U.PORTO

MESTRADO
MULTIMÉDIA - ESPECIALIZAÇÃO EM MUSICA INTERATIVA E DESIGN DE SOM

DIGIT - DIGItal sTeps

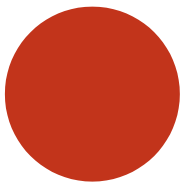
Luis Alberto Teixeira Aly

M
2016

FACULDADES PARTICIPANTES:

FACULDADE DE ENGENHARIA
FACULDADE DE BELAS ARTES
FACULDADE DE CIÊNCIAS
FACULDADE DE ECONOMIA
FACULDADE DE LETRAS

U.PORTO



Universidade do Porto

Faculdade de Engenharia

FEUP

DIGIT - DIGItal sTeps

Realização de som de passos no contexto de foley usando processos de síntese sonora concatenativa e controlo por intermédio do gesto

Luis Alberto Teixeira Aly

Mestrado em Multimédia - Música Interativa e Design de Som da Universidade do Porto

Orientador: Gilberto Bernardes (Doutor)

Coorientador: Rui Luis Nogueira Penha (Professor Doutor)

Junho de 2016

DIGIT - DIGItal sTeps

Luis Alberto Teixeira Aly

Mestrado em Multimédia - Música Interativa e Design de Som da Universidade do Porto

Aprovado em provas públicas pelo Júri:

Presidente: Doutor Matthew Edward Price Davies

Vogal Externo: Doutor Luís Gustavo Pereira Marques Martins

Orientador: Doutor Gilberto Bernardes de Almeida

Palavras-chave: síntese concatenativa; gesto; desenho de som; foley; passos

RESUMO

Com reconhecido impacto significativo, o *foley* de som acrescenta à imagem em movimento um significado sónico através da manipulação expressiva de objetos, fazendo coincidir a sua performance com a imagem em movimento. Este processo é moroso e exige muitos recursos materiais. Apesar das evoluções tecnológicas sentidas nas últimas décadas no domínio dos média no sentido da automatização dos processos de pós-produção áudio estas não ocorreram na prática do foley. Recentemente foram desenvolvidas propostas digitais para a concretização destes sons, e mais concretamente para a produção do som de passos que tendem a contrariar esta tendência. No entanto estas soluções ainda apresentam limitações, tais como: (1) a exaustão das propostas no caso de recurso a bibliotecas de som, (2) a limitação expressiva do MIDI no caso do recurso a *plugins* para a produção do som de passos e (3) alguma falta de realismo no caso do recurso ao áudio procedimental.

Nesta dissertação, de forma a dar resposta às limitações levantadas anteriormente apresento o DIGItal sTeps (DIGIT), um sistema digital interativo para a concretização do som de passos num contexto de *foley* de som. DIGIT baseia-se no conceito de perfil acústico de um gesto produzido pelo impacto da mão numa superfície captado por um microfone de contacto. Através da descrição do perfil acústico destes gestos e do seu mapeamento para parâmetros de controlo de síntese concatenativa, pretende-se a realização de uma pesquisa automática em bases de sons de passos pré-gravados, e a posterior escolha do segmento mais adequado à descrição acústica do gesto.

DIGIT potencia a expressividade e exploração do gesto como forma de recriação do som de passos em ambiente de pós-produção áudio de um filme ou jogo, dando um seguimento à prática tradicional do *foley* de som, que através da exploração táctil de objetos físicos, e reconhecimento das suas qualidades acústicas, os adapta às necessidades temáticas e narrativas da imagem em movimento.

DIGIT apresenta vantagens em relação às soluções anteriores na medida em que utiliza a descrição do perfil acústico do gesto (nas suas múltiplas dimensões) como controlo do sistema visando desta forma contornar as limitações impostas pelos controladores MIDI usados em aplicações de foley digital. DIGIT dispõe ainda de um motor de reprodução áudio granular que permite uma alteração das

características originais de um som de forma a criar mais possibilidades a partir de um pequeno número de amostras. Esta funcionalidade contribui para a diminuição do tamanho da base de som de passos pré-gravados que compõe o sistema, ao contrário das soluções como bibliotecas de som ou *plugins*, assim como potencia a diversidade de resultados.

ABSTRACT

Sound foley provides to the moving image a sonic meaning with recognized impact, through the expressive manipulation of objects in synchronisation with the image. This process is time consuming and requires a lot of physical means. Despite the great technological advances of the audiovisual industry over the last decades, towards the automation of the post-production tasks, sound foley still remains a manual task with deep roots on the traditional basis of the technique. More recently, academic and industrial research has been addressing this area and proposing some digital tools to assist the process of creating foley, and more specifically footstep sounds. Yet, these solutions have marked limitations, such as the (1) lack of diversity in the possible solutions offered by sound libraries, the (2) lack of expressive control and number of dimensions when using a MIDI keyboard as a control strategy of plugins to produce footstep sounds and the (3) lack of realism in procedural audio. In this dissertation, and in order to respond to the abovementioned limitations, I present DIGItal sTeps (DIGIT), an interactive digital system for footstep sounds in a sound foley context. DIGIT is based on the concept of the acoustic response of a gesture produced by the impact of the hand on a surface and captured by a contact microphone.

DIGIT enhances the expressiveness of gestures as a way of recreating footstep sounds in an audio post-production environment for movies or games, in line with the traditional practice of sound foley, which through tactile exploration of physical objects, and recognition of its acoustic qualities, adjusts to the narrative of the moving image.

DIGIT shows advantages over previous solutions in the use of audio descriptions of an acoustic gesture profile (in its multiple dimensions) as the main control of the system, thus expanding the number of dimensionalities of control of MIDI controllers. DIGIT also has a granular audio playback engine that enables the user to change unique features of the playback sound in order to create more possibilities from a small number of samples. This feature contributes to a decrease in the size of the sound database of pre-recorded footstep sounds that comprise the system unlike the solutions presented by sound libraries or plugins.

Agradecimentos

Gostaria de aproveitar este espaço para incluir pessoas e instituições que de uma forma mais ou menos direta contribuíram para a realização deste trabalho.

Ao meu orientador Gilberto Bernardes por me conduzir neste percurso académico novo para mim de uma forma tão próxima, com um espírito sempre construtivo, atento e com muita paciência. Obrigado Gilberto.
Ao meu co-orientador Rui Penha pela paixão que demonstra pelas novas tecnologias multimédia aplicadas ao som que se torna contagiante.

Ao Eduardo Magalhães por me ter incentivado de todas as formas possíveis a integração no Mestrado de Multimédia, e me ter aberto as portas a esta fantástica experiência académica. Aos professores, colegas de mestrado e amigos com quem tive o prazer de partilhar estes três anos de experiências. Sem qualquer ordem de importância: Alexandre Clément, Mariana Sardón, Jorge Pandeirada, Manuel Brásio, Urbano Ferreira, Javier Tomás, Rodrigo Constanzo, Cédric Camier, Dr. George Sioros, Prof. Matthew Davies, Prof. Marcelo Caetano, Prof. António Coelho, João Jacob, aos colegas do primeiro ano do Mestrado e à pessoa que nos conhece a todos pelo nome próprio Marisa Silva.

Ao Palácio do Bolhão - Escola de Artes (Glória Cheio e Pedro Carvalho), Oficina dos Violinos e Sr. Joaquim do Leroy-Merlin pelo apoio logístico.

Aos meus fabulosos pais que sempre me apoiaram nas minhas decisões ao longo de toda a minha existência. Obrigado!
À Sandra pela sua paciência, crítica, perspicácia e espírito criativo que me contaminou e alimentou ao longo do processo de elaboração deste trabalho.

Project "TEC4Growth - Pervasive Intelligence, Enhancers and Proofs of Concept with Industrial Impact/NORTE-01- 0145-FEDER-000020" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF)

Luis Aly

ÍNDICE

Lista de Figuras

Lista de Tabelas

"Ao atribuir-se significado a um ruído, este torna-se comunicação"

— Sonnenschein

1 Introdução

Esta secção contextualiza historicamente a prática do *foley* de som atendendo aos recursos materiais e humanos necessários para a sua realização. Este enquadramento é central à minha investigação de forma a apresentar as bases teóricas, os seus limites e as motivações para desenvolver o projeto aqui descrito e, mais concretamente, o sistema digital DIGIT. O capítulo termina com a apresentação dos objectivos e metodologia aplicada da minha investigação.

1.1 Contexto da investigação

A evolução e expansão do conceito de imagem em movimento de Muybridge (1855) teve como resultado inúmeras inovações tecnológicas que alteraram os média. Nos anos 1920, a inclusão de som a acompanhar imagem em movimento vem estabelecer um ramo novo dentro dos média que passa a ocupar-se pela criação de efeitos de som que é designado de *foley* (Uri & Doyle, 2013).

A técnica de foley deriva o seu nome de Jack Foley, que trabalhou como duplo e como diretor assistente para a Universal Studios nos anos 1920 (2012). Em 1927 a Warner Brothers lança *The Jazz Singer*, o primeiro filme a ter som, mais concretamente canto vocal sincronizado com imagem, a Universal Studios pressentiu que a sua nova produção, *Show Boat* também de 1927, iria ser suplantada visto se tratar de um filme mudo.

Foi então que Jack Foley e a sua equipa foram incumbidos da responsabilidade de desenhar todos os sons diegéticos[1], e.g. passos, roupa ou um bater de porta, de forma sincronizada com o filme já editado (Yewdall, 2012), técnica esta conhecida como *retroffiting*.

[1] Sons diegéticos são todos os sons que deverão estar sincronizados com a imagem para que o espectador os ouça como pertencentes ao mundo que se pretende pela narrar.

Desde então o *foley* passou a fazer parte do processo de pós-produção áudio de um filme. É a fase onde todos os sons diegéticos (Chion, 2009) são criados e gravados, numa linha de tempo sincronizada com o filme. A integridade dos sons diegéticos contribui para o que Landy (1994, p.49) designa de *something to hold on to fator*, ou seja, no contexto da música electroacústica o compositor pode lançar uma âncora ao ouvinte que lhe permita aceder ao real do que está a ouvir, i.e. ao ouvinte é-lhe dado a perceber que um determinado som que está ser processado eletronicamente pertence a um instrumento real específico. Desta forma pretende-se potenciar a compreensão do intuito da peça musical em questão.

Neste sentido o foley é um elemento essencial do design de som para o estabelecer de uma realidade física da imagem em movimento. No caso do som de passos este é o som que liga o espectador ao espaço que é retratado pela imagem, criando uma presença física do personagem.

O foley é executado por um artista de foley que torna os sons quotidianos mais proeminentes para as várias cenas contribuindo para acentuar o impacto da narrativa. O artista de foley através de práticas de escuta reduzida (Schaeffer, 1966), explora criativamente de forma táctil objetos e adereços quotidianos tentando extrair destes um significado acústico-narrativo para as imagens.

Apesar das evoluções tecnológicas que se deram no cinema o foley continua a ser um processo manual que assenta nas capacidades técnico-interpretativas do artista de foley: a imagem em movimento é projetada no ecrã, e o artista de foley mimetiza-a de uma forma sónica através de uma manipulação expressiva de objetos, de forma a fazer coincidir a sua performance com a imagem.

A arte do foley não está limitada a sons humanos. Em muitos casos a produção não dispõe de qualquer fonte de som, e.g. filme de animação ou um jogo digital. Para estas formas particulares de imagem em movimento a paisagem sonora é uma tela branca, permitindo ao artista de foley manufacturar cada evento sónico, seja um som do mundo criado pela narrativa, seja uma criação única que contribui para a diegese da história.

O foley de som tem como ponto de partida a ideia de objecto encontrado[2], (Camic & Camic, 2016) e é um processo que envolve a interação entre a estética, a cognição, a emoção, a mnemónica e a ecologia. Os fatores criativos na procura, descoberta e utilização de objetos encontrados pauta toda a atividade do foley, no sentido em que o objecto não requer especiais características de acústica musical,

[2] Objetos encontrados são objetos quotidianos que através de uma interação estética adquirem um novo significado

mas somente características acústicas relevantes que possam ser adaptadas à narrativa.

No que respeita à relação entre a prática do foley e as características acústicas dos objetos encontrados não existem estudos sistematizados que apresentem resultados objectivos sobre quais os melhores objetos para a realização de determinados efeitos de som. Trata-se de uma atividade que não dispõe de um quadro de práticas sistematizadas como na produção musical, sendo que o que existe é um repositório de experiências de vários profissionais da área (Viers, 2011).

O foley é executado num palco de foley que consiste num (1) espaço dotado de propriedades acústicas que beneficiam a escuta dos sons produzidos pelos movimentos dos artistas de foley, (2) um palco de som segmentado por áreas com vários tipos de pisos onde os artistas de foley podem recriar diferentes tipos de som de passos; e (3) um espaço contíguo onde se pode encontrar armazenado todo o tipo de materiais e adereços que estão acessíveis ao artista de foley durante a gravação e que o ajudam à eficiência e controlo na gestão de tempo e de recursos.

O próprio Jack Foley era conhecido por passar muito tempo no meio do palco de foley a seguir a ação do filme e imitar todos os movimentos e passos de um determinado ator/atriz (Yewdall, 2012). No contexto de produção de um filme o som de passos funciona como a âncora que vai ajudar o espectador a criar uma presença física e real dos personagens no ecrã.

Uri e Doyle (2013) dividem uma sessão de foley em três categorias distintas. A primeira categoria são os sons de movimentos e referem-se a todo e qualquer movimento físico feito pelos personagens no ecrã, e.g. o som originado por um casaco quando é retirado pelo personagem ou o som produzido por calças de neve enquanto se caminha. Sem estes sons o mundo criado pela narrativa iria parecer estéril e demasiado *perfeito*.

A segunda categoria é o som de passos, e, provavelmente, aqueles pelos quais o artista de foley é mais conhecido. De entre os muitos desafios da tarefa de produção do som de passos, a principal reside no facto de, mesmo que o personagem retratado no ecrã vá de um ponto A ao ponto B, o artista de foley não pode caminhar livremente pelo palco visto o microfone estar fixo na sua captação. É na execução desta categoria de sons que são levadas ao extremo as capacidades técnicas de imitação de um artista de foley e onde são adoptadas técnicas como o seguimento pelos ombros do personagem e não pelos pés na simulação do caminhar.

Finalmente a terceira categoria são os sons específicos que constituem tudo em que o personagem toca, ou que resulte da ação física do mesmo, e.g. som de

uma porta bater. É nesta categoria que o artista de foley estende a sua criatividade até ao limite na procura das soluções para os vários sons que são propostos (Viers, 2011).

Nas categorias de foley acima descritas os artistas de foley investem tempo e esforço em algo secundário na medida em que os sons que compõem uma sessão de foley pretendem somente tornar verosímil o mundo criado pela narrativa visual seja num filme tradicional, de animação ou num jogo digital.

1.2 Enquadramento e motivação da investigação

DIGIT tem como objectivo a exploração de tecnologias multimédia para a realização de foley, mais concretamente, para a realização da segunda categoria do foley, i.e. o som de passos, em ambiente de pós-produção áudio. DIGIT pretende contornar limitações de controlo de soluções digitais já presentes no mercado recorrendo ao gesto em superfícies como forma de controlar um processo digital de geração de som de passos.

Assim a minha investigação pretende dar resposta à seguinte pergunta:

como realizar som de passos de forma digital e interativa de forma a garantir a continuidade do saber-fazer e expressividade patentes nas práticas e resultados do foley tradicional?

Recentemente foram desenvolvidas propostas digitais para a concretização do som de passos. Estas soluções vêm simplificar a tarefa de foley na medida em que reduzem um palco de foley a um sistema digital num computador e dão acesso a uma base de dados extensa de gravações.

No entanto as limitações que estas aplicações apresentam, principalmente na forma de controlo por intermédio de um controlador MIDI, resultam numa falta de expressividade. A minha investigação pretende abordar estas limitações através da exploração do gesto para conduzir uma geração automática de som de passos, que consiga, na medida do possível, potenciar a expressividade e exploração como forma de recriação do som de passos em ambiente de pós-produção áudio para imagem em movimento. Desta forma, enquadro a minha investigação na prática tradicional do foley de som através da exploração táctil de objetos físicos, e

reconhecimento das suas qualidades acústicas, adaptando-os às necessidades temáticas e narrativas da imagem em movimento.

1.3. Objetivos e metodologia de investigação

Tendo como base o conceito de perfil acústico do toque ou gesto[3] pretendi o desenvolvimento de uma aplicação que apresente vantagens em relação às soluções digitais anteriores na medida em que (1) utiliza a descrição do perfil acústico do gesto (nas suas múltiplas dimensões) como controlo do sistema visando desta forma contornar as limitações impostas pelo protocolo de comunicação MIDI (2) dispõe ainda de um motor de reprodução áudio que permita uma alteração das características originais de um som de forma a criar mais possibilidades a partir de um pequeno número de amostras e (3) redução do espaço necessário para dispor de um palco de foley.

Assim a minha investigação tem como objectivos (1) a escolha de uma superfície ideal para realização do interface com o DIGIT, (2) o levantamento de quais os descritores do tipo MPEG7 (Kim, Moreau, & Sikora, 2006) mais relevantes para descrição do perfil acústico e (3) estabelecer um modelo de interação gestual que permita que este seja conduzido para a criação de som de passos.

O desenvolvimento deste modelo em Max[4] tem como base a síntese concatenativa de som através da aplicação de um microfone de contacto numa superfície e consequente reconhecimento e classificação do perfil acústico do toque/gesto de acordo de acordo com parâmetros estabelecidos pelos descritores áudio.

A metodologia na base da minha investigação, tem como objectivo o desenvolvimento de uma aplicação para geração de passos em ambiente de pós produção.

Através da descrição do perfil acústico destes gestos e do seu mapeamento para parâmetros de controlo de síntese concatenativa, pretende-se a realização de uma pesquisa automática em bases de sons de passos pré-gravados para a seleção do segmento que mais se assemelhe à descrição acústica do áudio de entrada. A metodologia aplicada desdobra-se nas seguintes tarefas:

[3] Nesta tese por perfil acústico entende-se a detecção e descrição acústica de movimentos da mão realizados numa superfície (Braun et al., 2015)

[4] https://cycling74.com/

(1) escolher uma superfície de interface para o DIGIT. Nesse sentido procederei à análise das propriedades acústicas de várias superfícies, sob a interação de uma partitura de gestos previamente definidos. Com esta abordagem pretendo aferir qual a superfície mais adequada e quais as propriedades mais adequadas a serem utilizadas pelo módulo de descrição do perfil acústico do gesto.
(2) definir qual a técnica de captação e respetiva colocação do microfone;
(3) interpretação por parte de um sistema digital de características acústicas de gestos numa superfície (definida em 1) através dos descritores de áudio. Será feito também o processamento do sinal da entrada do microfone de contacto e uma análise em tempo real do sinal de entrada do microfone configurada com os descritores mais relevantes adquiridos aquando da fase 1.
(4) pesquisa e reconhecimento numa base de sons de passos, de segmentos semelhantes ao perfil acústico de um sinal de entrada. Efetuarei uma pesquisa automática e classificação de entre uma base de sons. Será ainda necessário efetuar uma ponderação dos valores adquiridos à entrada e na base de dados de forma a poder explorar a base de sons mais uniformemente.
(5) reprodução áudio que se aproxima dos conceitos de síntese granular, mais concretamente *brassage* (Roads, 2004). Por fim foi realizado processamento de limitação (igual ao presente na entrada de sinal) e um módulo de reverberação tipo *plate* para a simulação de espaços acústicos diferentes.

2 Revisão bibliográfica

Neste capítulo descrevo o estado-da-arte das várias áreas abordadas na minha investigação. Este capítulo está estruturado da seguinte maneira.

Na secção 2.1 abordo a descrição do perfil acústico e a taxonomia dos descritores adoptados na minha investigação. Na secção 2.2 descrevo processos de síntese concatenativa. Na secção 2.3 descrevo as várias possibilidades de controlo usados em síntese concatenativa. Na secção 2.4 são referidas aplicações musicais de síntese concatenativa. Na secção 2.5 são referidas aplicações que utilizam uma forma de controlo de síntese sonora com base em interfaces tangíveis. E, finalmente, na secção 2.6 são apresentadas algumas soluções digitais para a realização de foley e mais concretamente para a realização do som de passos.

2.1 Descrição do perfil acústico

A descrição do perfil acústico é feita por um sistema de descritores áudio tipo MPEG-7.[5] O recurso a este protocolo de descritores tem como objectivo poder aferir qual o perfil acústico de um sinal de entrada.

Para sistematizar o vasto tema dos descritores de áudio, que se estendem por variados contextos de aplicações, irei recorrer à taxonomia descrita no projeto Cuidado (Peeters, 2004). Esta fonte revelou-se de extrema importância na área pela concentração de informação acerca deste tema onde as definições semânticas dos vários descritores são elencadas, assim como as respetivas fórmulas de cálculo para os vários descritores apresentados. Acresce ainda que é com base na taxonomia proposta no projeto Cuidado que o sistema de descritores a ser utilizado pelo DIGIT se baseia.

[5] MPEG-7 (Motion Picture Experts Group) são um standard que define a descrição e sistematização de descritores (meta data) para conteúdos multimédia, o que permite que os utilizadores possam procurar, identificar, filtrar e explorar conteúdos áudio visuais (Kim et al., 2006)

2.1.1 Taxonomia dos descritores áudio

Os descritores de áudio do tipo MPEG-7 são classificados de acordo com o tipo de informação oferecida pelos descritores, assim como pelo seu nível de abstração. De acordo com Schwarz (2000) os descritores áudio podem ser organizados em três classes: (1) categórica (pertencente a uma classe); (2) estática (um só valor); e (3) dinâmica (apresentando a evolução temporal).

A distinção entre os vários descritores áudio tem em conta quatro perspetivas:

(1) a estabilidade ou dinâmica da característica a ser extraída;

(2) a extensão de tempo a que se aplica o descritor distinguindo os descritores globais—calculados para todo o sinal, ou seja, para todo o conjunto de janelas, e.g. o ataque de um som, e que requerem uma localização prévia do evento— e os descritores instantâneos que são calculados para cada janela de análise;

(3) qual o nível de abstração que uma característica representa em relação a outra característica;

No que respeita ao nível de abstração os descritores são divididos em: (3.1) descritores de baixo nível são descritores que são computados através de meios simples, e.g. amplitude, centroíde ou obliquidade. São descritores que se baseiam no modelo da informação; (3.2) descritores de médio nível já requerem alguma interpretação da informação, interpretação essa é feita durante uma fase de análise; e finalmente (3.3) os descritores de alto nível que são centrados no utilizador e expressam atributos semânticos como o estilo ou género musical. Esta classe de descritores implica aprendizagem com base no modelo do utilizador, e não só com base no modelo da informação;

(4) qual o processo de extração das características, ou seja, se são calculadas: (4.1) a partir do domínio temporal do sinal, e.g. taxa de cruzamento de zero (*zero-crossing rate*); (4.2) a partir de uma transformada de Fourier no sinal, e.g. centroíde espectral; (4.3) a partir de um modelo do sinal, e.g. modelo fonte-filtro; ou (4.4) a partir de uma modelação do sistema auditivo, e.g. bancos de filtro Bark.

O conjunto de características a serem analisadas relacionam-se pelas respetivas (1) características da envolvente temporal; (2) características temporais instantâneas; (3) características de energia; (4) características da envolvente espectral; e, por fim, (5) características da dinâmica espectral.

Depois de uma análise estatística dos resultados de várias gravações efectuadas para avaliação do comportamento de diversas superfícies, aferi que a classe de descritores que continham em si mais informação que pudesse ser

interpretada pelo sistema seriam as referidas nos pontos (2), (3) e (4). As características referidas em (1) e (5) estão fora do âmbito da investigação. Esta opção tem a ver diretamente com a característica peculiar do som dos gestos pretendidos na interface do DIGIT. Estes consistem em sons de curta duração (cerca de 450 milissegundos), portanto com uma evolução temporal inexistente, onde o que interessa retirar dos descritores é a característica temporal instantânea de cada som.

2.1.2 Características de energia

Em Peeters (2004) as características de energia desdobram-se em: (1) energia global como estimação da potência do total do sinal a um determinado momento; (2) energia harmónica como estimação da potência do sinal com conteúdo harmónico a um determinado momento; e (3) energia de ruído como estimação da potência do sinal com conteúdo não harmónico a um determinado momento.

2.1.3 Características da envolvente espectral

São uma classe de descritores que extraem medidas do sinal de áudio no domínio das frequências e indicam propriedades da distribuição de energia no espectro de frequências. Estes descritores são calculados, a partir de um sinal contínuo, com recurso ao FFT (Transformada Rápida de Fourier):

- *centroide espectral*[6] (*spectral centroid*)—o centro de gravidade da envolvente espectral de um som a cada janela de análise;
- *desvio padrão espectral* (*spectral spread*)—a dispersão da distribuição da energia espectral em relação ao centroide espectral;
- *obliquidade espectral* (*spectral skewness*)—medida da assimetria da distribuição em torno de uma média;
- *curtose espectral* (*spectral kurtosis*)—medida de dispersão que caracteriza o *achatamento* da distribuição energética do envelope espectral;

[6] Nesta tese adopto uma nomenclatura para referir-me aos descritores em português baseada em (Monteiro, 2012).

- *medida de nivelamento espectral* (*spectral flatness*)—indica a homogeneidade da distribuição de energia no espectro, ou seja, quanto a envolvente espectral é próxima de *plano*;
- *inclinação espectral* (*spectral slope*)—fornece uma estimativa da energia espectral computada por regressão linear no espectro da magnitude;
- *deslizamento da frequência* (*spectral roll-off*)—define como o valor de frequência que assinala o ponto em que a somatória da energia dos componentes espectrais abaixo desse ponto contém 95%[7] do total da energia do espectro
- *irregularidade espectral* (*spectral irregularity*)- mede o grau de diferença entre magnitudes de componentes espectrais adjacentes, descrevendo se o contorno da envolvência espectral é suave ou se apresenta picos.
- *fluxo espectral (spectral flux)* - mede a diferença de magnitude entre janelas sucessivas de análise o qual retorna valores baixos quando o sinal tende a ser estático e valores altos quando o sinal tende a sofrer mais alterações.[8]

Com este levantamento sobre a pertinência de vários descritores num contexto de análise de sons não musicais pretendi estabelecer um levantamento que fosse significativo para o tipo de análise de sinais de entrada que poderei usar no DIGIT.

2.2 Síntese concatenativa

Em (Schwarz, 2006) os métodos de síntese concatenativa pressupõe a utilização de uma extensa base de sons pré gravados, segmentados em unidades, e onde um algoritmo de seleção de unidades encontra a unidade que melhor coincide com o som ou controlo de entrada. O segmento selecionado é designado de alvo, esta seleção dos alvos é realizada de acordo com uma taxonomia de descritores das unidades que são características extraídas das fontes de sons. Posteriormente as unidades selecionadas são concatenadas.

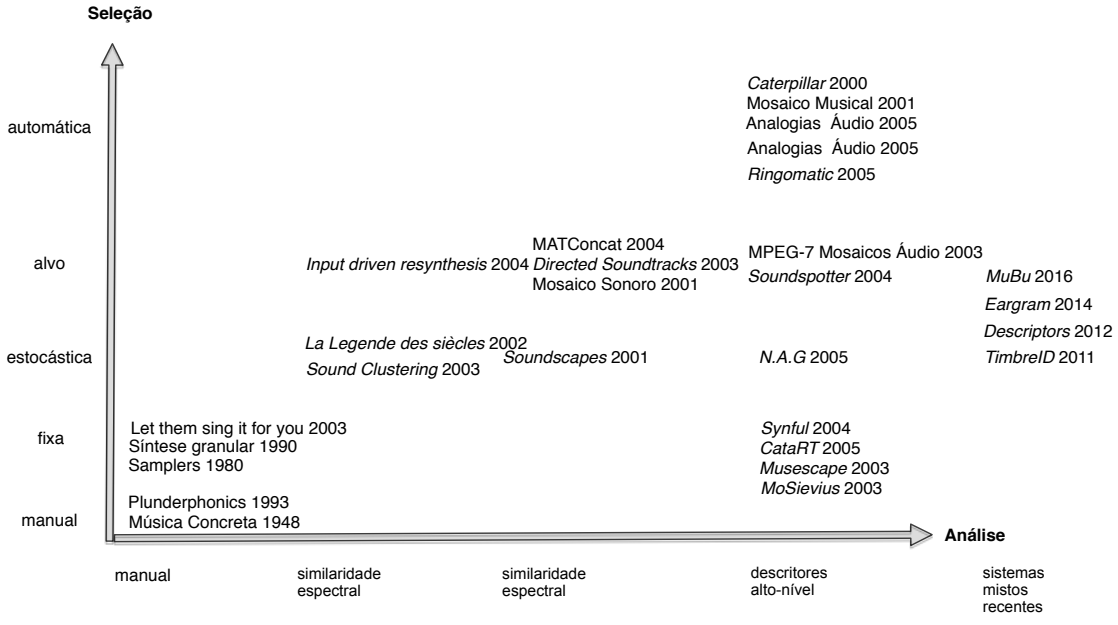
[7] Esta variável pode assumir outros valores.

[8] Descritor específico utilizado pela biblioteca **descriptors** é mkl (*Modified Kullback Leibler*) que pretende ser uma medida mais avançada que a do fluxo espectral medindo a entropia relativa entre duas distribuições probabilísticas *P* e *Q* de uma forma simétrica (valores de um janela anterior e valores de janela seguinte).

A síntese concatenativa ao utilizar gravações de sons baseia o seu funcionamento, não num sistema de regras pré definido e.g. como num modelo de síntese generativa, mas num sistema em que as regras são inferidas pela própria informação.

Ainda Schwarz (2006) define quatro contextos de aplicações da síntese concatenativa (1) instrumentos de síntese alto nível onde são realizadas transições muito naturais ao selecionar as unidades de um determinado contexto de *matching*, e onde a informação atribuída aos sons é a base de para a exploração da base de sons, (2) resíntese de áudio de acordo com descritores áudio, sendo normalmente referido como mosaico sonoro, (3) síntese de texturas e ambientes sonoros usadas em instalações ou para pós produção áudio para cinema e onde são geradas as bandas sonoras a partir de bibliotecas de sons, ou em situação de extensão de uma paisagem sonora de forma a faze-la coincidir com a imagem, e (4) síntese livre a partir de bases de dados heterogéneas, e que se for realizada em tempo real permite a exploração de um *corpus* sonoro de forma interativa.

Figura 1: Perspectiva histórica da síntese concatenativa (2006)



Os sistemas de síntese concatenativa apresentados na Figura 1 estão organizados cronologicamente e de acordo com dois aspectos (seleção e análise) que quando combinados indicam o nível de automatismo e abstração. No eixo *x* é

indicado o grau de estruturação da informação obtida pela análise dos sons de base e a meta data, e no eixo *y* o grau de automatização da seleção.

No que respeita à síntese das unidades selecionadas são usados os métodos como a aplicação de um simples *crossfade* de concatenação entre as unidades, a aplicação de técnicas de síntese de fala de forma a reduzir as descontinuidades entre os segmentos, e a utilização de modelos mais avançados de sinal como o modelo aditivo sinusoidal mais ruído ou o recurso a técnicas de síntese granular (Schwarz, Grégory, Bruno, & Sam, 2006).

O que se pretende com a aplicação de técnicas de síntese concatenativa é uma exploração dinâmica e interativa de uma base de sons onde métodos de pesquisa automática, ou manual, funcionam como um meio de exploração interativa dessa mesma base.

2.3 Controlos à disposição da síntese concatenativa

Segundo Schwarz (2012), os tipos de controlos do processo de síntese concatenativa organizam-se em dois grupos: (1) controlos posicionais 2D e 3D, e.g. controladores MIDI X-Y, superfícies de multi toque, sistemas de captura de movimento ou acelerómetros e (2) controlo por análise de áudio de entrada, e.g. microfones de contacto e análise áudio de entrada com recurso a um sistema de descritores.

Dentro destes, o controlo por análise de áudio de entrada é o mais relevante para a minha investigação e consiste na recuperação de unidades de áudio numa base de sons através da descrição do perfil acústico da entrada de acordo com um sistema de descritores standard tipo MPEG-7.

Os sensores acústicos têm sido usados em atividades de reconhecimento como por exemplo a detecção de presença e atividade humana dentro de uma sala, encontrando semelhanças entre um som captado e uma base de sons previamente criada (Braun, Krepp, & Kuijper, 2015).

Em sistemas que recorrem a um só microfone Zamborlin (2009) utiliza o sinal de um microfone de contacto para reconhecimento de diferentes modos de interação com superfícies, e onde, por um lado, cada modo está associado a um filtro ressoador que é usado para processar a entrada, e por outro lado, modos semelhantes controlam a intensidade dos filtros (Françoise, 2015).

Em sistemas que recorrem a mais do que um microfone (Braun et al., 2015) demonstra a aplicação da técnica de *location template matching* utilizando métodos

de aprendizagem máquina onde as várias características do sinal áudio são correspondidas com características treinadas e aprendidas pelo sistema. Em (De Sanctis, Rovetta, Sarti, Scarparo, & Tubaro, 2006) a aplicação da técnica de *time delay of arrival* mede os tempos de atraso da chegada do sinal a cada um dos microfones, deduzindo a partir destes dados em que local da superfície foi executado um gesto.

2.4 Aplicações musicais de síntese concatenativa e respectiva forma de controlo

Esta secção está organizada tendo em conta as formas de controlo de síntese concatenativa descritas na secção anterior.

A partir de controlos posicionais 2D no earGram[9] (Bernardes, Guedes, & Pennycook, 2012) e CataRT[10] (Schwarz et al., 2006) apresentam uma navegação num espaço projetado no ecrã do computador. Por intermédio do rato, o utilizador navega nas coordenadas X e Y no ecrã do computador acedendo desta formas às várias unidades de áudio que compõem o corpus sonoro.

Em Schwarz, Tremblay e Harker (2014) altera-se o paradigma de controlo (a partir da análise do áudio de entrada) e de navegação (interação do gesto numa superfície de contacto) na base de sons com um recurso a microfones de contacto colocados em superfícies sólidas, e onde através de técnicas de impulso resposta dá-se a convolução em tempo real do sinal do microfone com os vários sons que compõem o corpus sonoro (Harker & Tremblay, 2012).

Em *Tangible Scores*,[11] Tomás (2014) define o seu sistema como um sistema de partituras tangíveis que são compostas por uma camada física metálica incorporada na configuração do instrumento digital com a intenção de conduzir gestos tácteis e movimentos e para controlar os resultados do sistema. Em Polotti (2005) o objetivo final é o de reencontro da energia acústica de uma superfície excitada com a geração de som sintetizado.

Schwarz (2012) avança a possibilidade de detecção da posição da mão num plano 2D. De igual forma (Duindam & Leeuw, 2015) discutem a utilização do controlador digital Tingle[12] numa tentativa de recapturar a sensação táctil acústica

[9] https://sites.google.com/site/eargram/, último acesso a 29 de Junho 2016

[10] http://imtr.ircam.fr/imtr/CataRT, último acesso a 29 de Junho 2016

[11] http://interface.ufg.ac.at/blog/projects/tangible-score/, último acesso a 29 de Junho 2016

[12] http://www.nupky.com/, último acesso a 29 de Junho 2016

através do feedback háptico deste controlador, proporcionando desta forma novas oportunidades de expressão musical.

Na mesma linha de controlo de síntese concatenativa, DIRTI[13] (Savary, Pellerin, Cahen, Ateliers, & Massin, 2013) explora as possibilidades oferecidas pelas superfícies tangíveis com recurso a materiais físicos granulares ou líquidos de modo a criar um design de interação onde questões de feedback auditivo e aprendizagem sensorial e motora são potenciadas de uma forma expressiva.

Um projeto de interesse especial interesse para a minha investigação é C-C-Combine[14] onde é explorada um aplicação de mosaico sonoro com recurso a sinal áudio de entrada.

2.5 Controlo de síntese sonora com base em interfaces tangíveis

Nesta secção pretendo referenciar aplicações que apesar de não terem na sua base princípios de síntese concatenativa, abordam a descrição acústica do gesto como forma de controlo de processos de síntese sonora.

Quando se aborda a questão dos interfaces tangíveis é inevitável a referência ao projeto *Reactable*[15] (Kaltenbrunner, Jordà, Geiger, & Alonso, 2006) que desenvolveu estratégias onde o toque é partilhado entre utilizadores de forma a incentivar a colaboração na criação de uma peça musical. O gesto nesta aplicação controla parâmetros de um determinado modelo de síntese.

Em *Legos*[16] (Tajadura-Jim & Bianchi-Berthouze, enez and Nadia, E. Furfaro, 2015) os autores estudam o efeito do feedback auditivo para apreensão de gestos em sistemas interativos. Este sistema baseia-se em experiências de tecnologias sonoras e musicais, assim como na neurociência e controlo motor.

O projeto *Geecos*[17] realizado no IRCAM utiliza microfones de contacto em superfícies construídas para o propósito e onde é explorada a expressividade do gesto como condutor de um modelo de síntese modal aumentando virtualmente a acústica da superfície de interação.

[13] http://www.userstudio.fr/projets/dirti-for-ipad/, último acesso a 29 de Junho 2016

[14] http://www.rodrigoconstanzo.com/combine/, último acesso a 29 de Junho 2016

[15] http://reactable.com/, último acesso a 29 de Junho 2016

[16] http://legos.ircam.fr/, último acesso a 29 de Junho 2016

[17] http://medias.ircam.fr/x45a1fc, último acesso a 29 de Junho 2016

A expressão comercial do conceito de superfície tangível e ubíqua encontra-se concretizada em *Mogees.*[18] (Bevilacqua et al., 2009). Trata-se de uma tecnologia que torna objetos físicos triviais em autênticos instrumentos musicais, ao converter vibrações acústicas das superfícies de contacto em reprodução de sons que constam no sistema. O sensor é acompanhado por uma aplicação que deteta e analisa as propriedades físicas do objetos, aumentando-as de forma a criar sons musicais. Qualquer superfície pode ser utilizada, sendo que o sistema faz coincidir de forma adaptativa o sinal de entrada do sensor acústico ao tipo de síntese inerente ao sistema.

De entre as aplicações para iOS que operam sobre o perfil acústico destaco Impaktor, [19] um sintetizador de ritmos onde os impulsos acústicos reais capturados pelo microfone são utilizados como fonte de excitação para avançados módulos de síntese modal capazes de simular o comportamento de membranas, címbalos, metalofones e cordofones.

Uma outra aplicação iOS de especial importância para a minha investigação pelo facto de abordar a utilização do gesto como forma de controlo para gerar som de passos é iFoley Footsteps[20] onde o utilizador recria os sons de passos através da navegação no espaço táctil do dispositivo móvel. Esta aplicação apresenta também a possibilidade de alteração do tipo de piso onde ocorre o som de passos assim como um módulo de reverberação para processamento de espaços acústicos diferentes.

Projetos independentes como Stretta,[21] os impulsos são captados por um microfone de contacto e convertidos em parâmetros de síntese ou *sampling*, e Pulse[22] onde é realizado um *mapping* inteligente que converte as ondas acústicas em notas MIDI que podem ser utilizadas para controlo de instrumentos electrónicos.

[18] http://mogees.co.uk/, último acesso a 29 de Junho 2016

[19] http://www.beepstreet.com/ios/impaktor, último acesso a 29 de Junho 2016

[20] http://www.ifoleyapp.com/, último acesso a 29 de Junho 2016

[21] https://cycling74.com/practical-max/practical-max-1/#.V2wpwSMrK2x, último acesso a 29 de Junho 2016

[22] http://www.tetmusic.com/, último acesso a 29 de Junho 2016

2.6 Soluções e limites de ferramentas digitais para a realização de foley

Conforme o descrito no Capítulo 1 o foley continua a ser um processo manual que assenta nas capacidades técnico-interpretativas do artista de foley. Nesta secção o que se pretende é dar a conhecer outras possibilidades de realização de foley, nomeadamente e, mais recentemente, algumas soluções digitais que foram desenvolvidas para facilitar a tarefa do foley, com especial foco na realização do som de passos pela incidência do tema no meu projeto.

O recurso a bibliotecas de som que incluem diversos sons de passos em vários tipos de superfícies, ou pisos, sempre foram ferramentas tradicionais da pós produção áudio. Exemplo destes produtos são comercializados por empresas como a Sound Ideas[23] ou pela companhia Pro Sound Effects[24]. No entanto esta solução revela-se uma opção limitada pela rápida exaustão de propostas apresentadas por esta solução e dificuldade de seleção das mesmas (Viers, 2011).

No sentido de simplificar a tarefa de realização do som de passos foram desenvolvidas recentemente aplicações que têm como função dar alternativas digitais para a concretização destes tipo de sons.

Destas ferramentas digitais destaco a aplicação AudioStepsPro, uma ferramenta comercial da empresa AudioGaming[25], e a Boom Library Virtual Foley Artist[26]. Ambas as aplicações são *sampler plugins* que se baseiam em centenas de amostras de sons de passos previamente gravados.

Estas aplicações dispõem de controlos como controlo sobre a sequência de passos esquerdo e direito, definição de valores de aleatoriedade na escolha das amostras que vão ser reproduzidas por parte do software ou mesmo controlo de parâmetros como a irregularidade dos passos. Estas opções permitem diferenciar passos de corrida e passos de andar normal. Este controlo é efetuado através de comandos que seguem o protocolo de comunicação MIDI.

Apesar destas aplicações virem simplificar a realização do som de passos, apresentam duas limitações que pretendo colmatar com a minha investigação: (1) a falta de expressividade, que advém pelo facto de serem controladas por um

[23] https://www.sound-ideas.com/, último acesso a 29 de Junho 2016

[24] http://www.prosoundeffects.com/, último acesso a 29 de Junho 2016

[25] http://lesound.io/product/audiosteps-pro/, último acesso a 29 de Junho 2016

[26] https://www.boomlibrary.com/boomlibrary/products/virtual-foley-artist-footsteps, último acesso a 29 de Junho 2016

controlador MIDI, que faz uma truncagem de valores grande, eliminado desta forma pequenas variações e subtilezas durante a execução destes efeitos; e (2) a limitação que estas aplicações apresentam no que respeita à criação de sons de passos não reais, e.g. para uma qualquer criatura ficcionada de um jogo digital.

Fora do âmbito comercial também foram desenvolvidas algumas propostas para a realização de som de passos, mais concretamente através de técnicas de áudio procedimental. No contexto do design de som, o uso de áudio procedimental tem a sua expressão máxima no trabalho de Andy Farnell ("Designing Sound | The MIT Press," 2010) onde o autor propõe a realização de sons de passos de forma totalmente processual, não se preocupando com a total definição do som, mas dando somente as características fundamentais que permitem ao ouvinte identificar o som em questão fazendo uso de efeitos psicoacústicos (e.g. *cocktail party effect*).

3 Superfícies de Controlo para Síntese Concatenativa

Sendo que o DIGIT assenta num paradigma de controlo de processos de síntese concatenativa com base em análise do áudio de entrada, tive a necessidade de abordar várias possibilidades de superfícies físicas para poder aferir a melhor superfície que pudesse servir de interface do DIGT.

Devido à falta de estudos sistematizados sobre quais os melhores materiais, formas, dimensões e geometria, para a realização desta tarefa foram elaboradas faseadamente as experiências que passo a descrever:

(1) levantamento das várias possibilidades de materiais e formas que pudessem ser utilizadas como interface do DIGIT;

(2) definição de uma taxonomia de gestos que pudessem servir de amostra representativa das várias possibilidades de interação;

(3) análise das respostas acústicas de cada material quando sujeito à taxonomia de gestos definida em (2).

A realização destas experiências teve como objectivo encontrar uma superfície que apresentasse mais qualidade - e, por qualidade entendo como sendo a superfície que apresente uma resposta acústica o mais coesa para mesmos gestos e, ao mesmo tempo, que apresente uma resposta acústica o mais distinta para gestos diferentes.

3.1 Análise da Qualidade Expressiva de Superfícies de Controlo de Síntese Concatenativa

3.1.1 Definição de superfícies de controlo

O foley é uma prática exploratória, tanto ao nível da gestualidade usada para interpretar os efeitos de som, como também ao nível da escolha dos utensílios que são explorados para a sua criação.

Com a ideia de objecto encontrado procedi a um levantamento de possíveis superfícies que poderiam servir de interface para o DIGIT, tendo em conta características físicas como o material que as compunha e respectivas dimensões.

De entre uma panóplia quase infinita de possibilidades restringi os meus testes a seis superfícies com respostas acústicas diferentes (Henrique, 2002): caixa de madeira, kalimba madeira, tampo de mesa de madeira, placa de metal, caixa de metal e um jarro de vidro.

O processo de escolha das superfícies a serem testadas foi guiado inicialmente por uma escuta intuitiva e pelo conhecimento particular do comportamento acústico destes vários materiais resultante da minha experiência profissional na área do foley. Esta escolha pretendeu enquadrar o foley num ambiente digital potenciando a experimentação inerente e o modo de operação da prática.

Figura 2: Superfícies usadas para testes de interface: (a) caixa de madeira, (b) kalimba de madeira, (c) tampo de mesa madeira, (d) placa de metal, (e) caixa de metal e (f) jarro de vidro. É usado um rato MagicMouse da Apple (11.43cm de comprimento) para servir de referência às dimensões dos objetos

a

b

c

d

e

f

Para capturar a resposta acústica de vários gestos nas superfícies optei pela utilização de um microfone de contacto. A opção por esta técnica de captação teve dois objectivos (1) possibilidade de realizar uma captação definida de gestos subtis executados na interface e (2) pelas características de transdução do microfone de contacto que tornam o dispositivo imune à captura de sons que não tenham origem somente naqueles que advêm da interação com a própria superfície.

No entanto, tendo em conta que a captação áudio com recurso a microfones de contacto diverge da forma como naturalmente o som é percepcionado pelo sistema auditivo humano, na medida em que o microfone de contacto não utiliza o ar como canal de transmissão para as vibrações acústicas mas sim a vibração do física dos materiais quando sujeitas a uma interação, foram realizados testes de comparação que consistiram na execução de uma sequência de gestos captada por um microfone de contacto e por um microfone de condensador tradicional.

Desta forma pretendeu-se aferir quais as principais diferenças entre estas duas técnicas de captação e quais as características acústicas mais relevantes para o design do DIGIT.

3.1.2 Definição de uma taxonomia de gestos

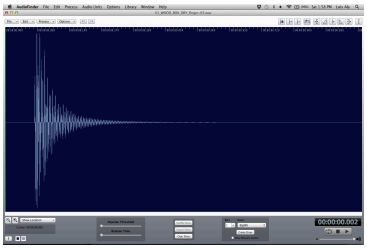
Sendo a prática tradicional de foley naturalmente a de uma exploração manual dos objetos de forma a retirar destes conteúdo sónico, o gesto é algo que está inerente ao *foley*, foi delineada na minha investigação uma taxonomia de gestos que servisse de amostra representativa das várias possibilidades de exploração táctil das várias interfaces. Esta amostra recorre não só ao gesto produzido pelas mãos diretamente nas superfícies como recorre a adereços como

forma de criar vibrações que não são possíveis somente através da interação com a mão.

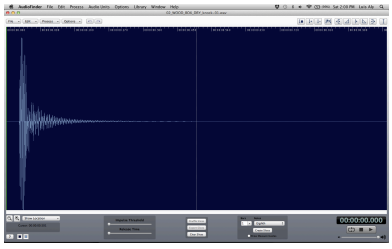
Neste sentido foram realizadas gravações áudio, a uma frequência de amostragem de 44.1kHz e uma precisão de 16 bits. As gravações realizadas é tiveram um tratamento do sinal do microfone através de *gate* e limitação de volume de forma a evitar distorções de amplitude (Schwarz et al., 2014). Foram realizadas as seguintes gravações:

- toque com um só dedo (a)
- bater com nó do dedo (b)
- arranhar com unha (c)
- arranhar com tira metálica fina (d)
- arranhar com tira metálica grossa (e)
- toque múltiplo com unhas (f)
- toque abafado com a palma da mão (g)

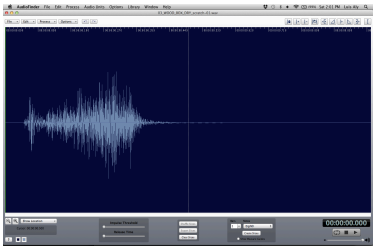
Figura 3: Representação da forma de onda (amplitude) das várias interações de gestos: (a) toque com um só dedo, (b) bater com nó do dedo, (c) arranhar com unha, (d) arranhar com tira metálica fina, (e) arranhar com tira metálica grossa, (f) toque múltiplo com unhas e (g) toque abafado com a palma da mão



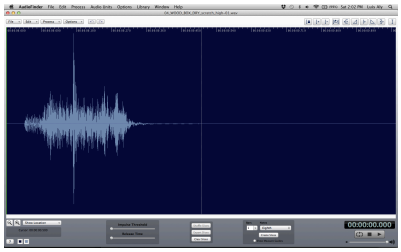
a



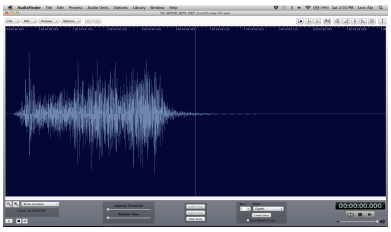
b



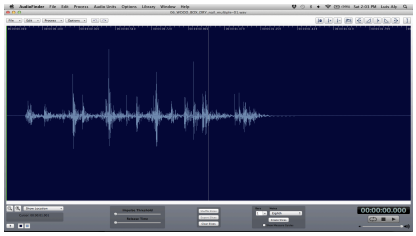
c



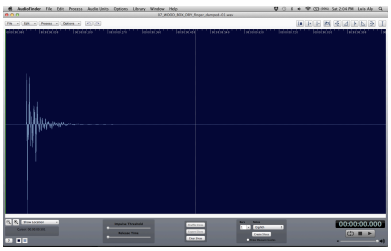
d



e



f



g

Na Figura 3 represento a evolução temporal da amplitude da gravação das várias sequências simples de gestos captadas por um microfone de contacto. A partir de uma análise atenta destas representações deduz-se que cada gesto tem uma impressão acústica diferente, sendo que estas pequenas variações na interação são perfeitamente apercebidas pelo microfone de contacto. Estas representações serviram como suporte à decisão de utilização do microfone de contacto no interface do DIGIT em prejuízo de um microfone de condensador.

3.1.3 Comparação de Gestos de Controlo de Síntese Concatenativa de Som nas Várias Superfícies

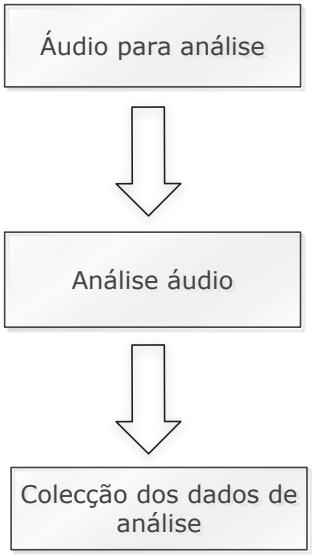
Para uma avaliação objectiva das qualidades das várias superfícies testadas procedi a uma análise estatística da informação adquirida pelos descritores com o objectivo de comparar as várias superfícies e poder, desta forma, avaliar a qualidade da expressividade de cada uma das superfícies experimentadas.

Por expressividade, no contexto da minha investigação, entende-se a resposta acústica particular de uma superfície quando sujeita à interação por intermédio de um gesto. O que pretendi foi encontrar uma superfície que fosse o mais linear possível dentro de um quadro de gestos iguais, e que fosse o mais diferente possível ao executar um âmbito de gestos diferentes dando, desta forma, ter a possibilidade de uma mesma superfície poder ser utilizada como interface do DIGIT com o objetivo de conseguir um âmbito mais alargado de propostas de som de passos avançadas pelo DIGIT.

Para realizar a avaliação das superfícies foi feita uma gravação dos sete gestos definidos na secção 3.1.2, com oito repetições de cada gesto, de forma a aumentar a robustez da representação de cada gesto nas seis superfícies definidas na secção 3.1.1.

Para a descrição áudio das gravações recorri ao ambiente de programação Max e à biblioteca de objetos externos **descriptors** de Alex Harker.[27] A escolha desta biblioteca prende-se com o facto de poder ter a possibilidade de uma descrição de sinais de áudio em tempo-real e em tempo não real. O digrama seguinte representa o fluxo em Max para a análise das gravações realizadas.

Figura 4: Representação do fluxo seguido para a análise das gravações



Para cada gesto individual foram extraídas as características do **descriptors** referentes aos seguintes dez descritores: energia, deslizamento da frequência, fluxo espectral, mkl, *loudness*, centroíde logarítmico, desvio padrão espectral logarítmico, obliquidade espectral logarítmica, curtose espectral logarítmica e irregularidade espectral. A análise utilizou os seguintes parâmetros: uma janela de análise com 2048 amostras (44.1kHz e 16 bits), sobreposição de 25% entre janelas consecutivas e um tipo de janela Hanning. Foi feita uma análise dos primeiros 250 ms de cada gesto de forma a retirar a informação mais relevante que diz respeito às características iniciais de um som, ou seja a transiente de ataque.

[27] http://www.alexanderjharker.co.uk/Software.html, último acesso a 29 de Junho 2016

Sendo que o objecto **descriptors** calcula por defeito uma média de valores ao longo do tempo de análise pretendido, estes valores foram extraídos de forma a poderem ser comparados sob duas perspectivas:

(1) quais os descritores mais pertinentes, ou seja, que descritores apresentavam mais sensibilidade, por sensibilidade entendo mais variabilidade de resultados perante gestos diferentes.

(2) dar a possibilidade de avaliar de forma mais objectiva qual a superfície com melhor qualidade, ou seja, qual a superfície de entre todas que apresentasse quais os descritores que para um mesmo gesto apresentassem mais proximidade e em relação a gestos diferentes mais afastamento. Este procedimento é explicado em pormenor na secção seguinte.

3.1.4 Avaliação da Qualidade Expressiva das Superfícies

Nesta fase da investigação realizei dois tipos de teste para (1) avaliar a qualidade expressiva das várias superfícies definidas na Secção 3.1.1 e (2) definir o conjunto de descritores a ser utilizados no DIGIT para a descrição do áudio com base de na correlação entre a descrição do perfil acústico hápticos e perceptivos.

Para uma avaliação objectiva das qualidades das várias superfícies testadas procedi a uma análise estatística da informação adquirida das várias repetições dos gestos enunciados na Secção 3.1.2, com o objectivo de aferir a melhor superfície para o DIGIT. Mais concretamente, medi a compactação (intra-aglomerado) e afastamento (inter-aglomerado) dos gestos usando o índice Davies-Bouldin (Davies & Bouldin, 1979), uma técnica de validação e classificação de aglomerados (Saitta, Raphael, & Smith, 2007). Cada aglomerado é aqui interpretado como um gesto. No que respeita à classificação dos aglomerados importa reter que, no índice Davies-Bouldin, configurações ótimas—com o mínimo de distâncias intra-aglomerados e máximo de distâncias extra-aglomerados—correspondem a valores baixos.

No segundo teste pretendi aferir quais os descritores que apresentam melhor correlação entre as características extraídas dos gestos através do microfone de contacto e as mesmas características extraídas a partir do microfone de condensador.

O objetivo foi o de encontrar os descritores que ofereçam uma correlação alta entre o que é percebido sensorialmente e o que é captado pelo sensor háptico. Desta forma, a experimentação e estímulo perceptual (capturado através do

microfone de condensador) é garantido no perfil acústico capturado pelo microfone de contacto.

Na minha investigação utilizei dois coeficientes de correlação para comparar os dados provenientes dos dois microfones: os coeficientes de KendallTau e de SpearmanRho (Bolboaca, 2006). Ambos estes coeficientes de correlação medem o grau em que duas variáveis seguem o mesmo nível de ordenação.

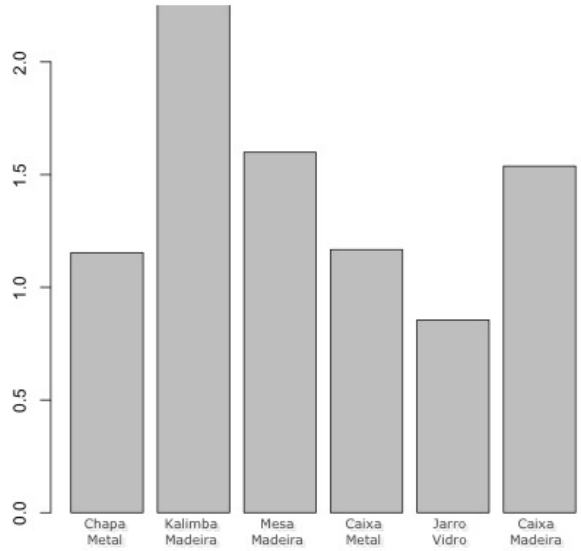
Desta forma pretendi que o design do DIGIT fosse ao encontro da prática tradicional do foley, na medida em que o DIGIT teria de responder de forma eficaz a pequenas variações de interação, e fazer com que essas variações fossem percebidas pelo sistema e traduzidas em, e.g. sons de passos diferentes ou com sons de passos com volumes diferentes.

3.1.5 Resultados da Avaliação da Qualidade Expressiva das Superfícies

Na Figura 5 apresento os resultados dos primeiros testes efectuados para medir a qualidade expressiva das superfícies na interação com os gestos pré-definidos, indicando para tal o índice de Davies-Bouldin para cada superfície.

Os resultados do teste definem três superfícies como ótimas (que minimizam o índice de Davies-Bouldin) para suportar o objectivo proposto para a interface do DIGIT: X y Z. Dentro destas, acabei por adoptar a caixa de metal como interface para o DIGIT, porque, apesar de não ser a superfície que apresenta o melhor resultado, é a que em termos de reprodutibilidade se torna mais viável, pela possibilidade de expansão, escalar o sistema e por razões de segurança que na realização da tarefa do foley tem de se ter em conta porque materiais como o vidro representam algum tipo de perigo para o artista de foley.

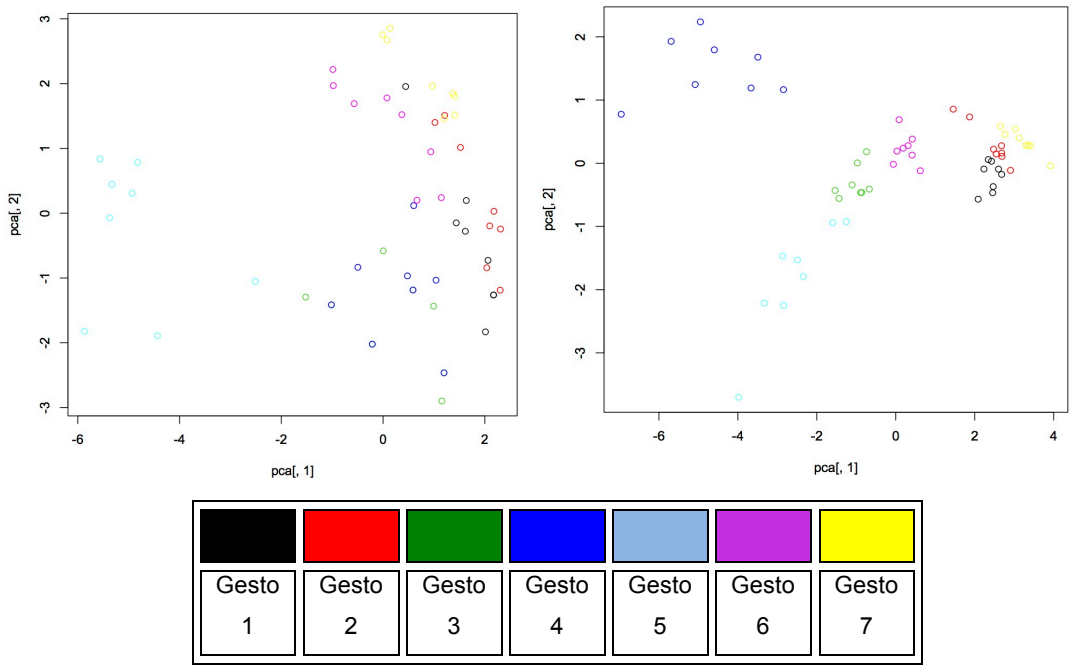
Figura 5: Visualização da informação pelo índice Davies-Bouldin para todas as superfícies de interface



Para proceder à visualização do espaço de descrição dos gestos representados por vectores multidimensionais apliquei uma estratégia que reduz os dados a dois dos seus componentes principais (usando a técnica de Análise dos Componentes Principais). As representações da Figura 6 mostram os vários gestos representados num espaço a duas dimensões para a melhor e pior superfícies.

Cada gesto é representado por uma cor diferentes enquanto que gestos diferentes (representados por um cor diferente) deverão ocupar zonas diferentes do espaço, por conseguinte, apresentar uma maior distância euclidiana.

Figura 6: Visualização dos gestos no espaço de descrição em 2D da superfície (kalimba madeira) com o resultado mais baixos no que respeita à qualidade do aglomerado (esquerda) e da superfície escolhida como interface do DIGIT (caixa de metal) (direita)

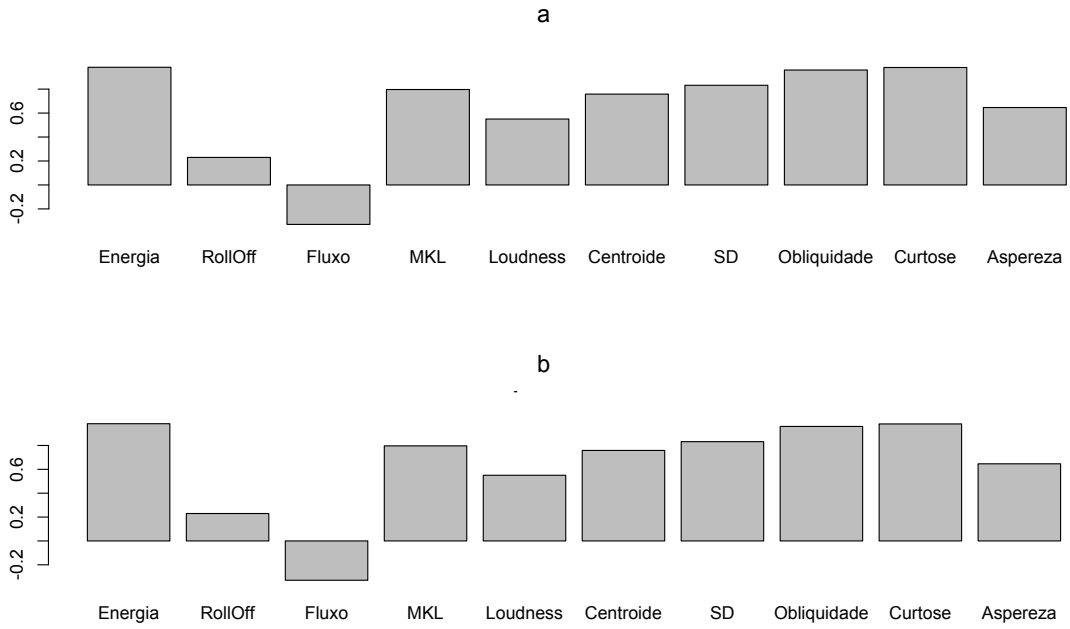


Na tabela 1 e Figura 7 apresento os valores das classificações referentes ao segundo teste, que mede os coeficientes de correlação de KendallTau e SpearmanRho, com o objectivo de aferir a escolha dos descritores com melhor correlação entre o microfone de contacto e o microfone de condensador a serem usados como características representativas dos segmentos de áudio de passos no DIGIT.

Tabela 1: Visualização dos resultados dos testes de correlação Kendall Tau e SpearmanRho. A negrito estão identificados os descritores escolhidos para a tarefa de caracterização das unidades no DIGIT.

	Energia	RollOff	Fluxo	MKL	Loudness	Centroíde	SD	Obliquidade	Curtose	Aspereza
Kendal Tau (τ)	**0.928**	0.187	-0.233	**0.662**	0.440	**0.548**	**0.674**	**0.841**	**0.897**	0.459
Spearman's Rho (ρ)	**0.928**	0.229	-0.329	**0.796**	0.550	**0.758**	**0.831**	**0.959**	**0.980**	0.646

Figura 7: Resultados dos testes de correlação (a) Kendall τ e (b) Spearman ρ



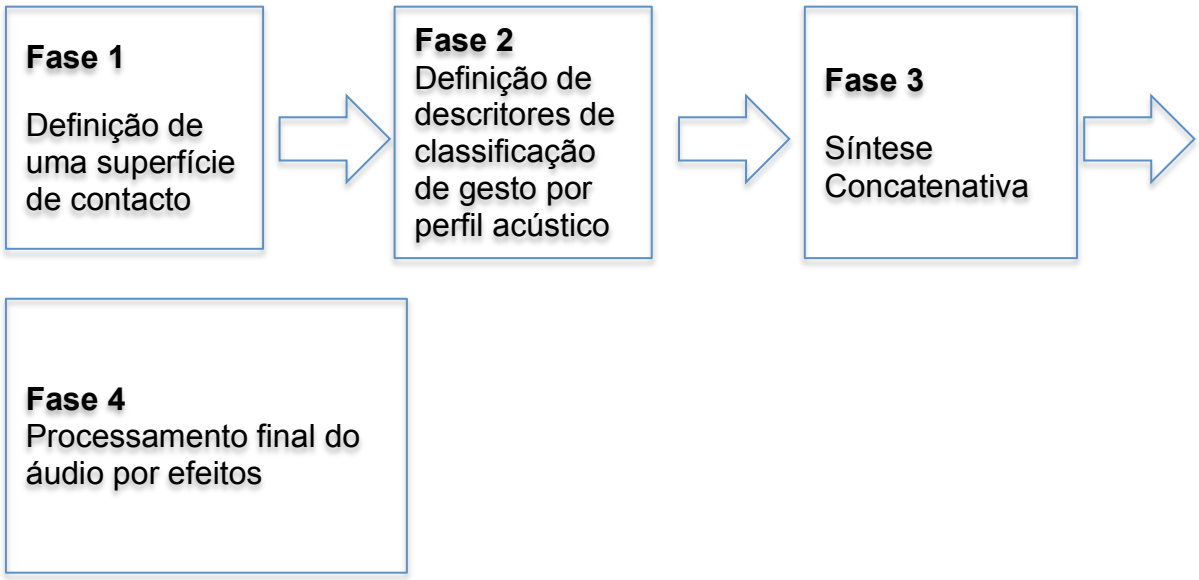
Assim, e como resultado dos testes elaborados, cheguei à conclusão de que os descritores que iriam ser usados pelo DIGIT para definição do perfil acústico do gesto seriam os cinco descritores com maior correlação: energia, MKL, centroíde espectral logarítmico, desvio de padrão espectral logarítmico, obliquidade espectral logarítmica e curtose espectral. Apesar do centroíde logarítmico não apresentar os melhores resultados decidi utilizar este descritor pela falta de um descritor que represente a variação de altura do som, relevante para a tarefa de exploração de objetos no domínio do foley.

4 DIGIT: Geração de Passos por Síntese Concatenativa

Neste capítulo descrevo a arquitetura do sistema DIGIT. Em maior detalhe, descrevo as seguintes fases da implementação (1) a descrição e classificação do áudio de entrada de acordo com os descritores mais relevantes de acordo com as experiências descritas no capítulo anterior, (2) pesquisa automática numa base de de sons composta por gravações de sons de passos, (3) reprodução dos segmentos selecionados de acordo com métodos de síntese concatenativa por intermédio de um motor granular, e, finalmente, (4) o processamento final do áudio por intermédio de efeitos como reverberação.

A figura seguinte representa um diagrama de fluxo de sinal do DIGIT desde a superfície de contacto, a interface, até ao resultado final áudio do som de um passo num determinado espaço acústico.

Figura 8: Metodologia



4.1 Criação de uma base de sons de passos

De forma a cobrir um maior número de cenários possíveis durante a realização do som de passos em ambiente de pós produção, criei várias bases de sons no DIGIT que cobrem uma variedade de gravações de sons de passos com o mesmo tipo de calçado (sapato de sola dura), pisos e *gait.* [28]

A variedade de pisos permite usar o DIGIT quando ocorrem mudanças do espaço físico retratado na imagem. Gait diferentes têm como objectivo dar a possibilidade de o DIGIT responder a alterações de velocidade de movimento de um personagem, ou mesmo a intenção do andar, se mais apressado ou de passo mais arrastado. Então foi construída uma base de dados de som de passos composta pelos seguintes sons individuais:

- 300 passos em madeira a uma velocidade média
- 300 passos em madeira a uma velocidade rápida
- 300 passos em madeira a uma velocidade lenta
- 300 passos em gravilha a uma velocidade média com irregularidades
- 300 passos em gravilha a uma velocidade média
- 300 passos em gravilha a uma velocidade rápida
- 300 passos em gravilha a uma velocidade lenta
- 300 passos em gravilha a uma velocidade média com irregularidades
- 300 passos em cimento a uma velocidade média
- 300 passos em cimento a uma velocidade rápida
- 300 passos em cimento a uma velocidade lenta
- 300 passos em cimento a uma velocidade média com irregularidades
- 300 passos em água a uma velocidade média
- 300 passos em água a uma velocidade rápida
- 300 passos em água a uma velocidade lenta
- 300 passos em água a uma velocidade média com irregularidades

O elevado número de amostras (um passo por segundo durante cinco minutos que perfaz um total de 300 passos por classe) potencia o sistema na medida em que minimiza as possibilidades de repetição do mesmo som quando

[28] Modo de andar e é definido como o padrão de movimento dos ombros dos animais, inclusive do homem, durante a sua locomoção em pisos sólidos

executados dois gestos de características muito semelhantes, sendo que uma pequena variação na descrição do áudio de entrada implica a escolha, por parte do sistema, de um segmento diferente da base de dados.

Os sons de passos que compõem a base de dados (cada um entre 250 milissegundos e o máximo de 1 segundo) foram gravados a uma frequência de amostragem de 44.1kHz e uma precisão de 16 bits com recurso ao software AudioStepsPro.

De referir o sinal do microfone foi sujeito a uma limitação de volume de forma a evitar distorções de amplitude, e a um processo de *gate* que tem por objectivo eliminação de *floor noise* da captação. O *threshold* do gate revelou-se um parâmetro fundamental para afinação da sensibilidade do DIGIT funcionando como um parâmetro que define o espaçamento mínimo entre passos.

4.2 Descrição e classificação do áudio de entrada e da base de dados

Após colecionar as várias gravações de passos procedi à descrição do conteúdo do áudio de cada passo na base de sons. Cada passo foi descrito pelos seis descritores definidos na secção 3.1.4: energia, MKL, centroíde logarítmico, desvio padrão espectral logarítmico, obliquidade espectral logarítmica e curtose espectral logarítmica. Esta análise foi configurada com uma janela de 2048 amostras e 25% de sobreposição de janelas de análise. A coleção de valores foi depois compilada numa lista e indexada ao segmento de som de passo respectivo de forma a que quando se chamar o ficheiro de som que irá ser cruzado com o perfil acústico este traga consigo a análise respectiva.

Para uma eficaz comparação entre os vários descritores usados no sistema, minimizando os efeitos impostos na comparação pela diferenças de escala dos vários descritores, procedi a normalização de cada descritor para média 0 e desvio padrão 1. (ver apêndice A.1 Patch de normalização).

Esta estratégia de normalização foi também aplicada ao áudio de entrada e efetuada numa fase de calibração do sistema. Durante a fase de calibração os valores passam por uma fase de normalização de forma a poderem ocupar um espaço mais próximo daquele que é ocupado pela base de sons possibilitando assim uma exploração mais uniforme da base de sons e permitir a adaptação do DIGIT ao perfil de cada utilizador no contexto de foley.

Figura 9: Fluxo de sinal para a análise do áudio de entrada

Seguidamente com recurso ao objecto **bonk**[29] (Puckette, Apel, & Zicarelli, 1998) foi realizado um módulo para detecção de ataques para reconhecimento por parte do DIGIT da entrada de sinal e fazer desta forma despoletar uma nova pesquisa na base de dados.

O sinal de entrada foi então analisado e foram retirados os descritores de energia, MKL, centroíde logarítmico, desvio padrão espectral logarítmico, obliquidade espectral logarítmica e curtose espectral logarítmica com uma janela de 2048 amostras e 25% de sobreposição de janelas de forma a fazer a análise em tempo real semelhante à que foi feita previamente à base de sons.

4.3 Seleção Automática dos Segmentos

Depois de normalizados os dados do áudio de entrada e da base de dados procedi à implementação do módulo de pesquisa automática. A pesquisa automática foi feita com recurso ao objecto externo Max **zsa.distance** (Malt & Jourdan, 2008).[30] Este objecto calcula de forma eficaz a distância Euclidiana entre o vector de

[29] crca.ucsd.edu/~msp, último acesso a 14 de Junho 2016

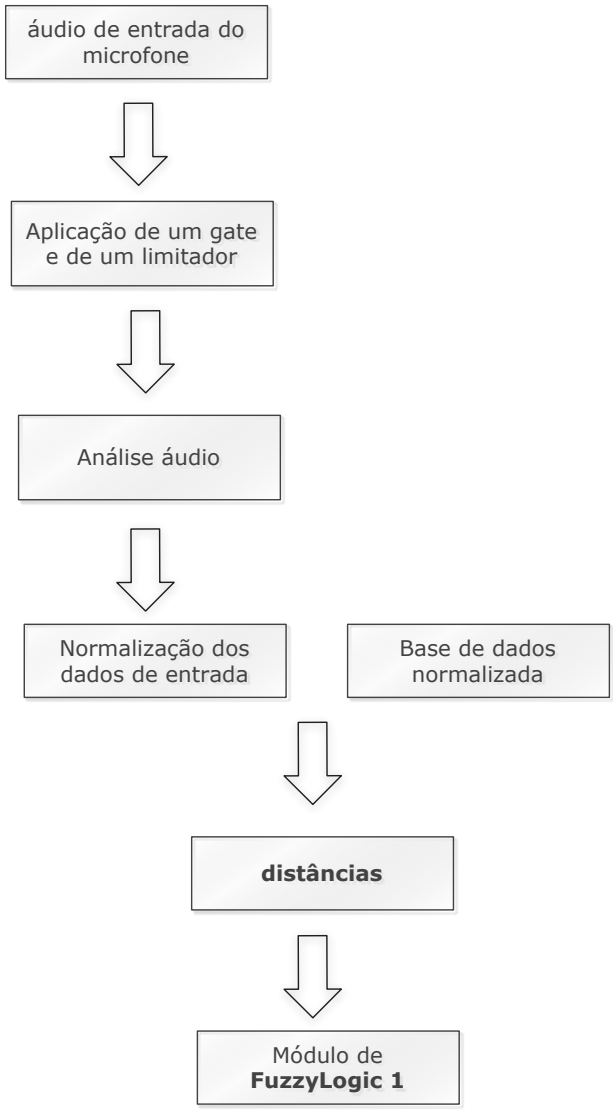
[30] http://www.e--j.com/index.php/what-is-zsa-descriptors/, último acesso a 27 de Junho 2016

características de entrada e os vectores de características que representam cada passo na base de sons. Assumo como vetor mais semelhante o que minimiza a distância em relação ao vetor de entrada.

Uma dificuldade inerente aos métodos de síntese concatenativa que baseiam em áudio de entrada é a da exploração de uma forma o mais uniforme possível da base de dados. Para minimizar esta limitação criei um módulo designado de **FuzzyLogic** (ver apêndice A.2 Patch FuzzyLogic) que tem como função aplicar uma penalidade a um segmento já escolhido e colocá-lo no fim da distância na lista dos próximos quatro candidatos, ou seja, depois de ser escolhido um segmento este passa para o fim de uma lista de quatro possíveis candidatos que mais se assemelhem com o sinal de entrada.

A introdução deste módulo permite que pequenas variações, ou subtilezas, na interação com a interface do DIGIT apresentam menos probabilidades de dar um resultado repetido aquando da escolha do segmento. Desta forma o DIGIT está menos sujeito a repetições que soam por vezes não naturais num contexto de foley, sendo que cada passo apesar de parecer igual ao anterior contém em si diferenças subtis as quais o espectador está muito atento pelas razões apresentadas no Capítulo 1.

Figura 10: Diagrama de comparação entre o áudio de entrada e a base de dados



4.4 Reprodução Áudio dos Segmentos da Base de Sons

Seguidamente à escolha do segmento na base de sons que mais se assemelha às características do perfil acústico do sinal de entrada, este elemento vai ser reproduzido por um motor granular. A opção por este sistema de reprodução áudio apresenta vantagens em relação a uma reprodução linear em três níveis (1) a possibilidade de reproduzir várias vozes (16 vozes no caso do DIGIT) do mesmo segmento, ou seja criar uma polifonia, (2) a alteração da velocidade de leitura do áudio por voz, o que permite uma alteração do pitch original do som, e a construção de som de passos mais complexos possibilitando a utilização do DIGIT em

ambientes de criação de passos não-reais, e ainda, (3) a possibilidade de poder escolher somente uma parte temporal do segmento que vai ser reproduzido. (ver apêndice A.3 Reprodução Áudio)

Tanto a velocidade de reprodução como o tamanho do segmento dispõem de controlos de âmbito, o que significa que as escolhas feitas não estão alocadas a um só valor. A cada nova detecção de entrada a escolha que o DIGIT faz é a de aleatoriamente adquirir um novo valor dentro do âmbito destes parâmetros definido pelo utilizador seguindo a metodologia proposta na síntese granular (Roads, 2004).

Ainda no sentido da ampliação das possibilidades do DIGIT foi desenvolvida uma duplicação das bases de sons com o intuito de criar aquilo que a minha investigação designa de sons compósitos, ou seja, oferecer a possibilidade de aceder a duas bases de sons de forma a poder trabalhar o som de passos por camadas, indo ao encontro de uma prática muito utilizada pelos designers de som como forma de dar mais "corpo" a um som e para que este se destaque mais na mistura final do áudio.

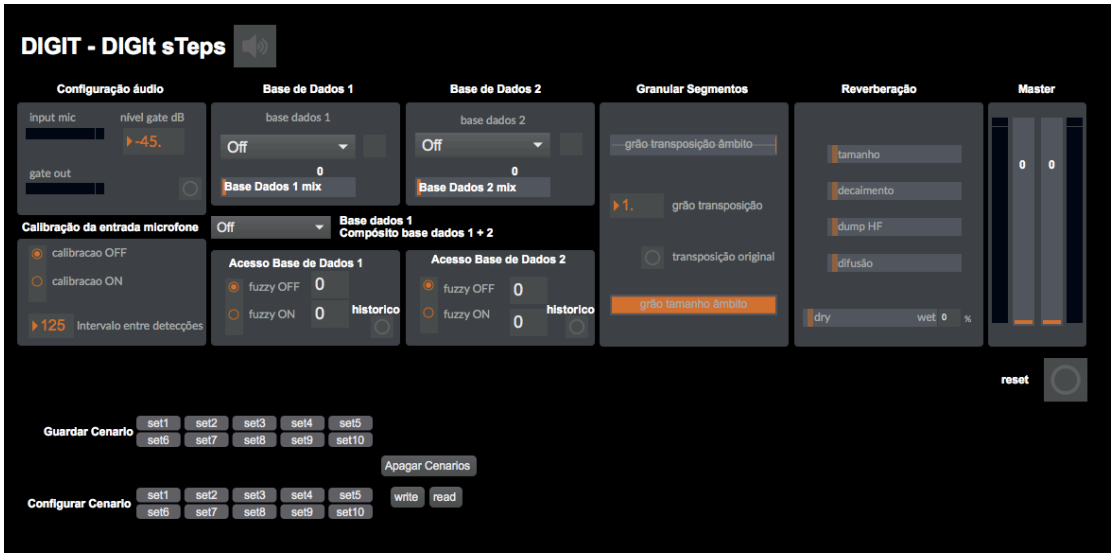
Com a duplicação da base de sons o DIGIT permite o acesso a duas camadas, sendo que uma pode ser composta e.g. som de passos em gravilha e outra em cimento, que podem ser misturados podendo desta forma, por um lado, criar som de passos que se destaquem mais na mistura final do áudio através da duplicação da energia dos passos que estão ser concretizados, e por outro lado, dar a possibilidade de construção de som de passos mais compostos, e.g. som de passos em cimento dão o ataque do som, e passos em gravilha dão a textura dos passos, aumentando assim a panóplia de possibilidades oferecidas pelo DIGIT.

Numa fase final da geração do áudio foi colocado uma reverberação tipo *plate*[31]. Trata-se de uma reverberação em estéreo que tem por fim a dar a possibilidade de poder explorar diferentes acústicas onde ocorrem os sons dos passos. Por fim, foi colocado um limitador de sinal para evitar distorções de amplitude à saída.

A figura seguinte representa o interface de utilização do DIGIT com os módulos organizados de acordo com o fluxo de sinal no sistema, de cima para baixo e da esquerda para a direita. Ainda no fundo do interface foi configurada uma secção onde é possível guardar memórias dos vários parâmetros de forma a poder utilizar em cenários futuros de pós produção essas mesmas configurações.

[31] *Plate reverb* ou reverberação por intermédio de uma chapa consiste placa de metal grande e fina suspensa numa moldura por molas. Um transdutor é colocado no centro da placa suspensa induz vibrações na placa e cujo e efeito é capturado por *pickups* (Roads, 1996)

Figura 11: Interface de utilização do DIGIT



O desenho da interface no DIGIT teve como objetivo a apresentação de uma forma clara todas as funcionalidades do sistema. Está organizado de acordo com o fluxo de sinal desde o reconhecimento do áudio de entrada até ao controlo de volume final. O DIGIT dispõe ainda da possibilidade de guardar o que designo de cenários para futuras utilizações. O que pretendi foi apresentar de uma forma clara as funcionalidades do DIGIT mantendo fora do âmbito do utilizador as configurações mais complexas através de uma automatização destas ao arranque da aplicação.

4.5 Resumo e conclusões

No sistema DIGIT pretendi abrir a possibilidade de inclusão do perfil acústico do gesto como forma de controlo de um sistema digital para a realização de foley. Pela descrição acústica do gesto e, através de uma pesquisa automática e exploratória de uma base de sons pré gravados, foi dada continuidade à premissa tradicional da exploração táctil de objetos por parte de um artista de foley como forma de conseguir determinados efeitos de som.

Para a realização do som de passos o paradigma da sua execução é alterado, literalmente dos pés para as mãos, e onde a realização do som de passos passa a ser operado através do controlo gestual. No entanto esta alteração de paradigma não foi feita sem desafios e dificuldades que se levantaram ao longo da

investigação: a escolha do interface, discutido no Capítulo 3, as particularidades da captação por intermédio de microfones de contacto, qual a taxonomia de descritores a serem utilizados pelo DIGIT, e permitir uma exploração gestual o mais abrangente possível da base de sons que compõe o DIGIT. Para a realização desta tarefa foi necessário recorrer à programação de módulos estatísticos cujos resultados serviram de base e refinamento à programação dos módulos de processamento de sinal.

Foi também criado um site https://sites.google.com/site/luisaly/DIGIT onde se encontra alojada uma versão aberta do DIGIT disponível para download, links para projetos que inspiraram o desenvolvimento do DIGIT e recursos de carácter informativo para o desenvolvimento do DIGIT. Pretendo que este site seja atualizado de forma regular e que persista por um prazo de dois anos. Qualquer alteração será notificada no site.

Os principais objectivos foram conseguidos e realizada uma prova de conceito na forma de uma aplicação funcional onde puderam ser testadas as premissas teóricas apresentadas pela investigação no Capítulo 2.

5 Avaliação do sistema DIGIT

5.1 Desenho dos testes de avaliação

A avaliação de design é uma questão recorrente na investigação. Em (Johnston, 2011) a avaliação não deve ser vista puramente como um exercício, mas antes como um estudo mais alargado entre o utilizador e as suas práticas criativas no contexto de utilização de uma nova aplicação. Johnston defende ainda em (2011) ser necessário alargar o âmbito daquilo que constitui a avaliação e reconhecer que apesar da ergonomia e a eficiência serem factores importantes no desenvolvimento de novos interfaces, estes não são os principais fatores determinantes da qualidade do interface a ser testado.

Jordà (2005) defende que uma avaliação de um interface humano-computador deve suportar interações em que os utilizadores possam ter múltiplas interpretações sobre a finalidade do interface. Desta forma a avaliação muda o seu foco para a identificação, coordenação, estímulo e análise de processos de interpretação em prática criativa.

Para avaliar o DIGIT foram realizados testes de usabilidade de sistemas e um questionário que pretende aferir a qualidade da experiência de utilização do DIGIT tendo em conta da utilização do sistema no contexto de foley.

Foi realizada também uma sessão de testes em ambiente prático de criação de efeitos de som para a jogos digitais, inserida na unidade curricular do presente ano de Design de Som do mestrado Multimédia na Faculdade de Engenharia do Porto. O DIGIT foi disponibilizado neste contexto e estudado por estes utilizadores para a realização de efeitos de som em contexto prático. A este grupo heterogéneo foi somente feito o questionário de avaliação qualitativa.

Os testes foram elaborados a partir da seguinte metodologia:

1) período de treino de cerca de 2 minutos, por parte do participante, na operação do DIGIT. Aqui foi veiculado ao participante informações sobre o funcionamento do DIGIT assim com a forma de interagir com a superfície que serve de interface;

2) realização de uma tarefa prática de sincronização do som de passos do DIGIT com uma sequência de imagens em movimento pré gravada e que consiste em dois cenários de piso (cimento e gravilha) com três velocidades cada (lenta, média e rápida) com o intuito de simular uma situação real de foley. Aos participantes foram

pedidas tarefas como *acompanha o movimento do personagem, acompanha as velocidades de movimento do personagem* ou *simula a mudança de direção do movimento do personagem.* De referir que a experiência foi levada a cabo com a utilização de auscultadores;

3) foi feita a gravação vídeo desta sessão de testes de forma a poder fazer uma análise posterior com o fim de registar depoimentos do participante durante a sessão

4) realização de um questionário sobre a usabilidade do sistema utilizando para isso a escala standard *system usability scale*[32] (SUS). Esta escala de Likert apresenta-se como uma forma de medição da usabilidade, e consiste num questionário com 10 perguntas e 5 opções de resposta desde Concordo Plenamente até Discordo Completamente cujo o resultado é calculado de acordo com uma fórmula específica. A classificação tem uma escala que vai dos 40 a +90 sendo entre 40 e 50 implica uma classificação **F**, entre 50 e 60 **D**, entre 60 e 75 **C**, entre 75 e 80 **D**, entre 80 e 90 **A**, e, finalmente, +90 **A+**. Trata-se de um questionário que permite avaliar uma grande variedade de produtos e serviços, incluindo hardware, software, dispositivos móveis, websites e aplicações;

Tabela 2: Adaptação do SUS aos propósitos de testes ao DIGIT

Penso que gostaria de utilizar o DIGIT:
Achei o DIGIT desnecessariamente complexo:
Achei o DIGIT de fácil utilização:
Penso que iria necessitar de suporte técnico de forma a poder utilizar o DIGIT:
Achei que as várias funcionalidades do DIGIT estavam bem integradas:
Penso que há muitas inconsistências no DIGIT:
Imagino que a maior parte das pessoas aprenderiam a usar o DIGIT:
Achei o DIGIT muito complexo e confuso
Senti-me confiante ao usar o DIGIT:
Necessitei de aprender muitas coisas antes de poder usar objectivamente o DIGIT:

5) realização de um questionário mais vocacionado para a aferição da qualidade da experiência em contexto criativo.

A recolha de dados teve como objetivos:

1) aferir as possibilidades da introdução de um sistema como o DIGIT em práticas de foley;

2) aferir o grau de expressividade dos resultados obtidos nas tarefas propostas;

[32] https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html, último acesso a 28 de Junho 2016

3) aferir a eficácia (como a relação entre o efeito obtido e os objectivos pretendidos) e eficiência (a relação entre a satisfação dos resultados e os recursos materiais e de tempo utilizados nas várias práticas) do DIGIT
4) comparação qualitativa com outros sistemas e práticas semelhantes

5.1 Resultados e Conclusões

Realizaram os testes e questionários um total de 8 participantes (2 do género feminino e 6 do género masculino), com idades compreendidas entre os 28 e os 42 anos. Três participantes afirmaram ter experiência na prática tradicional de foley, todos os participantes utilizaram bibliotecas de som de passos e cinco participantes afirmaram ter experiência na utilização de ferramentas digitais para a concretização de foley. Nenhum dos participantes afirmou ter problemas de escuta. Nenhum dos participantes nos testes foi pago.

Dos resultados obtidos no questionário[33] SUS sobre a usabilidade do sistema com recurso à escala standard SUS o sistema DIGIT obteve um resultado de 84.5, ou seja, uma classificação de A (Anexo A.4). Este resultado permite confirmar a estabilidade do sistema no que respeita à usabilidade.

Nos questionários qualitativos[34] foi aferida a qualidade da experiência do DIGIT. No que respeita à eficiência e eficácia do sistema foi a opinião geral de que o DIGIT é um sistema responsivo, de fácil utilização, de retorno imediato e de mapeamento gesto/resultado sonoro consistente. No entanto também forma apontados pormenores que poderiam melhorar o desempenho do DIGIT. Segue, algumas respostas dos participantes:
"*bastante fácil de usar o sistema e encadear a cadência dos passos" (Participante 4)*;
"*boa resposta dos sons faz com que estejamos a interagir coma superfície*, *retorno imediato de resultados, facilitou a exploração de expressividade" (Participante 1)*;
"Hit and miss, alguns resultados bons outros ... nem tanto" (Participante 2);
"Senti-me limitada em algumas abordagens menos convencionais, aqui talvez devido á interatividade, considerei interessante o alargar de possibilidade sensível o toque" (Participante 7);

[33] Resultados do SUS no anexo A

[34] Questionários qualitativos em anexo

"Tive alguma dificuldade inicial em perceber a relação entre a minha ação e o resultado de reprodução áudio. mas depois disto considerei um método muito eficaz" (Participante 7);
"Facilita imensamente o trabalho e temos o retorno imediato do resultado" (Participante 8);
"Não consegui recriar o arrastar do pé quer no cimento quer na gravilha" (Participante 8);
"O DIGIT é mais eficiente. Uma vez resolvidas algumas limitações de mapeamento tem o potencial para substituir o palco de foley em determinados projetos" (Participante 2);
"Facilitou a exploração da expressividade, principalmente pela velocidade do retorno e por uma forma que não é cansativa ... chega ser divertido!" (Participante 3).

No âmbito da realização dos testes forma também deixadas algumas sugestões por parte dos participantes:

"Analisar a correspondência entre o feedback háptico e expectativas de resultado áudio de forma a melhorar o processo de input" e *"Integração de geradores procedimentais de passos" (Participante 1);*
"Exploração de outras texturas de superfícies" (Participante 2);
"Arranjar uma forma de o sistema conseguir interpretar outras superfícies virtuais de forma a aumentar a acústica destas" (Participante 4).

6 Conclusões e Trabalho futuro

No Capítulo 1 fiz uma contextualização histórica da prática do *foley* de som atendendo aos recursos materiais e humanos necessários para a sua realização, assim como uma apresentação dos objectivos e metodologia aplicada da minha investigação.

No Capítulo 2 descrevi o estado-da-arte das várias áreas abordadas na minha investigação, a descrição do conceito de perfil acústico e apresentei o conjunto de descritores adoptados pela minha investigação. Descrevi os processos de síntese concatenativa e as suas várias possibilidades de controlo. Referi aplicações musicais de síntese concatenativa e outras de controlo de síntese sonora com base em interfaces tangíveis. Apresentei também as várias soluções digitais existentes para a realização de foley.

No Capítulo 3 apresentei o levantamento das várias possibilidades de materiais e formas que pudessem ser utilizadas como interface do DIGIT; estabeleci a definição de uma taxonomia de gestos que pudessem servir de amostra representativa das várias possibilidades de interação nas superfícies e a análise estatística das respostas acústicas de cada material quando sujeito à taxonomia de gestos definida para os testes às superfícies.

No Capítulo 4 descrevi a arquitetura do sistema DIGIT nas suas várias fases da implementação: (1) a descrição e classificação do áudio de entrada de acordo com os descritores mais relevantes, (2) pesquisa automática numa base de sons de passos, (3) reprodução dos segmentos selecionados de acordo com métodos de síntese concatenativa e, finalmente, (4) o processamento final do áudio por intermédio de efeitos como reverberação.

No Capítulo 5 foram descritos os testes experimentais realizados ao DIGIT e os métodos de avaliação usados.

No Capítulo 6 pretendo apresentar as conclusões sobre o trabalho realizado assim como perspectivas de trabalho futuro tendo como ponto de partida o trabalho desenvolvido no DIGIT.

No contexto da minha tese apresentei o sistema DIGIT para a realização de som de passos em contexto de foley com recurso a técnicas de síntese concatenativa tendo como meio de controlo a análise de um sinal de áudio. A

avaliação do sistema demonstrou que a abordagem aqui referida tem potencial para a realização de foley de som de passos. As possibilidades de interpretação de um perfil acústico revelou-se um dos elemento que mais relevância no DIGIT e, de acordo com os testes e questionários realizados, foi das particularidades do sistema que criou mais impacto nos participantes/utilizadores.

O facto de se recorrer a uma forma de expressão tão natural como o gesto numa superfície, prática corrente no foley tradicional, potenciou o DIGIT a superar as limitações apontadas em sistemas relacionados como o AudioSteps. Mais expressividade do que com recurso a um controlador MIDI o que permite que subtilezas de interação pudessem ser exploradas e com resultados que corresponderam em grande parte às expectativas dos utilizadores.

Um elemento crucial desta investigação foi a decisão numa fase preliminar de adoptar um microfone de contacto para capturar o perfil acústico de um gesto, mesmo consciente das limitações perceptuais que esta opção acarreta. Conclui que o facto de haver uma ligação háptica entre o gesto e o som que é produzido ajuda a diluir as diferenças entre o tipo de "escuta" de um microfone de contacto e o que os sistema auditivo humano percepciona, i.e. a escuta é complementada com o gesto

O DIGIT comprovou pela sua arquitetura ser capaz de reproduzir grande variabilidade sem no entanto recorrer a uma extensa base de dados de sons como em aplicações relacionadas. Apresenta-se como uma solução mais célere quando comparado com técnicas já existentes, principalmente em relação a práticas tradicionais de foley

O DIGIT apresenta possibilidades de expansão e adaptação a outros contextos de criação de efeitos de som a partir do momento que conceptualiza a exploração do gesto como condução de processos de síntese concatenativa. Portanto no contexto dos novos media criativos a relevância do tema e das questões que coloca julgo ser de todo o interesse da comunidade.

A presente investigação deixa uma série de questões em aberto que pretendo abordar no futuro em versões atualizadas da aplicação DIGIT. Entre o trabalho futuro, destacaria os quatro principais tópicos: (1) exploração de reconhecimento espacial do toque, por intermédio de técnicas de *Time Delay Of Arriving*, de forma a conduzir esta mesma localização para controlos de parâmetros de síntese, ou para a busca de um outro tipo de sons que componham a base de dados, e.g. som do passos a virar de direção ou de parar, (2) com recurso a técnicas de aprendizagem máquina poderá permitir um controlo mais previsível por parte do utilizador, na medida em que podem ser explorados diferentes sons em diferentes zonas da superfície de contacto, (3) recurso a técnicas de *audio fingerprint* que

permitam uma maior robustez na detecção de sinais hápticos, (4) a exploração de métodos de síntese procedimental onde o perfil acústico é conduzido para um modelo de síntese de passos como os modelos apresentados por Andy Farnell e (5) modelação física de superfícies de forma a aumentar a acústica virtual de superfícies reais dando a possibilidade de exploração de resultados diferentes.

Num contexto de criação de outros efeitos de som num contexto de foley que não somente som de passos, o DIGIT abre possibilidades de exploração de técnicas de aprendizagem máquina no sentido de criação de ambientes sonoros mais complexos e inverosímeis. Em contexto de recriação do som de uma tempestade em ambiente de pós produção poderemos associar um tipo de gesto estão a uma camada específica do design de som, e.g. som do vento, e um outro gesto poderá estar a controlar uma camada diferente, e.g. som dos trovões.

No contexto dos novos media criativos julgo que a relevância do tema que abordei na minha tese e as questões que foram colocadas e respondidas na investigação serem de todo o interesse da comunidade.

7 Bibliografia

Bernardes, G., Guedes, C., & Pennycook, B. (2012). EarGram : an Application for Interactive Exploration of Large Databases of Audio Snippets for Creative Purposes. *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR)*, (June), 265–277.

Bevilacqua, F., Zamborlin, B., Sypniewski, A., Schnell, N., Guédy, F., & Rasamimanana, N. (2009). Continuous realtime gesture following and recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5934 LNAI*, 73–84. doi:10.1007/978-3-642-12553-9_7

BOLBOACĂ, Sorana-Daniela; JANTSCHI, L. (2006). Pearson versus Spearman, Kendall's Tau Correlation Analysis on Structure-Activity Relationships of Biologic Active Compounds. *Leonardo Journal of Sciences*, (9), 179–200. Retrieved from http://ljs.academicdirect.ro/A09/179_200.pdf

Braun, A., Krepp, S., & Kuijper, A. (2015). Acoustic tracking of hand activities on surfaces. *Proceedings of the 2nd International Workshop on Sensor-Based Activity Recognition and Interaction - WOAR '15*, (JUNE), 1–5. doi:10.1145/2790044.2790052

Camic, P. M., & Camic, P. M. (2016). From Trashed to Treasured : A Grounded Theory Analysis of the Found Object From Trashed to Treasured : A Grounded Theory Analysis of the Found Object, (May). doi:10.1037/a0018429

Chion, M. (2009). *Film, a Sound Art*. Columbia University Press. Retrieved from https://books.google.com/books?id=_YWRPQAACAAJ&pgis=1

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *1*(2), 224–227. doi:10.1109/TPAMI.1979.4766909

De Sanctis, G., Rovetta, D., Sarti, a., Scarparo, G., & Tubaro, S. (2006). Localization of tactile interactions through TDOA analysis: Geometric vs. inversion-based method. *European Signal Processing Conference*, *2*(Eusipco), 2–5.

Designing Sound | The MIT Press. (n.d.). Retrieved October 14, 2015, from https://mitpress.mit.edu/books/designing-sound

Duindam, R., & Leeuw, H. (2015). Tingle : A Digital Music Controller Re-Capturing the Acoustic Instrument Experience, 3–6.

Françoise, J. (2015). Motion-Sound Mapping by Demonstration, (July). doi:10.13140/RG.2.1.5035.0248

Harker, A., & Tremblay, P. A. (2012). The HISSTools Impulse Response Toolbox: Convolution for the Masses. *Proceedings of the International Computer Music Conference*, 148–155. Retrieved from http://eprints.hud.ac.uk/14897/

Henrique, L. (n.d.). *Acustica Musical*.

Johnston, A. (2011). Beyond Evaluation: Linking Practice and Theory in New Musical

Interface Design. *Proceedings of the International Conference on New Interfaces for Musical Expression*, (30 May - 1 June 2011), 280–283. Retrieved from http://www.nime2011.org/proceedings/papers/G18-Johnston.pdf

Jordà, S. (2005). Digital Lutherie Crafting musical computers for new musics ' performance and improvisation. *Departament de Tecnologia*, *26*(3), 531. Retrieved from http://dialnet.unirioja.es/servlet/tesis?codigo=19509

Kaltenbrunner, M., Jordà, S., Geiger, G., & Alonso, M. (2006). The reacTable*: A collaborative musical instrument. *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE*, 406–411. doi:10.1109/WETICE.2006.68

Kim, H.-G., Moreau, N., & Sikora, T. (2006). *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*. Retrieved from https://books.google.com/books?hl=pt-PT&lr=&id=eQM_Ip2nN8YC&pgis=1

Landy, L. (1994). The "something to hold on to factor" in timbral composition. *Contemporary Music Review*, *10*(2), 49–60. doi:10.1080/07494469400640291

Malt, M., & Jourdan, E. (2008). Zsa.Descriptors: a library for real-time descriptors analysis. *5th Sound and Music Computing (SMC'08)*.

Monteiro, a. C. (2012). Criação e performance musical no contexto dos instrumentos musicais digitais.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO IST Project Report*, *54*(0), 1–25. doi:10.1234/12345678

Polotti, P., Sampietro, M., Milano, D.-P., & Vaudoise, H. E. (2005). Acoustic Localization of Tactile Interactions for the Development of Novel Tangible Interfaces. *8th International Conference on Digital Audio Effects*, (OCTOBER), 20–23.

Puckette, M. S., Apel, T., & Zicarelli, D. D. (1998). Real-time audio analysis tools for Pd and MSP. *Analysis*, *74*, 109–112. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.6961

Roads, C. (1996). *The Computer Music Tutorial*. Retrieved from https://books.google.com/books?hl=pt-PT&lr=&id=nZ-TetwzVcIC&pgis=1

Roads, C. (2004). *Microsound*. MIT Press. Retrieved from https://books.google.com/books?id=YxnqcAR7xjkC&pgis=1

Saitta, S., Raphael, B., & Smith, I. F. C. (2007). A bounded index for cluster validity. *Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition*, 174–187. doi:10.1007/978-3-540-73499-4_14

Savary, M., Pellerin, D., Cahen, R., Ateliers, E. Les, & Massin, F. (2013). Dirty Tangible Interfaces — Expressive Control of Computers with True Grit, 2991–2994.

Schaeffer, P. (1966). Traité des Objets Musicaux. Retrieved from http://philpapers.org/rec/SCHTDO-15

Schwarz, D. (2000). A system for data-driven concatenative sound synthesis. *Digital Audio*

Effects (DAFx), 97–102.

Schwarz, D. (2006). Concatenative Sound Synthesis: The Early Years. *Journal of New Music Research*, *35*(1), 3–22. doi:10.1080/09298210600696857

Schwarz, D. (2012). The Sound Space as Musical Instrument : Playing Corpus-Based Concatenative Synthesis. In *New Interfaces for Musical Expression (NIME)* (pp. 250–253).

Schwarz, D., Grégory, B., Bruno, V., & Sam, B. (2006). Real-time corpus-based concatenative synthesis with catart. *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06)*, (September), 1–7.

Schwarz, D., Tremblay, P. A., & Harker, A. (2014). Rich Contacts: Corpus-Based Convolution of Contact Interaction Sound for Enhanced Musical Expression. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 247–250. Retrieved from http://www.nime.org/proceedings/2014/nime2014_451.pdf

Tajadura-Jim, A., & Bianchi-Berthouze, enez and Nadia, E. Furfaro, F. B. (2015). Sonification of Surface Tapping Changes, 48–57.

Tomás, E., & Kaltenbrunner, M. (2014). Tangible Scores: Shaping the Inherent Instrument Score. *Proceedings of the International Conference on New Interfaces for Musical Expression*, 609–614. Retrieved from http://www.nime.org/proceedings/2014/nime2014_352.pdf

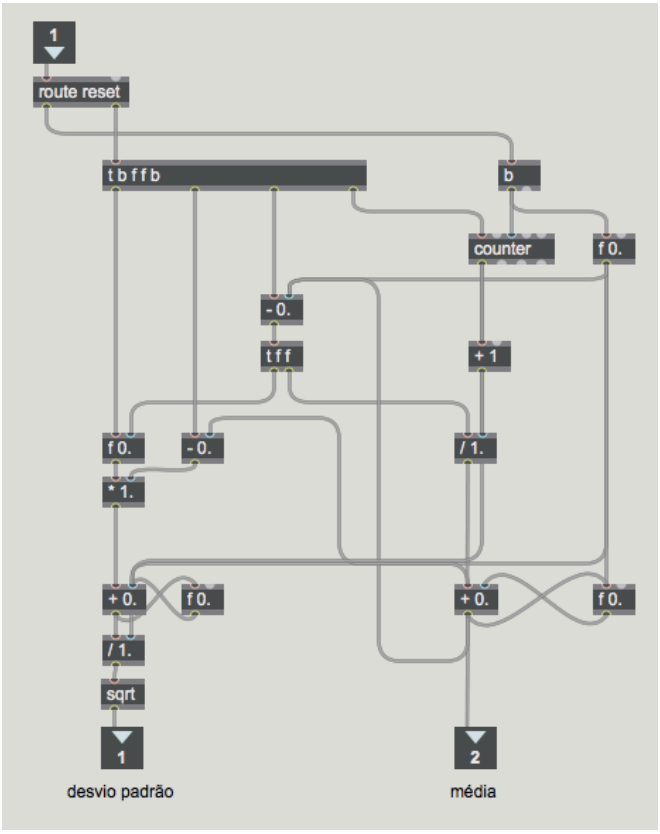
Uri, D., & Doyle, J. (2013). Subtlety of Sound : A Study of Foley Art.

Viers, R. (2011). *Sound Effects Bible*. Retrieved from https://books.google.com/books?hl=pt-PT&lr=&id=8mocQmOfPz4C&pgis=1

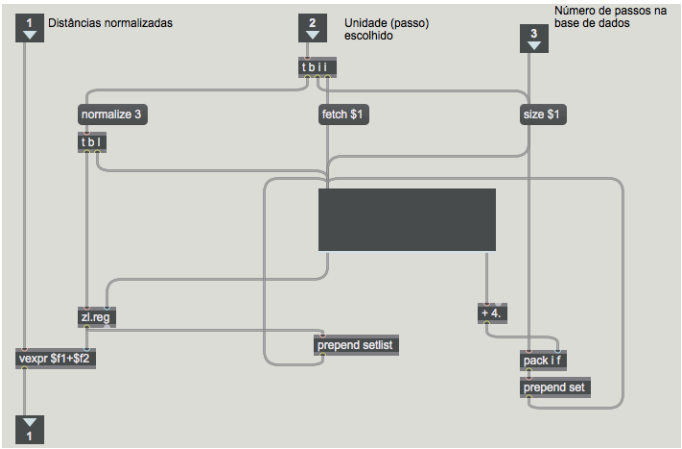
Yewdall, D. L. (2012). *The Practical Art of Motion Picture Sound*. Retrieved from https://books.google.com/books?hl=pt-PT&lr=&id=0nUqBgAAQBAJ&pgis=1

8 Anexo A

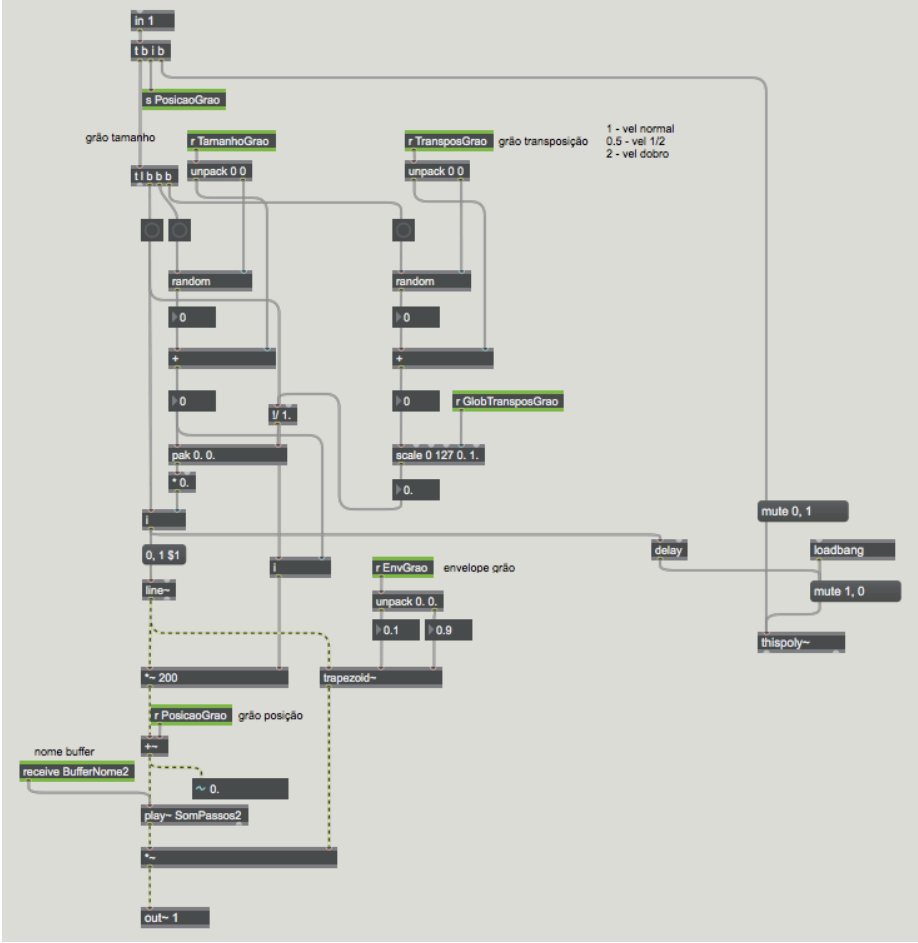
Anexo A.1 Patch de Normalização



Anexo A.2 Patch FuzzyLogic



Anexo A.3 Reprodução Áudio



Anexo A.4 Avaliações SUS

Participante 1		Participante 2		Participante 3		Participante 4		Participante 5		Participante 6		Participante 7		Participante 8		
UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	UserScore	SUS Score	SUS Score
4	3	5	4	5	4	5	4	4	3	4	3	4	3	4	3	
1	4	2	3	1	4	1	4	1	4	1	4	1	4	1	4	
5	4	4	3	5	4	5	4	4	3	4	3	3	2	4	3	
2	3	2	3	4	1	2	3	2	3	1	4	1	4	1	4	
3	2	3	2	3	2	5	4	5	4	5	4	3	2	3	2	
2	3	2	3	1	4	2	3	2	3	3	2	2	3	3	2	
4	3	5	4	5	4	4	3	4	3	4	3	2	1	4	3	
1	4	1	4	1	4	1	4	1	4	1	4	2	3	2	3	
4	3	3	2	5	4	4	3	4	3	4	3	3	2	4	3	
1	4	1	4	2	3	1	4	1	4	1	4	1	4	1	4	
	82.5		80		85		90		85		85		70		77.5	84.5