, S.; Koutroumbas. (2009) “Pattern Recognition”. Academic press, Third Edition. 967 pgs.
- Thiede, T. (1999). “Perceptual Audio Quality Assessment using a Non-Linear Filter Bank”. Vom Fachbereich Elektrotechnik der Technischen Universität. Berlin zur Erlangung des akademischen Grades. genehmigte Dissertation.

Composição de um Tema Musical utilizando a abordagem do Lindenmayer System e da Geometria Fractal

Dirceu de F. P. Melo^{1*}

¹Departamento de Ciências Aplicadas IFBA - Campus Barbalho
Rua Emidio Santos, s/n – 40.301-015 Salvador, Ba

Abstract. *This paper describes the application of Lindenmayer System in the composition of a musical theme, according to the fractal structure of the Koch curve. For this we provide a brief theoretical discussion of fractals, L-System, and the LOGO turtle geometric approach. Finally, it demonstrated a geometric-algorithmic paradigm transposition for the musical one, and held a discussion about the musical choices and techniques used in the composition.*

Resumo. *Este trabalho descreve a aplicação do Lindenmayer System na composição de um tema musical, segundo a estrutura fractal da Curva de Koch. Para isso é realizada uma breve discussão teórica sobre fractais, Sistema -L, e a abordagem geométrica da tartaruga LOGO. Finalmente, é demonstrada uma transposição do paradigma algorítmico-geométrico para o musical, e realizada a discussão sobre as escolhas musicais e técnicas usadas na composição.*

1. O que são Fractais?

Fractais são conjuntos ou estruturas de forma extremamente fragmentada, e que preservam a mesma estrutura em todas as escalas. O termo *fractal* foi cunhado por Bernoid B. Mandelbrot, e tem raiz na palavra *fractus*, que significa quebrar, fragmentar. Um exemplo clássico de um conjunto fractal é a curva de Von Koch, que é formada da seguinte maneira: um segmento de reta, chamado de iniciador, é subdividido em três segmentos de igual tamanho, daí retira-se a parte central, e substituído por dois segmentos iguais, formando um triângulo equilátero. Repetindo as operações, sucessivamente, para cada segmento, obtém-se a curva de Koch no limite para o qual tende esta estrutura. A Figura 1 mostra a evolução da curva de Koch em quatro iterações.



Figure 1: Quatro iterações da Curva de Koch. Fonte: O autor

1.1. O Sistema L

No ano de 1968, Aristid Lindenmayer, um botânico alemão, elaborou um novo tipo de reescrita, denominado de *Lindenmayer System* (Sistema-L), ou *Sistema de Reescrita Paralela*. A reescrita é uma técnica utilizada para definir objetos complexos, e consiste na substituição sucessiva, de partes de um objeto inicial simples, usando um conjunto de regras de reescrita ou regras de produção [Lindemayer and Prusinkiewicz, 2004]. O Sistema-L é um tipo particular de Sistema Dinâmico Simbólico formado pelos seguintes elementos [Wright, 2013]:

*Supported by Instituto Federal de Educação Ciência e Tecnologia da Bahia

- Alfabeto: o alfabeto é um conjunto V finito de símbolos formais, geralmente caracteres como A, F, X, Y , etc.
- Axioma ou Iniciador: é uma cadeia W de caracteres formais de V . O conjunto formado por uma cadeias de caracteres, é também chamado de *palavra*, é denotado por V^* .
- O comprimento $|W|$ de uma palavra W é o número de símbolos da palavra. Também é possível o mapeamento de produções de a para uma palavra vazia, denotada por f , ou para o próprio a .
- Produções: uma produção ou regra de reescrita é um mapeamento de um símbolo $a \in V$ em uma palavra $W \in V^*$.

1.2. Sistema L na Abordagem Gráfica da Tartaruga

A Interpretação da tartaruga é dada basicamente deste modo: um estado da tartaruga é definido pelo conjunto (x, y, α) , onde as coordenadas cartesianas (x, y) representam a posição da tartaruga, e o ângulo α , chamado de *heading*, é interpretado como a direção para qual a tartaruga está voltada. Dado o tamanho do passo d e o incremento angular δ , a tartaruga pode responder a comandos representados pelos símbolos da Tabela 1:

Table 1: Comandos para definir o estado da tartaruga

F	Dê um passo à frente de tamanho d . O estado da tartaruga muda para (x', y', δ) , onde $x' = x + d.\cos\delta$ e $y' = y + d.\sen\delta$.
f	Dê um passo à frente de tamanho d sem traçar uma linha.
-	Vire um ângulo δ à esquerda. O próximo estado da tartaruga é $(x, y, \alpha + \delta)$.
+	Vire um ângulo δ à direita. O próximo estado da tartaruga é $(x, y, \alpha - \delta)$.
[Empilha o estado atual da tartaruga.
]	Desempilha e torna atual a última informação empilhada.

2. Aplicação do Sistema-L na Composição Musical

Para realizar a composição do tema musical foi necessário pensar na tranposição do paradigma visual para o musical. Uma vez que a curva de Koch tem como iniciador ou axioma, um segmento de reta que, a medida que vai sendo iterado, vai imergindo no espaço, foi definido que, para o universo musical o tamanho do iniciador deve ser o tamanho de um compasso completo. Assim, uma linha da partitura assinalada por um tom musical, com duração de um compasso, equivale ao segmento iniciador. Utilizou-se um sistema onde cada 15^0 de deslocamento, corresponde a um salto de $1/2$ tom, assim em 12 semitons de uma oitava, na escala cromática, tem-se o deslocamento total de 180^0 . Nessa construção girar um ângulo $\delta = 60^0$ à esquerda, a partir de uma determinada posição, equivale a subir quatro semitons (ou dois tons), a partir de uma determinada nota. O ângulo de partida α , nesse caso é 0^0 e, a sua interpretação musical é a nota de partida n_0 =sol 3, escolhida a critério do compositor. As tabelas 2 e 3 mostram os comandos associados ao movimento da tartaruga na esfera gráfica e na esfera musical.

Table 2: Comandos para definir o estado da tartaruga na construção da curva de Koch

F	Dê um passo à frente de tamanho d
-	Vire um ângulo δ à esquerda
+	Vire um ângulo δ à direita

A curva de Koch utiliza a seguinte regra de reescrita:

Table 3: Comandos para definir a execução no Sistema Musical

T	Execute a nota n_i de duração d
-	Desça o Intervalo I.
+	Suba o Intervalo I

$$w : F$$

$$p : F \mapsto F - F + +F - F$$

$$\alpha = 0^0 \text{ e } \delta = 60^0$$

Fazendo a equivalencia com o sistema musical, a regra fica assim:

$$w : T$$

$$p : T \mapsto T + T - -T + T$$

$$n_0 = \text{sol3} \text{ e } I = 2\text{tons}$$

O tema é construído em três iterações pois, nessa composição, houve a escolha de se trabalhar com uma unidade mínima de semicilcheia. A Figura 2 mostra a evolução do iniciador até a terceira iteração, associando ao desenho da curva de Koch em da nível.



Figure 2: Evolução do tema musical relativo à curva de Koch em três iterações.
Fonte: O autor

3. Conclusão

A técnica de reescrita L-System proporcionou uma maneira de exprimir musicalmente as características de auto-similaridade da geometria fractal, que auxiliaram na construção do tema principal da música *Fractus de Koch*. Um aspecto muito importante da utilização dessa técnica portanto, é que a realização da composição depende fundamentalmente da interferência do compositor, pois existem vários fatores que dependem da escolha criativa. Por fim, com a utilização dessa técnica, abre-se a possibilidade da para composição de temas musicais a partir de outras estruturas fractais.

References

- Lindemayer, A. and Prusinkiewicz, P. (2004). *The Algorithmic Beauty of Plants*. Springer.
- Wright, D. J. (2013). Dynamical systems and fractals lecture notes. <http://www.math.okstate.edu/mathdept/dynamics/lecnotes/lecnotes.html>.

Fractus de Koch

$\bullet = 100$ Dirceu Melo

1 *ff* *mp* *mf*

4 *ff* *ff*

6 *mp* *mf* *ff*

8 *mp*

10 *ff*

Figure 3: Fractus de Koch pg.1

10

mf *mp*

11

mp *mf* *mp*

12

mf *mp*

13

mf *mp*

14

mf *mp*

15

mp *p* *rit.* *rit.* *p*

Figure 4: Fractus de Koch pg.2

Using drawings and a guitar pedal to create one-man-band music: the Illusio

Jerônimo Barbosa¹, Filipe Calegario¹, Veronica Teichrieb¹, Geber Ramalho¹,
Giordano Cabral²

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
50.740-560 – Recife – PE – Brazil

²DEINFO – Universidade Federal Rural de Pernambuco (UFRPE)
52.171-900 – Recife – PE – Brazil

{jbcj,fcac,vt,glr}@cin.ufpe.br, giordanorec@gmail.com

Abstract. *This paper presents a new digital musical instrument, the Illusio, based on an augmented multi-touch interface that combines a traditional multi-touch surface and a device similar to a guitar pedal. Illusio allows users to perform by drawing and by associating the sketches with live loops. These loops are manipulated based on a concept called hierarchical live looping, which extends traditional live looping through the use of a musical tree, in which any music operation applied to a node affects all its children nodes.*

1. Introduction

A digital musical instrument (DMI) “consists of a control surface or gestural controller, which drives the musical parameter of a sound synthesizer in real time” (Miranda & Wanderley 2006). They can be separated in three parts: (a) the input module - that transforms a value from nature into a computable number; (b) the output module – responsible for the sound synthesis; and (c) the mapping module – a set of strategies responsible for mapping the input into the output module.

Although the DMI may have a simple model, analyzing its user experience (UX) may be a complex task (O’Modhrain 2011). Attempting to refine this UX analysis, experiments by designing, evaluating and analyzing DMIs were performed based on potential user’s feedback (Barbosa et al. 2011; Barbosa et al. 2012). This led us to consider aspects beyond the traditional input, output and mapping ones, such as: (d) its body – e.g. ergonomics, appearance and portability of the device’s physical body; (e) adhesion – any reason that would lead someone to be interested in using this instrument, like price, references of virtuosi players; (f) and user profile – eg. how much suitable is the instrument to the musical style it plays.

Traditional aspects have been already explored: new interface technologies are opening a wide range of possibilities to new DMIs inputs (Miranda & Wanderley 2006), the great diversity of technologies related to sound synthesis (Roads 1996) brings valuable opportunities to be applied to DMIs outputs and tools are concerned with turning the mapping task more efficient. However, new DMIs that are focused on exploring aspects like body and adhesion are far insufficient and, if we want to build more engaging instruments, all these UX aspects should be considered in the design

process. Incorporating this idea to our design process and aiming to explore both body and adhesion aspects, we set the following guidelines to the creation of a new DMI:

- **Playfulness & Visual aesthetics** – related to the adhesion aspect, the DMI should resemble a playful environment where the user is encouraged to freely explore musical ideas. It should be visually expressive in a live performance, aiming to impact the audience;
- **Flexibility** – related to the body aspect, the DMI should be easily customizable by users concerning their context and intention.

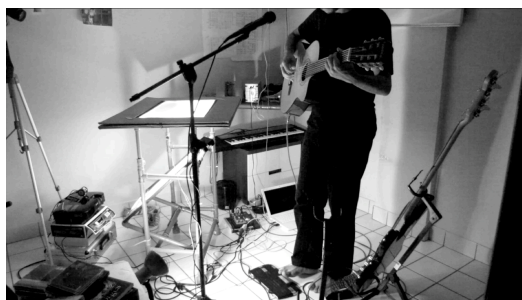


Figure 1. User performing with the Illusio

2. Description

As a result, we propose the Illusio¹ (Figure 1), a DMI which is based on new interaction concepts – the sketch, the mockup and the navigation – that allow users to draw their own graphical interface (GUI) by sketching on an empty surface and by relating these sketches with real-time recorded loops during the performance (the mapping module). For that, it uses an augmented multi-touch interface as input module (that mixes a multi-touch surface with a guitar pedal), and the hierarchical live looping concept (Berthaut et al. 2010) as the output module. Each one of its interaction concepts is further described in the following subsections.

2.1. Sketches

A sketch is a drawing made by the user and it is responsible for storing live loops. Users may create sketches by pressing and moving their fingers against the multi-touch surface as shown in Figure 2.

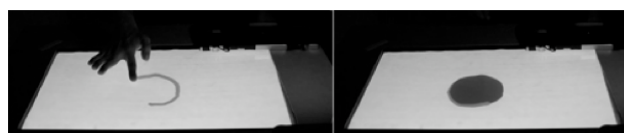


Figure 2. User drawing a sketch

A sketch is considered finished when the shape is closed. Thereafter, it can be selected by touching the surface. In this case, they can also be freely moved by dragging the finger. After created, the sketch is ready to be associated with any loop. This can be achieved through its selection and by recording a new loop using the pedal. Sketches that are associated with loops have a timer inside them, so the user can distinguish from sketches not yet associated with loops.

¹ Video demo available in: <http://vimeo.com/25641586>

When this association is completed, it is possible to play or stop loops by using the pedal – that also allows users to delete any sketch at any time. While a spinning line in the center of the sketch is used to represent playing loops, a point is used to represent stopped loops, as shown in Figure 4.

2.2. Mockups

Mockups are a subtype of sketch that, instead of live loops, stores other sketches (including other mockups as well). Thus, any music operation applied to a given mockup will affect all sketches stored inside it.



Figure 3. Grouping two sketches into a mockup

Mockups aim at gathering different loops into a single structure, triggering them together – by playing or stopping. Besides, it allows reducing the visual complexity of sketches. There are two ways for creating them:

- By selecting sketches already created and then by using the pedal, like shown in Figure 3;
- By opening an empty sketch – thereby, not associated with any loop – and then creating a new sketch inside of it (concept to be explained in the next section).

2.3. Navigation

Sketches and mockups are organized in a tree structure. Thus, navigation is the concept that allows users to navigate through this structure and to edit sketches in real-time.

For navigation, users should open or close objects. Both can be achieved by applying pinch gestures on the surface. To open a sketch, users should pinch inside the sketch area. To close a sketch, the gesture should be made where there are no sketches. When a sketch associated with a loop is open, it is possible to visualize the loop's waveform and then change it by drawing a new waveform. Besides, it is also possible to apply sound effects. When an empty sketch is open (with no loop associated) a new stage is shown. There, it is possible to create new sketches, which will automatically become children of the former sketch – that, in turn, will be transformed into a mockup.

Finally, when a mockup is open, its content (sketches and mockups inside the mockup) is displayed on the screen. These objects can once again be navigated as described above, until reaching the leaves.

2.4. Technical Description

Technically, Illusio comprises three components: the pedal, the interactive surface and the software.

The pedal component is responsible for triggering which functionality should be activated. It was built using a QWERTY USB keyboard, from which all but three keys were removed. The interactive surface component is responsible for allowing multi-touch interaction. For this, we used a Samsung Galaxy Tab 7 Plus. Both are shown in

Figure 4. The software component is the core of the instrument and can be separated in two different modules: (a) the black & white visual interface (illustrated in Figure 4), developed using Processing², responsible for handling user input gestures and represented by an empty screen containing rough sketches; and (b) the sound looper, developed using Openframeworks³/C++, responsible for dealing with the audio output. Both modules communicate via Open Sound Control (OSC) messages.

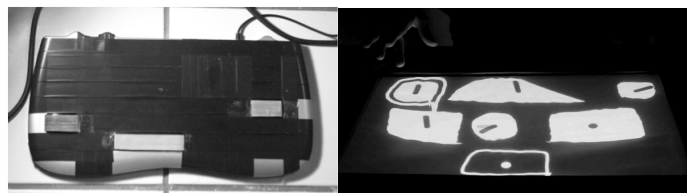


Figure 4. The pedal and the user interface

3. Conclusion

We presented here the Illusio – a new digital musical instrument that allows users to perform by drawing sketches and by associating them with live loops. As future work, we hope to improve the system by using the results achieved in a previous user test. We believe that this approach – a cycle of evaluation processes in which user feedback is constantly used to improve the system, always concerning all DMI's UX aspects – could be a promising way to build more effective and contextualized DMIs.

4. Acknowledgments

Special thanks to the Wowlabs group and the MustIC group. Illusio was developed with the partial support of Rumos Itaú Cultural Cybernetic Arts.

5. References

- Barbosa, J. et al., 2012. Considering Audience's View Towards an Evaluation Methodology for Digital Musical Instruments. In *NIME '12 Proceedings of the 2012 conference on New interfaces for musical expression*. Ann Arbor, USA.
- Barbosa, J. et al., 2011. Towards an evaluation methodology for digital music instruments considering performer's view: a case study. In *Proceedings of 13th Brazilian Symposium on Computer Music*. Vitória, Brazil.
- Berthaut, F., Desainte-Catherine, M. & Hachet, M., 2010. DRILE: an immersive environment for hierarchical live-looping. In *NIME '10 Proceedings of the 2010 conference on New interfaces for musical expression*. Sydney, Australia.
- Miranda, E.R. & Wanderley, M.M., 2006. *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, A-R Editions.
- O'Modhain, S., 2011. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1).
- Roads, C., 1996. *The Computer Music Tutorial*, MIT Press.

² <http://processing.org>

³ <http://openframeworks.cc>