ISEL – Instituto Superior de Engenharia de Lisboa

ADEETC – Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores

LEIM

Licenciatura em Engenharia Informática e Multimédia

Unidade Curricular de Projeto

Localização de eventos sonoros em ambiente de cidade

Carlos Almeida 38415

Orientador(es)

Professor Joel Paulo

Professor Gonçalo Marques

Setembro, 2017

Resumo

O mundo caminha em direção às cidades inteligentes (Smart Cities). Nesse âmbito é útil o desenvolvimento de tecnologias que simulem os sentidos do Homem. Neste caso pretende-se simular o sentido de audição e como tal ser capaz de detetar a localização de eventos sonoros ocorridos em ambiente de cidade. Foram estudadas duas topologias na abordagem a esta questão. A de Ambisonic, em que o dispositivo contém 4 microfones unidirecionais todos centrados no mesmo ponto, e a topologia em triângulo, em que 3 (três) microfones estão localizados nos vértices dum triângulo equilátero. Em situações reais, o ambiente sonoro tem geralmente níveis de ruído de fundo significativos e podem ocorrer eventos sonoros simultâneos. Desta forma estudaram-se também para estudar algoritmos de redução de ruído, detecção e separação de eventos. Neste projeto é testado o uso de filtros de ruído para efetuar uma melhor detecção de eventos e desempenho nos algoritmos de localização. Ainda assim, verificou-se que estes filtros sendo úteis na parte de detecção de eventos não era muito relevantes na fase de localização.

Índice

Lista de Figuras

Capítulo 1

Introdução

Certas partes do mundo caminham neste momento para um futuro abundante no conceito das *Smart Cities* . Estas serão, idealmente, cidades inteligentes que, com recurso a tecnologias informáticas, criaram melhor qualidade de vida para os seus habitantes e visitantes.

Um fator importante das *Smart Cities* será a simulação, informática, dos sentidos das pessoas. Um pouco espalhado pelo mundo estão a ser desenvolvidos sistemas informáticos que simulam a nossa visão e audição. Esses sistemas processam a informação retida muito mais eficientemente do que o ser humano. Existem trabalhos e projetos desenvolvidos no âmbito da localização de eventos sonoros (singulares e múltiplos). Foram testados métodos que usam energia e tempos de atraso, para esse efeito. Foram, também, realizados trabalhos no âmbito da detecção de eventos, filtros de redução de ruído e separação de eventos.

Neste projeto pretende-se explorar o processamento de informação auditiva. Também um pouco espalhado pelo mundo, começaram-se já a criar sistemas de microfones que detectam e localizam eventos sonoros.

Este projeto pretende realizar a localização de eventos sonoros em ambiente urbano. Além deste fator são também objetivos deste projeto o desenvolvimento de algoritmos de detecção de eventos, redução de ruído e separação de eventos.

Este trabalho pode, entre outras coisas, medir o ruído existente num dado local a uma dada hora ou dia, detetar a quantidade de tráfego existente numa determinada rua, detetar um disparo duma arma, uma pessoa a gritar ou um acidente automóvel. Outro uso possível para este projeto seria o de usar a

informação da localização para redirecionar uma câmara de vídeo-vigilância.

Pretende-se, numa fase mais avançada, que este sistema possa estar ligado a uma central de bombeiros e polícia e que estes possam receber alertas para eventos sonoros mais relevantes para cada entidade.

Na totalidade, este projeto engloba duas arquiteturas que abordam este desafio. Foram feitos algoritmos e testes para uma topologia de 4 microfones unidirecionais, conceito Ambisonic, e uma outra topologia baseada em 3 (três) microfones dispostos em triângulo. A maior diferença entre estas duas abordagens é que a primeira aborda o cálculo a localização do evento sonoro baseado na energia dos sinais enquanto a segunda trata a localização com base em tempos de atraso, entre cada par de microfones.

Foram também realizados testes com um algoritmo de separação de eventos sonoros. Esta característica pode ser muito importante, numa fase mais avançada do projeto, para fazer classificação de eventos. Esta última fase iria pressupor o recurso a algoritmos de Aprendizagem Automática ou *Machine Learning* .

Neste projeto não será, para já, abordada a questão do cálculo da distância a que ocorrem os eventos. Essa secção do projeto será explorada em trabalho futuro como estará referenciada mais à frente neste relatório na própria secção de "Trabalho Futuro".

Este trabalho vai estar dividido em 4 (quatro) fases. Redução de ruído, detecção de eventos, localização de eventos e separação de eventos.

Para a fase de "Remoção de Ruído" foram estudados os algoritmos de Subtração Espectral e Filtro de Wiener. A Subtracção Espectral é mais rápida mas menos eficiente do que o Filtro de Wiener.

Para a fase de "Detecção de Eventos" foi estudado o algoritmo de *Spectral Flux* .

Para a fase de "Localização" foram estudados dois algoritmos. Um algoritmo baseado em energia dos sinais, para a arquitetura Ambisonic, e outro baseado em tempos de atraso, *gcc-phat* (*Generalized Cross Correlation with Phase Transform*) , 3.1.3 , para a arquitetura em triângulo. De notar que, das arquiteturas estudadas neste projeto (Ambisonic e arquitetura em triângulo) , ambos os algoritmos apenas funcionam nas respetivas arquiteturas às quais foram aplicados. Ou seja, não se poderia ter aplicado algoritmos baseados em tempos de atraso na arquitetura Ambisonic nem aplicar algoritmos baseados

na energia do sinal à arquitetura em triângulo.

Para a fase de "Separação de Eventos" foi aplicado um algoritmo de separação baseado no de *ICA* (*Independent Component Analysis*), 3.1.4 , neste caso o *fastICA* .

Todos estes algoritmos serão explicados e exemplificados mais à frente.

O projeto tem esta forma porque faz sentido reduzir o ruído para uma melhor detecção de eventos. Os algoritmos de localização apenas fazem sentido ser aplicados em trechos de áudio que de facto contenham algum evento sonoro. A separação é a última a ser processada pois o seu processamento *apaga* informação relevante necessário aos algoritmos de localização.

Capítulo 2

Trabalho Relacionado

2.1 Detecção de Eventos

No trabalho [3] foram efetuadas várias abordagens para verificar o desempenho diferentes fórmulas matemáticas no processo de detecção de eventos. Neste projeto é importante detetar em que momentos do sinal analisado se encontram os eventos sonoros relevantes e que zonas do sinal contêm apenas ruído. O trabalho [3] testa 5 (cinco) abordagens de detecção de eventos. Partindo do princípio que todos os eventos num dado sinal áudio são perceptíveis, após análise da variação de amplitudes ou frequências, podemos concluir que será sempre possível obter um gráfico que nos mostra claramente aonde estão os eventos e a sua duração. O projeto referenciado em [3] fala em 5 abordagens:

- Spectral Flux (Subtração de Espectros consecutivos)
- Phase Deviation (diferença de fases em amostras consecutivas)
- Complex Domain (esta fórmula verifica a variação tanto de amplitudes como fase)
- Weigth Phase Deviation (esta fórmula é semelhante a Phase Deviation mas com pesos mais elevados atribuídos às fases com maiores amplitudes)
- Rectified Complex Domain (esta fórmula aplica uma retificação de onda a formula de Complex Domain atribuindo o valor 0 (zero) para valores abaixo dum determinado *thresholds*)

Figura 2.1: Dispositivo Ambisonic

Estas fórmulas são posteriormente sujeitas a alisamentos e são-lhe aplicados *thresholds* para identificar as amostras que correspondem a eventos sonoros. Os *thresholds* aplicados não serão os melhores em todos os casos possíveis de diferentes eventos e ambiente. Será sempre preciso adaptar os *thresholds* e parâmetros da melhor forma possível para cada caso. Este projeto conclui que a melhor fórmula para detetar eventos com menor erro possível é, duma forma geral, a primeira fórmula Spectral Flux, ou seja, a subtração espectral de amostras consecutivas. Este algoritmo encontra-se descrito em 3.1 .

2.2 Estimação da Direção de Eventos Sonoros em Conceito Ambisonic

O trabalho relacionado [2] envolve o desenvolvimento de um algoritmo de localização de eventos. Foi estudado a informação do sinal em função do tempo e da frequência. Isso é possível através das *STFTs - Short Time Fourier Transform*. Estas *STFTs* efetuam várias transformadas de fourier ao longo do tempo. Desta forma obtemos uma matriz m x n em que m representa as janelas de tempo enquanto n as de frequência. A análise no espectro da frequência é o que permite a este estudo detetar mais do que um evento ao mesmo tempo.

O microfone Ambisonic, demonstrado na figura 2.1 , capta o som proveniente de todos os ângulos em 4 canais unidirecionais, cada um apontado para na sua direção especifica definidos na lista seguinte.

- RF (*right front*) representa o microfone *frente direita*
- LF (*left front*) representa o microfone *frente esquerda*
- RB (*Right Back*) representa o microfone *trás direita*
- LB (*left back*) representa o microfone *trás esquerda*

Estas fórmulas encontram-se explicadas na secção 3.1 .

2.3 Remoção de Ruído com Subtração Espectral

Um outro trabalho realizado anteriormente que foi estudado, é referido em [1] que se centra no desenvolvimento do algoritmo de Subtração Espectral, sendo que se acentua mais especificamente sobre o uso deste mesmo filtro em sinais de fala. Apesar desse facto, o algoritmo pode-se aplicar a outros tipos de sinais de áudio. Este filtro centra-se na subtração de magnitudes entre amostra de sinal e amostra de ruído.

O documento referido oferece um diagrama que simplifica bastante a percepção dos processos a seguir para construir este filtro de ruido (2.2).

Como se pode ver este diagrama mostra que usando um "bloco" de amostras de ruído e subtraindo a sua magnitude á magnitude do sinal este perde o ruido existente, sabendo sempre que o ruido nunca será subtraído na sua totalidade, mas que este algoritmo pode oferecer um filtro relativamente eficiente para o "peso" de computação. Este digrama parte do princípio que o utilizador poderá indicar ao computador o que é ruido. Ou seja, o filtro não é adaptativo e precisa que lhe seja indicado "manualmente" o que são amostras de ruído, à partida. Assim o algoritmo só tem de subtrair a magnitude a todas as amostras do sinal. Este mecanismo está descrito mais á frente na secção 3.2 . Aí serão explicadas as alterações a este algoritmo para que se adaptasse a amostra de ruído.

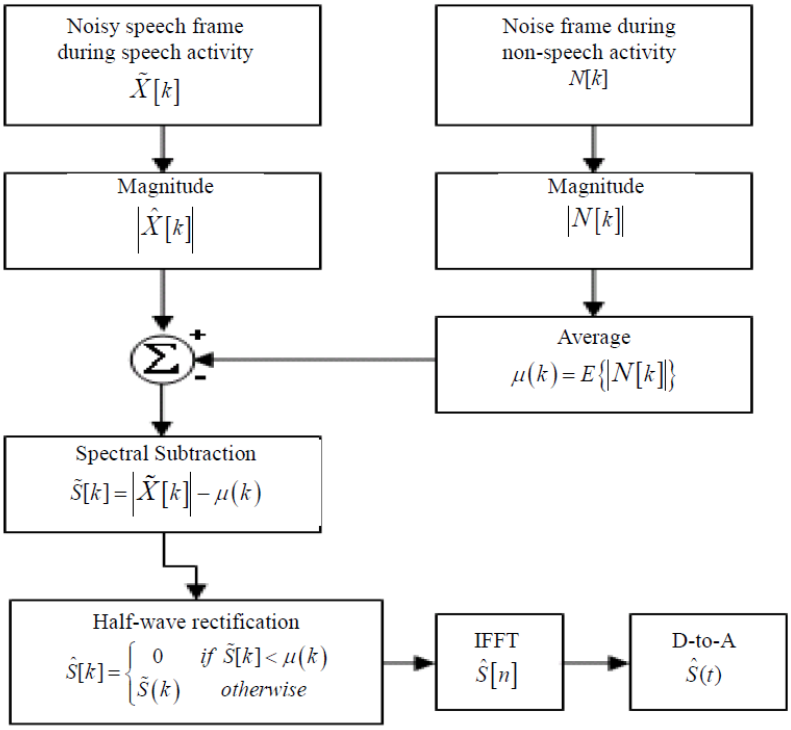


Figura 2.2: Diagrama de Subtração Espectral

Capítulo 3

Modelo Proposto

3.1 Fundamentos

Neste projeto foram estudados dois algoritmos de redução de ruído. O algoritmo de Subtração Espetral e uma abordagem ao filtro de Wiener. Contudo haveria uma falha nestes filtros se fossem aplicados, tal e qual a de definição original. Nenhuma destas abordagens se define como filtro adaptativo à partida. Isto significa que a amostra de ruído que é usada na redução do ruído ao longo de todo o sinal, não é atualizada com o decorrer de qualquer iteração durante o funcionamento do algoritmo. Então foi preciso alterá-los de forma que a remoção de ruído se adaptasse às variações de ruído de fundo ao longo de hora(s) ou dia(s), em ambiente urbano. Mais à frente irei explicar o conceito de ambos os filtros usados.

Para o funcionamento dos filtros de redução de ruído, assumi que os primeiros "blocos" de amostras, em cada ficheiro de teste, continham apenas ruído (sem eventos).

Quanto à detecção de eventos optei por usar o algoritmo *Spectral Flux* que revelou ter a eficiência necessária para o desenvolvimento do projeto. Após a detecção dos "trechos" de áudio que contêm eventos sonoros, estes "blocos" de amostras consecutivas são passados ao algoritmo de localização onde são identificados os ângulos dos eventos existentes. Estes "blocos" têm tamanho, em amostras, variável consoante as dimensões, dos eventos, retornada pelo algoritmo. Há um valor mínimo, de dimensão destes "blocos" de amostras, abaixo do qual os eventos são classificados como sendo apenas ruído. Isto é um passo necessário, pois acontecia com alguma regularidade que o algoritmo

retornasse um ou vários "pequenos eventos" que correspondiam a curtos momentos de maior ruído.

Em relação à localização de eventos para a arquitetura Ambisonic usei, como referido anteriormente, um algoritmo baseado nas diferentes energias dos sinais. Esse algoritmo está descrito nesta secção. Desenvolvi, também, uma fórmula diferente, da sugerida em 2 e referenciada em 3.13 , para calcular os múltiplos eventos que estará descrita na secção 3.2 .

Quanto à separação de eventos o algoritmo a ser usado foi o *fastICA* , explicado em 3.1.4 .

3.1.1 Filtros de Ruído

Subtração Espectral

Todos os filtros de ruído assumem que o sinal a ser processado consiste na soma de dois sinais diferentes. O ruído de fundo e o som do evento sonoro em si. Assim podemos definir uma fórmula de construção do sinal que nos irá ajudar a entender como reduzir o ruído existente no sinal.

$$x[n] = s[n] + r[n] \tag{3.1}$$

Em que x representa o sinal de áudio, r o ruído de fundo existente nos sinais de áudio captados e s representa o evento sonoro presente no trecho de áudio. Assim que conseguirmos definir a componente r podemos subtrair esta mesma ao sinal obtendo assim, apenas, a componente s . Subtração Espectral, por definição, é um algoritmo criado para ser aplicado em modo offline. Isto, porque o algoritmo necessita de saber que parte do sinal é apenas ruído. Com essa informação, é subtraída a magnitude do "bloco" de amostras, na frequência, de ruído a todas as *frames* do sinal. Desta forma foi necessário desenvolver este algoritmo de forma a que se adapte a cada momento de ruído no decorrer de todo o seu processamento. Na figura 2.2 é perceptível como funciona o algoritmo de subtração espectral. Como o esquema demonstra, é subtraído, ao sinal, uma amostra média de ruído. O "bloco" de amostras de ruído é "indicada" pelo utilizador. Após a subtração da amostra de ruído é feita uma transformada de *Fourier* Inversa *IFFT* e recuperado o sinal de áudio sem, ou com menos, ruído.

Filtro de Wiener

Este filtro é, como foi dito anteriormente, mais eficiente do que o algoritmo de Subtração Espectral mas é também mais "pesado" em termos computacionais. Dentro dos filtros de Wiener decidi usar a abordagem TSNR (Two Step Noise Reduction). Este conceito consiste em calcular o nível de SNR *a posteriori* (Signal to Noise Ratio) através da energia da *frame* atual e da amostra de ruído.

Este algoritmo assume que o primeiro "bloco" de amostras contém apenas ruído. Desta forma, e para o resto do seu processamento, usa esse mesmo "bloco" como a amostra de ruído para remover ao resto do sinal.

Após termos a nossa amostra de ruído, calcula-se o *SNR a posteriori* pela formula 3.2 .

$$SNR_{posteriori}(i) = (e(i-1)/r(i-1)) - 1; \tag{3.2}$$

Onde e representa a energia do sinal e r o ruído do sinal. Após este processo recorre-se a equação 3.3 para assegurar que o valor do *SNR* nunca é de facto igual a zero.

$$SNR_{posteriori}(n) = max(SNR_{posteriori}(n), 0.1) \tag{3.3}$$

Em que 0.1 é simplesmente um valor baixo próximo de zero mas não igual a zero. Isto serve para prevenir divisões por zero. De seguida calcula-se o *SNR a priori* através da fórmula em 3.4 .

$$SNR_{priori}(n) = \alpha * (e_{old}(n)/r_{samp}(n)) + (1-\alpha) * SNR_{posteriori}(n) \tag{3.4}$$

Em que e_{old} representa a energia calculada na iteração anterior.

$$\begin{cases} mag_x = |X| \\ mag_{new}(n) = (SNR_{priori}(n)/(SNR_{priori}(n)+1)) * mag_x(n); \end{cases} \tag{3.5}$$

Em 3.5 calcula-se a nova magnitude do sinal em que mag_x representa o módulo da *FFT* do sinal analisado na iteração corrente.

Em 3.6 calcula-se a atenuação do ruído sobre a energia do sinal, em que e_{new} define a energia calculada para a iteração corrente.

$$TSNR(n) = e_{new}(n)/r(n); \tag{3.6}$$

Em que r representa o ruído de fundo e $TSNR$ representa *two step noise reduction* .

$$\begin{cases} mag_{new}(n) = TSNR(n) * mag_x(n)) \\ fft_x(n) = mag_{new}(n) * \exp(i * phase_x) \end{cases} \tag{3.7}$$

Em que $phase_x$ representa a fase retirada da FFT ao sinal a ser analisado nesta interação.

O resultado obtido em 3.7 é somado a um array, o qual no final representará o sinal, na sua totalidade, com ruído reduzido.

3.1.2 Detecção de Eventos

Neste projeto foi usado, para detecção de eventos, o algoritmo de *SpectralFlux* mencionado em 2.1 . Assim, depois de calculadas a $STFT$ (*Short-Time Fourier Transform*) é aplicada a subtração da *frame* atual com a seguinte, isto para todas as frequências. É feita a soma de todos estes valores. Após este passo obtemos um gráfico com picos de amplitude correspondentes aos eventos sonoros, como mostra a figura **??**. Este gráfico é resultante da análise a um ficheiro de áudio que continha eventos sonoros de cães a ladrar.

O algoritmo *Spectral Flux* consiste em medir a mudança de amplitudes em cada janela de frequência como mostra a equação 3.8 .

$$SF(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|) \tag{3.8}$$

Em que H indica o processo demonstrado na equação 3.9 .

$$H(n) = \frac{SF(n) + |SF(n)|}{2} \tag{3.9}$$

$$g_\alpha(n) = max(f(n), \alpha g_\alpha(n-1) + (1-\alpha)f(n)) \tag{3.10}$$

Após o processamento efetuado em 3.9 ainda existe muito ruído para poderem ser detetados eventos sonoros eficientemente. Assim a função calculada em 3.9 passa por uma equação (3.10) de "alisamento" ou *smoothing* .

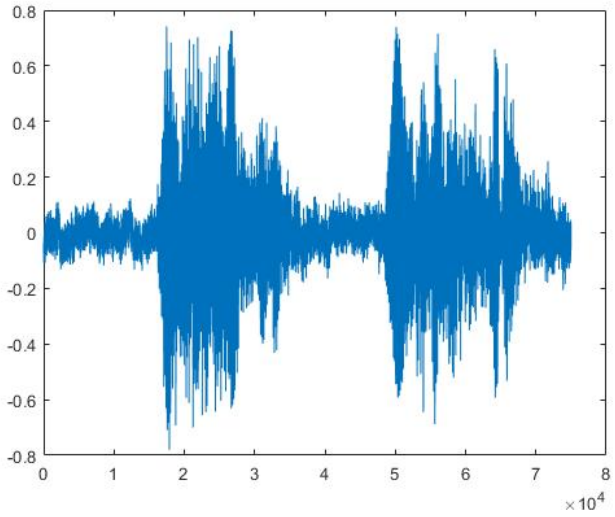


Figura 3.1: Sinal de áudio com dois eventos ("cães a ladrar")

Na figura 3.1 temos um sinal de áudio com dois evenos. A figura 3.2 mostra o resultado após os cálculos descritos nas equações 3.8 e 3.9 . Este gráfico não apresenta, exatamente definidos, todos os eventos apresentando muitas irregularidades. Seria difícil aplicar qualquer tipo de *thresholds* a este gráfico. Então esta lista de valores é "alisada" pela função expressa em 3.10 , que soma parte do "bloco" de amostras atual com o anterior. Este último tem sempre um *peso* muito elevado. O valor de α nessa mesma expressão deve ter valores entre]0, 1[e propõe-se que tenha valores perto de 1 para um melhor desempenho como filtro passa-baixo. Esta função permite-me obter um gráfico com variações mais suaves e definir mais claramente cada evento e a sua duração. Sabendo em que zona do sinal analisado se encontram os eventos sonoros que queremos localizar, passa a não ser preciso analisar todo o sinal. Assim elimina-se, ou tenta-se eliminar o máximo possível, os momentos de apenas ruído, que geraria resultados aleatórios no processo de localização.

Desta forma o conjunto de *frames* a ser analisada, em cada momento, adapta-se à dimensão do evento, sendo que existe um limite mínimo para a dimensão desta janela de análise. Assim se o evento detetado tiver duração (em número de amostras) abaixo dum determinado *threshold* este será desprezado, pois será interpretado como ruído de fundo. Assim os eventos que têm

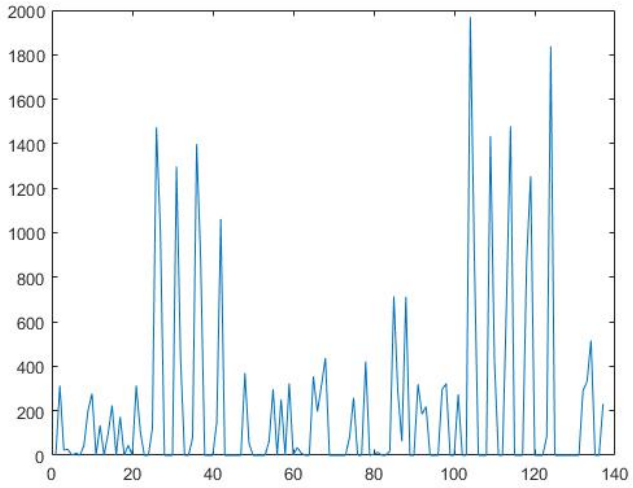
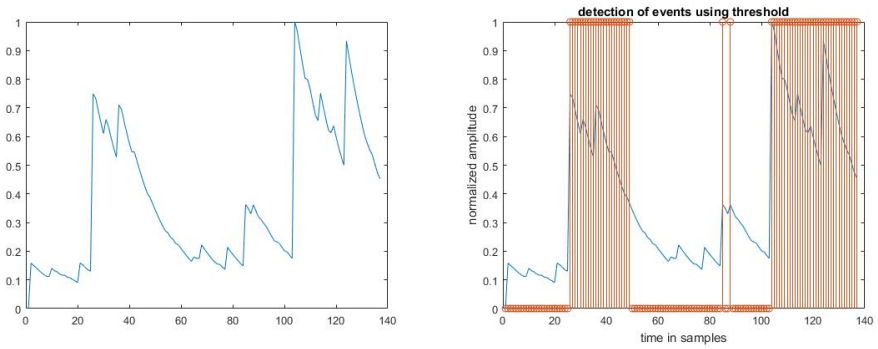


Figura 3.2: Detecção de Eventos - *Spectral Flux*

dimensão inferior a 30 milissegundos são eliminados. À partida, não haverá nenhum evento de dimensão no tempo inferior a esse *threshold*. Na figura 3.3 temos, à esquerda, a subtração de espectros consecutivos após um filtro passa-baixo, à direita, o gráfico que mostra, com linhas verticais laranja, os momentos dos eventos sonoros, sendo que no restante sinal de áudio é assumido que só existe ruído de fundo.



(a) Passo 2 (b) Passo 3

Figura 3.3: Detecção de Eventos

3.1.3 Localização de Eventos Sonoros

Conceito Ambisonic (baseado em Energia do Sinal)

O excerto de sinal que chega a esta parte do algoritmo é apenas um determinado "bloco" de amostras consecutivas. Apesar da a arquitetura Ambisonic permitir a localização do ângulo num espaço 3D, neste projeto apenas será usada a localização no plano horizontal, assumindo desta forma que geralmente todos os eventos sonoros ocorrem ao mesmo nível vertical.

Assim, neste projeto a formula de detecção do ângulo do evento propõe-se a analisar o sinal em função do tempo e da frequência. Podemos assumir que em cada janela de tempo e de frequência apenas um, dos eventos em simultâneo, sobressai. Desta forma podemos identificar todos os múltiplos eventos em simultâneo analisando todas as janelas de tempo e frequência, pois todos os eventos irão sobressair numa determinada janela.

Neste projeto também não será abordada a questão do cálculo da distância a que o evento ocorreu. Para a arquitetura Ambisonic seria preciso ter pelo menos dois destes aparelhos a funcionar em simultâneo para ser possível calcular a distância à qual ocorreu um determinado evento. Assim, usando um *array* de microfones ao longo dum local (conforme é pretendido para este projeto no futuro) poder-se-á contornar este problema.

Os quatro canais do aparelho Ambisonic (4 microfones) são processados, em 3.11 , para se obter sinais de energia nos respetivos eixos de referenciais, x , y e z , enquanto w representa a soma de todas as energias , ou seja representa um sinal omnidirecional.

$$\begin{cases} a(t) = 0.5 * ((LF - LB) + (RF - RB)) \\ b(t) = 0.5 * ((LF - RB) - (RF - LB)) \\ c(t) = 0.5 * ((LF - LB) + (RF - RB)) \\ d(t) = 0.5 * ((LF + LB) + (RF + RB)) \end{cases} \tag{3.11}$$

$$\begin{cases} Ix = Real(D^* * A) \\ Iy = Real(D^* * B) \end{cases} \tag{3.12}$$

Em 3.12 são calculadas as matrizes Ix e Iy , que contêm os vectores de energia em cada eixo. O sinal a contem a informação no eixo dos XX enquanto b contém informação no eixo dos YY. Nessa mesma equação, $D*$ é o conjugado de D e A e B são as matrizes dos sinais a e b após $STFT$s.

Por sua vez, estas matrizes permitem-nos calcular os ângulos para cada janela de frequência e para cada janela de tempo, obtendo uma matriz de mXn de ângulos, em que m representa as janelas de frequência e n representa as janelas de tempo. A equação 3.13 ilustra esse mesmo processo.

$$\begin{cases} \alpha = tan^{-1}(\frac{-Iy}{-Ix}) & \text{se Iy} \geq 0 \\ \alpha = tan^{-1}(\frac{-Iy}{-Ix}) - 180 & \text{se Iy} \leq 0 \end{cases} \tag{3.13}$$

Esta matriz de ângulos em função da frequência e tempo permite-nos encontrar múltiplos eventos em simultâneo.

Nas equações 3.14 e 3.15 as fórmulas para achar o(s) ângulo(s) resultante(s). A primeira foi estudada num trabalho relacionado descrito em 2.2 enquanto a segunda foi desenvolvida no presente projeto.

$$N(m,n) = \sum_{n=1,m=1}^{n=N,m=M} p[\alpha(n,m)|\theta] \tag{3.14}$$

$$\theta(m,n) = \alpha[Iy(m,n) + Ix(m,n) * i] \tag{3.15}$$

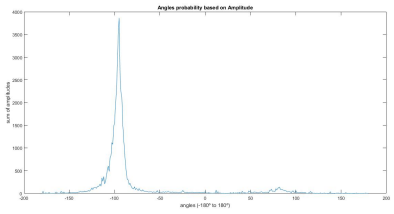
Na equação 3.15 , α representa o ângulo calculado, através do respetivo número imaginário. Ix é a matriz de intensidades para o eixo dos XX . Em 3.12 , D^* é o conjugado da $STFT$ de D e A é corresponde à $STFT$ de a . A segunda fórmula é respetiva ao eixo dos YY.

Com esta matriz de ângulos é possível então identificar o ângulo do evento a ser analisado. Haviam duas abordagens iniciais para estes cálculos. Calcular a probabilidade de um ângulo através dum histograma com os 360 graus ou calcular a probabilidade do mesmo baseado na soma das amplitudes desses ângulos.

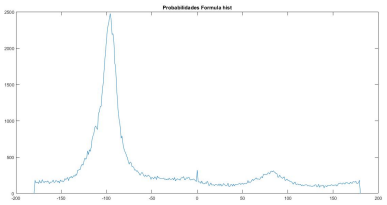
$$\begin{cases} h(i) = h(i) + 1 \\ SAmp(i) = SAmp(i) + W(i) \end{cases} \tag{3.16}$$

Em que h representa um histograma e $SAmp$ uma fórmula, desenvolvida neste projeto, que faz a soma de amplitudes de D ($STFT$ do sinal d calculado em 3.11)

Como se pode verificar, o histograma baseado na soma de amplitudes tem, consideravelmente, menos ruído do que o histograma "normal" . Desta forma,



(a) Soma das Amplitudes



(b) Histograma

Figura 3.4: Comparação de gráficos de ângulos para 'histograma' de amplitudes e histograma

neste projeto passou a ser usado a primeira opção em vez do histograma "normal" sugerido em 2.2 e noutros artigos relacionados semelhantes a este.

Ocasionalmente surgiu "picos" no gráfico que identificavam ângulos falsos. Ou seja, introduziam falsos positivos na detecção de ângulos. Isto poderia acontecer por introdução duma componente de ruído de elevada amplitude na *frame*.

O gráfico resultante de todas estas fórmulas contem bastante ruído e, por isso, é passado por um método que reduz algum ruído existente no gráfico.

Os picos detetados têm, assim, diferentes características que lhes permitem ser ignorados ou realçados. Se um pico for muito elevado, por comparação com os restantes, esse passa a ter mais relevo na identificação do ângulo de onde vem o evento. Nunca poderemos saber ao certo quantos eventos estão a ocorrer num determinado local a um determinado momento, por isso, não podemos descartar nenhum evento que se situe acima do *threshold*. Este *threshold* é definido pelo utilizador de acordo com o que for mais indicado para cada ambiente. Neste caso o *threshold* teve um determinado valor para os ambientes específicos de cidade em que os ficheiros foram gravados. Este mesmo poderia ter de ser ajustado para qualquer outro tipo de ambiente.

O desafio, nesta fase do projeto, passa por identificar quais os "picos" no gráfico que representam, de facto, eventos sonoros e aqueles que representam apenas ruído que por alguma razão sobressaiu naquele exato momento. Aqui definir o *threshold* trará consequências na localização dos eventos. Um *threshold* demasiado elevado pode acabar por neutralizar eventos menos sonantes, mas importantes ainda assim, enquanto um *threshold* demasiado baixo poderá criar eventos falsos.

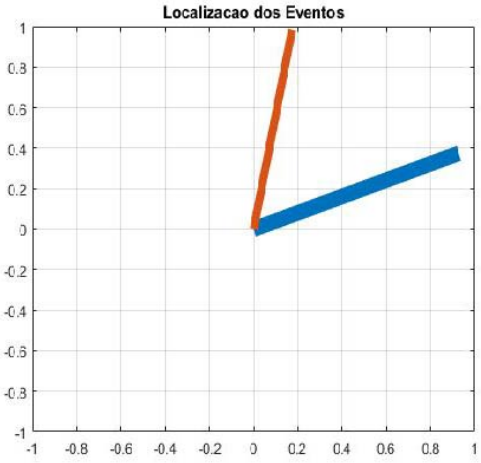


Figura 3.5: plano horizontal com dois eventos simultâneos

Após a detecção dos "picos" relevantes para os nossos casos de teste, os resultados eram apresentados num gráfico como ilustrado na figura 3.5 . Nesta figura o centro representa o ponto de coordenadas (0,0) num plano horizontal. Aqui existem dois traços que representam as direções de dois eventos sonoros detetados em simultâneo. As diferentes espessuras representam um evento mais ou menos sonante, em que quanto mais espesso o traço mais sonante terá sido o evento.

Arquitetura em Triângulo (Baseado em Atrasos do Sinal)

Para esta arquitetura, o algoritmo de localização desenvolvido foi o *GCC-PHAT* . Este algoritmo baseia-se na correlação entre dois sinais no tempo e na frequência. Esta será usada numa fórmula onde serão aplicados atrasos/adiantamentos de tempo que definirão um gráfico com um ou mais picos que identificam tempos de atraso presentes no respetivo par de sinais. Estes tempos de atraso são medidos em amostras. Com essas amostras e com a frequência de amostragem é possível identificar o ângulo no qual incidiu a onda sonora.

O primeiro passo é efetuar uma STFT (*Short Time Fourier Transform*) a cada um dos sinais. De seguida calcula-se a correlação entre os dois sinais através da expressão em 3.17 .

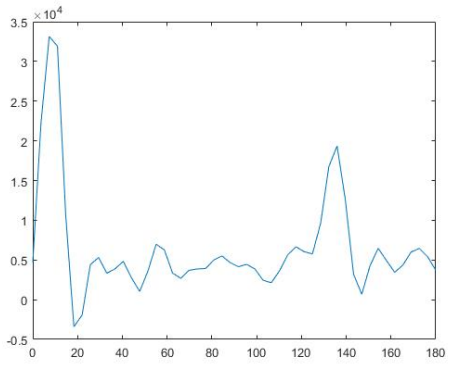


Figura 3.6: Correlação em função dos Tempos de Atraso

$$R_{xx}(t,f) = X(t,f) * Y(t,f)^H \tag{3.17}$$

E por último adiciona-se a fórmula que nos permite desenhar a correlação entre sinais ao longo dos diferentes tempos de atraso.

$$\phi^{gcc}(t,f,m) = \frac{R_{xx}(t,f)}{|R_{xx}(t,f)|} e^{-2\pi i f \tau(m)} \tag{3.18}$$

As expressões exemplificadas em 3.17 e 3.18 ocorrem uma vez por cada par de microfones. Sendo que nesta arquitetura existem 3 microfones, estas expressões ocorrem 3 vezes. Correlacionando desta forma os 3 pares de microfones.

Após este processamento obtemos uma matriz de dimensões t, f e m (t - tempo, f - frequência, m - tempos de atraso) que foi usada somando todos os resultados para as duas primeiras dimensões obtendo um vector final de dimensão 1xm. Se fizer o gráfico desta lista de valores será possível identificar os tempos de atrasos com maior correlação entre cada par de sinais, como mostrado na figura 3.6.

Sendo que a topologia é de 3 microfones separados em forma de triângulo, cada par estará definido para o seu próprio referencial. Desta forma tive de ajustar os referenciais de cada par de microfones todos para a mesma disposição de referencial (figura 3.7). Assim os referenciais passaram a estar alinhados com o referencial indicado com a letra “A”.

Após ajustar os 3 referenciais posso sobrepor os 3, respetivos, gráficos de pares de microfones e visualizar onde está a maior incidência de “picos”.

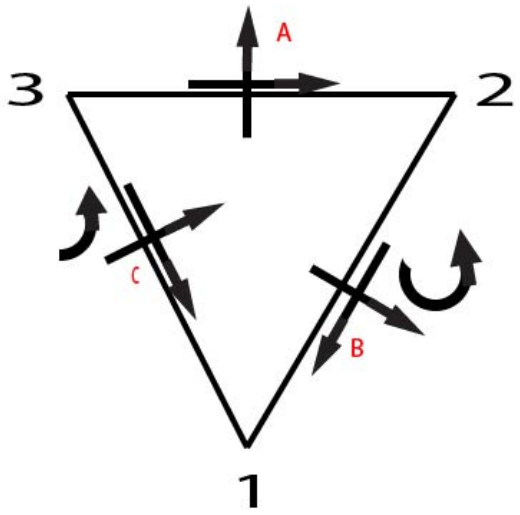


Figura 3.7: Ajuste de referenciais

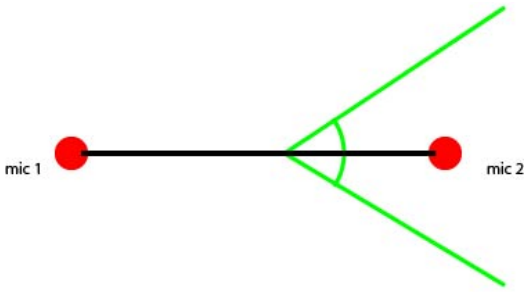


Figura 3.8: Ambiguidade de Ângulos em Pares de Microfones

Temos de ter em atenção os "picos" que representam ângulos falsos. Isto acontece porque cada par de microfones tem a capacidade de identificar o ângulo a que incidiu o sinal, mas não permite saber se veio da frente ou de trás dos microfones. A figura 3.8 ilustra essa particularidade desta topologia.

Esta ambiguidade é desfeita ao verificarmos os 3 pares de microfones, pois tempos de atraso, ou eventos, falsos estarão presentes apenas num dos gráficos. Os eventos sonoros, que sejam de facto reais, estarão ilustrados como "picos" dos 3 gráficos.

3.1.4 Separação de Eventos por ICA

Neste projeto foi abordada a temática da separação de eventos sonoros. Este tipo de algoritmos é bastante relevante para este projeto pois será importante,

em trabalho futuro, proceder à classificação automática de eventos. Desta forma é possível identificar programaticamente que tipo de evento sonoro ocorrera naquele preciso momento. Podem ser eventos variados como um acidente de viação, um disparo duma arma, cães a ladrar, sirenes, etc. Para uma classificação mais exata é necessária a separação dos eventos sonoros, pois se estiverem sobrepostos, mesmo que apenas sobreposto por ruído, são bastante mais difíceis de classificar.

Neste projeto, apenas foi usado o algoritmo *fastICA* que é uma versão de execução mais rápida em relação ao conceito básico do algoritmo de *ICA*.

Para termos noção de como funcionam os processos de mistura e separação estes processos estão exemplificados nas expressões 3.19 . A matriz A é duma mistura que simula a o que ocorre em ambiente real aquando da ocorrência de eventos sonoros pelos diferentes microfones em simultâneo. A matriz W representa os "pesos" que serão aplicados aos diferentes sinais (3 no caso da arquitetura em triângulo e 4 na de Ambisonic) para que ocorra a separação de cada evento sonoro ocorrido.

Importante salientar que o número de eventos simultâneos separados nunca será superior ao de sinais existentes. Desta forma, para a arquitetura Ambisonic seriam apenas possíveis de separar até 4 eventos em simultâneo enquanto que na arquitetura de triângulo seriam 3.

$$\begin{cases} y = Ax \\ x = W^T y \end{cases} \tag{3.19}$$

Este algoritmo, *fastICA* , consiste em 2 passos. Primeiro há uma centralização de dados seguido dum processo de *whitening*.

$$y_{i,j} = y_{i,j} - \frac{1}{M} \sum_{j'} y_{i,j'} \tag{3.20}$$

Com esta "centralização de dados" , referenciada pela expressão 3.20 , pretende-se que a média total de cada linha da matriz X tenha o valor esperado de 0.

Ainda como parte do primeiro passo, procede-se a fase de *whitening*. Esta fase pretende que as amostras de X sejam não correlacionadas entre si e tenham variância igual a 1, através da expressão em 3.21 .

$$WT = E * Diag^{-\frac{1}{2}} E^T X \tag{3.21}$$

Onde E define a matriz de *eigenvectors* e *Diag* a matriz diagonal

O segundo passo consiste em separar de facto os evento sonoros. Assim é repetido um determinado processo o número de vezes que for definido até termos a quantidade de eventos desejada. O processo de separação ocorre após processamento duma equação que vai obtendo valores para *pesos* que serão, no seguimento, aplicados ao sinal original. Enquanto esse conjunto de *pesos* continuar a retornar diferentes entre cada iteração seguida o processamento continua. A equação 3.22 demonstra como se obtém a matriz de *pesos*, W.

$$\begin{cases} W_p = \frac{1}{M} X_g (W_p^T X)^T - \frac{1}{M} g^{'}(W_p^T X) 1 W_p \\ W_p = W_p - \sum_{j=1}^{p-1} W_p^T w_j w_j \\ W_p = \frac{W_p}{|W_p|} \end{cases} \tag{3.22}$$

Como disse, este processo é repetido o número de vezes que for preciso para separar todos os eventos sonoros pretendidos, tendo em conta que se uma matriz tem n misturas então o *ICA* ou *fastICA* só poderão separar um máximo de n eventos sonoros.

Após o processo ilustrado em 3.22 falta apenas aplicar os ditos *pesos*, w, ao sinal original como demonstrado em 3.23.

$$\begin{cases} W = [w_1, w_2, ..., w_n] \\ S = W^{[}T] X \end{cases} \tag{3.23}$$

3.2 Abordagem

Este projeto, que na verdade se dividiu em dois, ao longo de dois anos, centrou-se sobretudo na localização onde passei de facto a maior parte do meu tempo de trabalho.

Aqui apliquei diferentes fórmulas. Na topologia Ambisonic segui as indicadas inicialmente pelo documento 2 .

$$\begin{cases} Ix = real(D^*. * A) \\ Iy = real(D^*. * B) \end{cases} \tag{3.24}$$

Em que W representa a $STFT$ do sinal w calculado em 3.11 . O documento referia também as seguintes fórmulas, para cálculo de ângulo.

$$\alpha(t,f) = \begin{cases} tan^{-1}(\frac{-I_y(t,f)}{-I_x(t,f)}) & I_y(t,f) > 0 \\ tan^{-1}(\frac{-I_y(t,f)}{-I_x(t,f)}) - 180 & I_y(t,f) <= 0 \end{cases} \tag{3.25}$$

Nesta fase alterei a equação 3.25 para a equação em 3.26.

$$azimuth(t,f) = \theta(I_x(t,f)i * I_y(t,f)) \tag{3.26}$$

Em que θ identifica a função de cálculo de ângulo através de um número imaginário.

Esta formula retorna uma matriz de tamanho dimFrequência$x$ dimTempo. Esta é composta por ângulos calculados para todas essas mesmas janelas de frequência e tempo. Assim, após o uso desta fórmula calculei os histogramas. Histograma é o método sugerida pelo documento referenciado em 2.2 .

Aqui, eu alterei ligeiramente a fórmula e em vez de somar o número de vezes de ocorrência de um ângulo, somei as amplitudes respetivas da matriz de energia total D que representa o sinal d após $STFT$, na equação 3.11 . Desta forma fui somando, a cada iteração, a amplitude 3.16 respetiva a cada ângulo em 3.26.

Em relação ao filtro de ruído de Wiener também fiz algumas alterações. É importante que a amostra de ruído vá sendo alternada e adaptando conforme as "flutuações" do ruído de fundo presente numa cidade, ao longo de um dia inteiro ou mesmo de semanas ou meses.

Assim, para este projeto foi desenvolvida uma forma de atualizar a amostra de ruído, ao longo do tempo, de acordo com o ruído verificado em cada preciso momento. É feita uma média do módulo de todas as FFT calculadas para o trecho de áudio da iteração corrente. Após fazer isto para a primeira amostra comparo este ruído com o sinal proveniente das iterações seguintes. Se essa média estiver entre um determinado *threshold*, ou limiar, em relação à amostra de ruído então assiná-lo a amostra corrente como se tratando de ruído. Aí adapto "suavemente" o o ruído a esta nova amostra de ruído através de um determinado *peso* como demonstrado em 3.27 .

$$r = \alpha * r + (1 - \alpha) * m_{|X|}; \tag{3.27}$$

Em que r representa a componente de ruído e m a média. O parâmetro α representa um *peso* entre 0 e 1 , não inclusive, e será ajustado pelo utilizador conforme desejado para um melhor desempenho.

3.3 Implementação do Modelo

Em termos tecnológicos usei o *Matlab* assim como o software Adobe Audition CC, que permitiu facilitar os precessos de edição de áudio. Este último permitiu-me todo o tipo de edição e montagem de ficheiros áudio bem como a gravação dos de teste.

Neste projeto apliquei algoritmos de filtro de ruído, detecção de eventos, localização de sons em topologia Ambisonic e em triângulo e ainda separação de eventos.

Aqui comparei o desempenho, em situações semelhantes, das duas topologias de microfones comparando a abordagem em função das energias dos sinais com a dos tempos de atraso.

Os ficheiros de áudio usados para teste tinham frequência de amostragem de 48000 amostras por segundo.

Capítulo 4

Validação e Testes

Para este projeto foram necessários ficheiros áudio de teste gravados nas topologias ou arquiteturas usadas neste projeto. Infelizmente não haviam ficheiros de teste disponíveis na *web* que pudessem ser usados aqui. Por isso foram gravados ficheiros de áudio de propósito para este projeto. Estas gravações foram feitas em câmara anecoica e em ambiente real, em locais fechados bem como em espaços semi-abertos.

Os ficheiros usados incluem eventos sonoros de "cães a ladrar" , "buzina" , "travagem de carro" , "disparo", "musica" e "fala", com e sem ruído. O ruído, quando presente nos ficheiros, foi acrescentado programaticamente e também obtido por gravações em ambiente real. Estes ficheiros estavam todos amostrados a 48000 amostras/segundo. Foram gravados em 4 e 3 canais simultaneamente para as arquiteturas de Ambisonic e em Triângulo, respetivamente.

Para os filtros de ruído efetuei testes com ficheiros com ruído artificial e real em ambiente de cidade.

Para o algoritmo de detecção de eventos, efetuei testes com todos os tipos de ficheiros áudio. Os melhores resultados foram, como seria de esperar, obtidos nos eventos que implicavam uma maior e mais abrupta mudança de amplitude. Assim os ficheiros com sons de "cães a ladrar" e "disparo de arma" foram os que obtiveram eventos mais definidos. Ainda assim este algoritmo verificou bons resultados com todo o tipo de eventos sonoros testados.

Para a fase de localização de eventos, efetuei testes para as duas topologias (arquiteturas).

Para a topologia Ambisonic efetuei testes com sons provenientes de ângulos

0, 90, 180 e -90 graus para todos os tipos de sons e em ambiente controlado (câmara anecoica) ou em cidade.

Para a segunda topologia em triângulo testei o algoritmo com ficheiros de sons emitidos a 0, 40 e 110 graus. Estes testes foram efetuados com misturas virtuais e reais em ambiente aberto e fechado (com maior reverberação). Isto dá-nos uma percepção de como o algoritmo funciona conforme se vão alterando as condições de trabalho, desde ambientes mais controlados até outros mais realistas. Os resultados foram melhores para ambientes de misturas virtuais e piores para reais em ambiente fechado.

Aqui os resultados foram claramente melhores para a arquitetura Ambisonic, revelando gráficos de ângulos muito mais definidos. Na arquitetura e algoritmo Ambisonic os "picos" dos gráficos de localização são, duma forma geral, mais pronunciados e existe menos ruído. Ainda assim, em ambiente real aberto, existe nos dois casos, algum ruído. Enquanto no algoritmo para Ambisonic existe muito ruído, e por isso mesmo mais "picos" falsos, para a arquitetura em Triângulo e algoritmo *GCC-PHAT* existem "picos" mais pronunciados mas com maior margem de erro no ângulo que deveria ter sido identificado. Assim, mesmo sendo fácil identificar os "picos" na arquitetura em triângulo, essa identificação dos ângulos poderá estar mais longe do valor real do que no caso do Ambisonic.

Para a fase de separação foram testados os ficheiros de testes da ambas as arquiteturas. Ainda assim os resultados não divergiram muitos. Havia, à partida, a limitação da arquitetura em triângulo só poder separar 3 eventos sonoros no máximo, por o número de microfones ser também 3. Ainda assim a separação propriamente dita foi semelhante para ambos os cenários.

Os testes e resultados estão relatados nas secções seguintes.

4.1 Filtros de Ruído

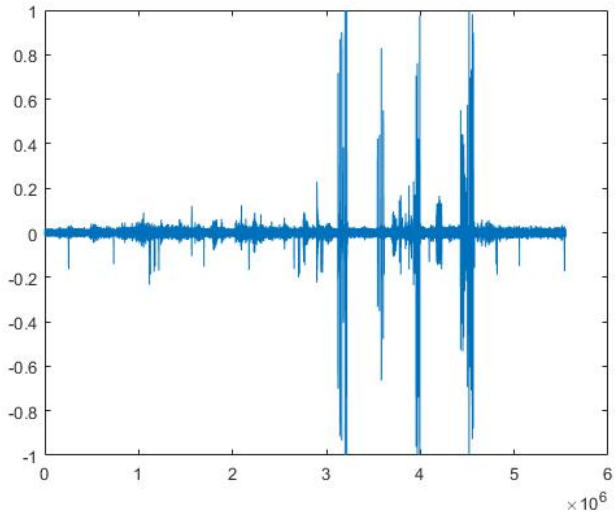
Neste projeto foi abordado a temática dos algoritmos de redução de ruído, neste caso, aplicados a ficheiros de áudio. Foram estudados os filtros de Subtração Espectral assim como o filtro de Wiener.

Na figura 4.1 são mostrados 3 sinais. O primeiro representa um sinal de áudio original sem qualquer filtro de ruído. Os 2 gráficos seguintes mostram o mesmo sinal de áudio após filtro de Subtração Espectral e Filtro de Wiener

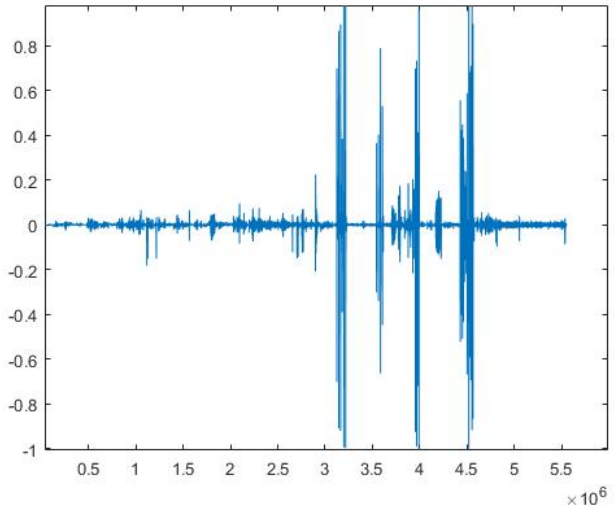
respectivamente.

É importante salientar que há uma "linha ténue" a separar um bom filtro que retira apenas ruído dum que remove ruído e também informação relevante do sinal para os algoritmos de localização. Por essa mesma razão, na maior parte dos testes realizados neste projeto foram feitos sem recurso a algoritmos de redução de ruído.

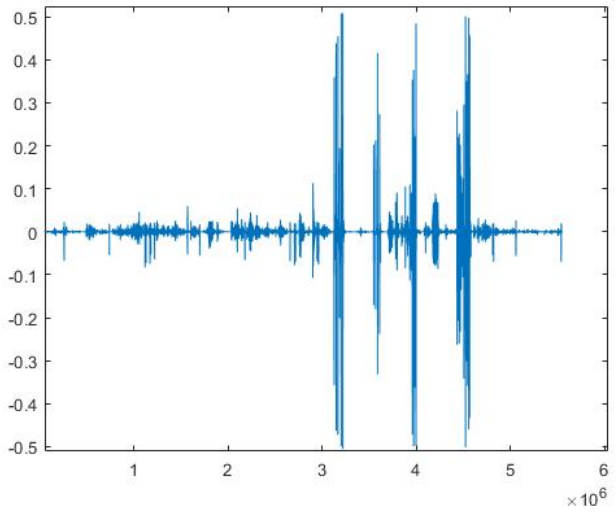
Os algoritmos de redução de ruído revelaram-se, ainda assim, sempre, ou quase, muito úteis para o bom funcionamento do algoritmo de Detecção de Eventos, *Spectral Flux* .



(a) Sinal de Áudio



(b) Sinal após filtro de Subtração Espectral



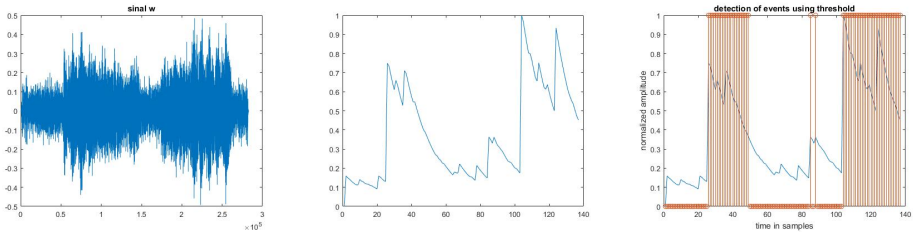
(c) Sinal após filtro de Wiener

Figura 4.1: Sinal de áudio para os diferentes filtros de ruído

4.2 Detecção de Eventos

Como descrito em secções anteriores, a detecção de eventos faz-se através do algoritmo de Spectral Flux, explicado em 2.1 e mais concretamente pela expressão em 3.8 .

Os gráficos na figura 4.2 mostra as diferentes fases do processo que ocorre para se processar a detecção de eventos. Primeiro temos um sinal de áudio com dois momentos de maior amplitude. De seguida temos a expressão *Spectral Flux* , aplicada tal como indicada em 3.8 . Por último temos a normalização desta mesma expressão anterior à qual foi aplicada um *threshold* acima do qual o sinal é identificado como evento sonoro, e abaixo do qual como apenas ruído. Aqui o *threshold* é definido pelo utilizador da melhor forma que se adeque ao ambiente no qual os microfones se encontram.



(a) Sinal de Áudio (b) HSF após *smooth* (c) Deteção de Eventos

Figura 4.2: 3 Fases do processo para detetar eventos

Tendo este gráfico, defini um *threshold*, ou limiar, que pode ser ajustado conforme o tipo de sons a ser analisado para um melhor desempenho. Aplicando esse *threshold* obtive o gráfico da figura 4.2c que mostra os trechos do sinal a serem analisados demarcados a laranja.

Os eventos, assinalados com as linhas verticais laranja, que sejam demasiado curtos são considerados resultados de ruído existente no sinal e são por isso descartados. Assim neste exemplo apenas dois eventos serão analisados, correspondendo aos trechos de maior largura.

4.3 Localização de Eventos

4.3.1 Arquitetura em Triângulo

Como referido antes os testes foram realizados para ângulos de 0, 40, e 110 graus.

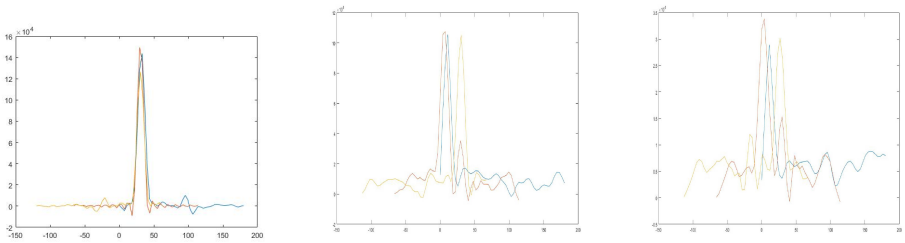
Como explicado anteriormente, a correlação e consequente cálculo dos tempos de atraso é feita com cada par dos 3 microfones. Assim teremos 3 pares e 3 gráficos com que representaram os ângulos calculados. Nestes gráficos os picos (máximo, ou máximos relativos caso haja mais do que um evento sonoro) identificarão os Ângulos obtidos após os cálculos.

Os ficheiros de teste foram originados em 3 ambientes diferentes. Primeiro efetuaram-se misturas virtuais, ou seja, ficheiros misturados programaticamente. Depois foram gravados ficheiros "reais" em ambiente fechado, dentro duma sala, e por ultimo foram gravados com misturas reais em ambiente aberto. Estes diferentes ambientes mostram diferentes graus de realismos em testes. Consequentemente veremos como se comporta o algoritmo em ambientes "controlados" ou realistas.

O esperado seria obter resultados mais concretos para misturas virtuais e menos para misturas reais em ambiente fechado.

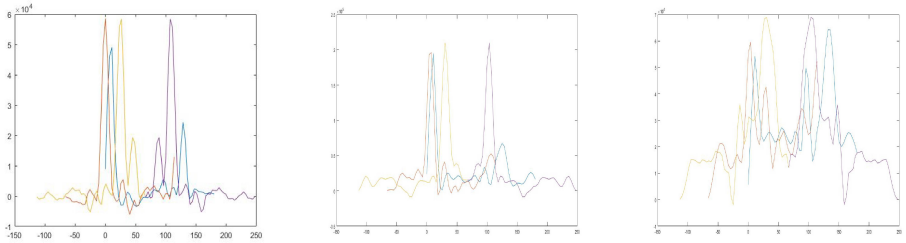
Passo a mostrar os resultados obtidos nas 3 situações para o mesmo tipo de ficheiros áudio.

Repito a demonstração mas agora para o ficheiro com sons de cães a ladrar e buzinas de carros (figura 4.3) .



(a) Mistura Virtual (b) Ambiente aberto (c) Ambiente fechado

Figura 4.3: Ficheiro de "ladrar e buzina" nos 3 ambientes estudados - 1 evento a 0 graus - Topologia em Triângulo - evento "ladrar" seguido de "música"



(a) Mistura Virtual (b) Ambiente aberto (c) Ambiente fechado

Figura 4.4: Ficheiro de "ladrar e buzina" nos 3 ambientes estudados - 2 eventos simultâneos a 0 e 110 graus - Topologia em Triângulo - eventos "ladrar" e "música"

4.3.2 Conceito Ambisonic

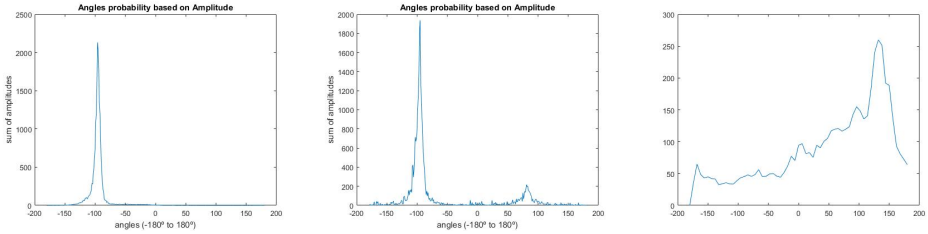
Para a topologia Ambisonic efetuei diferentes testes com ruído de cidade adicionado virtualmente e também ficheiros gravados em ambiente real de cidade. Efetuei testes com eventos sonoros de cães a ladrar, travagem de carro e sirene.

Passo a mostrar a comparação entre resultados obtidos em ambiente virtual sem e com ruído e ambiente real.

Na figura 4.5 são mostrados os gráficos para os 3 ambientes de testes para um único evento. Na figura 4.6 estão ilustrados, também, 3 gráficos para os referidos ambientes de teste mas para 2 eventos em simultâneo.

Como se pode verificar à medida que os testes se tornam mais reais, e menos de "laboratório" os resultados tornam-se mais ruidosos. Ainda assim o algoritmo foi capaz de classificar corretamente, na maior parte dos casos, os eventos sonoros presentes nos ficheiros gravados em ambiente real (cidade)
.

Após obter estes gráficos foi necessário fazer um "filtro" que retornasse apenas os "picos" de maior relevância. Assim criei um algoritmo que apenas identificava um aspaspico no gráfico como sendo um evento se tivesse uma determinada mudança de amplitude em relação aos valores da sua vizinhança. Desta forma elimina-se o evento falso ("pico") presente, por exemplo, na figura 4.5b a cerca de 90 graus. Ou mesmo todos os pequenos "picos" presentes na figura 4.5c retornando apenas o "pico" com maior amplitude e maior mudança de amplitude em relação à sua vizinhança, a cerca de 155 graus.

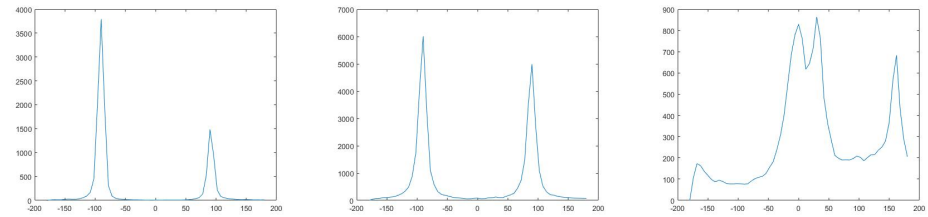


(a) Câmara anecoica - Evento a 270 graus - "disparo"

(b) Câmara anecoica com adição (virtual) de ruído - Evento a 270 graus - "disparo"

(c) Ambiente Real - Evento a cerca de 155 graus - "buzina"

Figura 4.5: 3 Ambientes de teste para topologia Ambisonic com 1 evento único



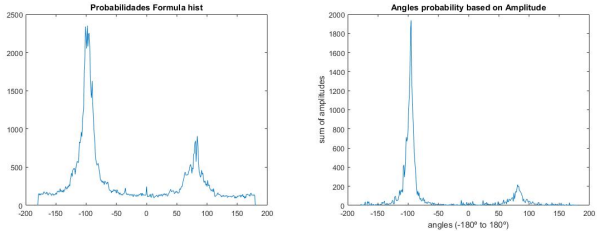
(a) Câmara anecoica - Eventos a 90 e 270 graus - "travagem" e "sirene"

(b) Câmara anecoica com adição (virtual) de ruído - Eventos a 90 e 270 graus - "travagem" e "sirene"

(c) Ambiente Real - eventos a cerca de 25 e 155 graus - "ladrar" e "buzina"

Figura 4.6: 3 Ambientes de teste para topologia Ambisonic com 2 eventos em simultâneo

Um outro aspecto abordado neste projeto foi a maneira como era calculado o ângulo do evento na arquitetura Ambisonic. Assim, como referido anteriormente, foram apresentados dois métodos de cálculo presente nas expressões 3.16 . Os diferentes resultados obtidos estão expostos na figura 4.7 . Como se pode verificar, o histograma proposto em 2.2 apresenta mais ruído e tornava-se em muitos casos menos perceptível de identificar os ângulos corretos.



(a) Histograma (b) Método de Soma de Amplitudes

Figura 4.7: Comparação dos dois métodos para calcular os ângulos dos eventos

4.4 Separação de Eventos

A separação de eventos, como referido anteriormente, foi efetuada através do algoritmo de *fastICA* .

Foram efetuados testes para ambas as arquiteturas de microfones. Entre as duas, a maior diferença terá sido a de que na arquitetura Ambisonic existem 4 microfones enquanto na de triângulo apenas 3. Como mencionado anteriormente, estes factos são relevantes porque os algoritmos de *ICA* separação um número máximo de eventos igual ao número de sinais analisados por este mesmo algoritmo.

Para a arquitetura Ambisonic, num primeiro teste, tentei separar quatro eventos sonoros, sendo eles som de "disparo" , "sirene" , "cães a ladrar" e "ruído de fundo".

A figura 4.10 mostra um dos 4 sinais originais de áudio. Nota que tratando-se dum ficheiro de áudio relativo à topologia Ambisonic o ficheiro de áudio contém 4 (quatro) canais. Deste modo o algoritmo *ICA* permite uma separação de 4 eventos sonoros máxima. Estes ficheiros de teste foram criados virtualmente em *software* próprio.

A figura 4.9 mostra o resultado obtido com um sinal de 4 eventos distintos após processamento por algoritmo *fastICA*.

O aspeto visual destes gráficos é animador pois o algoritmo *fastICA* retorna sinais bastante diferentes entre si, cada um salientado eventos sonoros distintos. Apesar desse facto, os resultados na prática não são assim tão esclarecedores como parecem ser a uma primeira vista da figura 4.9 mas ainda assim destacam com considerável relevância cada um dos eventos presentes

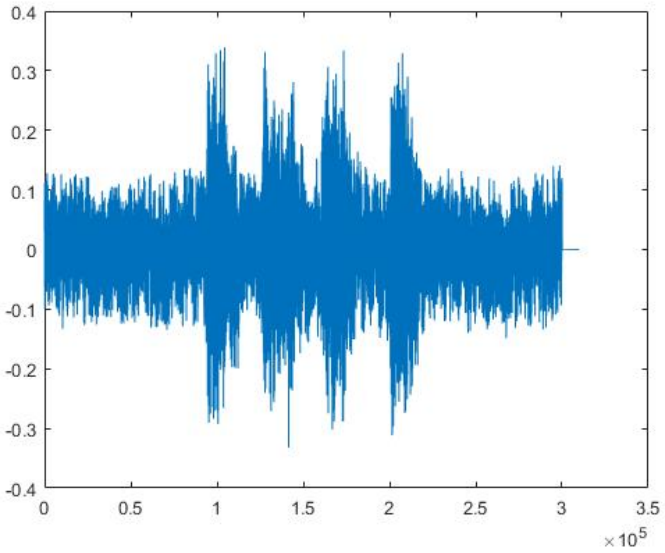


Figura 4.8: Sinal Áudio Original com 4 Sons Distintos - "disparo" , "sirene" , "ladrar" e "ruído de fundo"

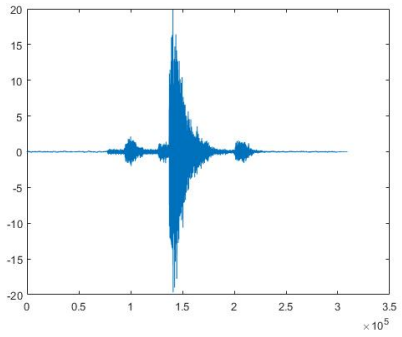
nos ficheiros áudio originais.

O nível de separação produzida neste teste seria, provavelmente, suficiente para fazer classificação de eventos através de algoritmos de *Aprendizagem Automática* ou *Machine Learning*. Isso será algo a ser desenvolvido numa fase mais avançada do seguimento deste projeto. Isto mesmo estará falado e referido na secção posterior deste projeto de *Conclusões e Trabalho Futuro*.

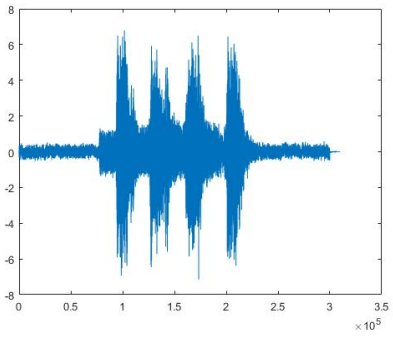
Agora mostro o resultado obtido para misturas reais gravadas em ambiente de cidade. Aqui existe apenas ruído de fundo e sons de cães a ladrar seguido de sons de buzina de carro.

Apesar dos gráficos se diferenciarem bem entre si, o algoritmo *fastICA* salientou diferentes eventos nos sinais de saída, o resultado prático não foi tão bom como no teste efetuado anteriormente para misturas áudio virtuais. Ainda assim penso que permitiria um bom desempenho de algoritmos para classificação aplicados aos ficheiros áudio respetivos aos gráficos da figura 4.11.

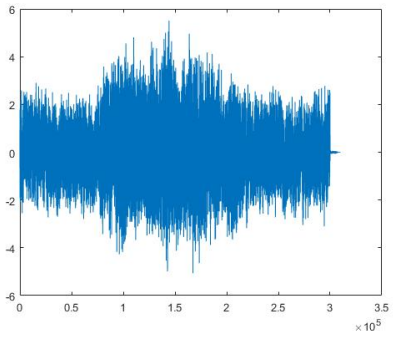
Não se verificaram diferenças de salientar entre a separação para topologia Ambisonic ou em triângulo. Para além da limitação de número de separações por terem diferentes números de microfones e consequentemente canais.



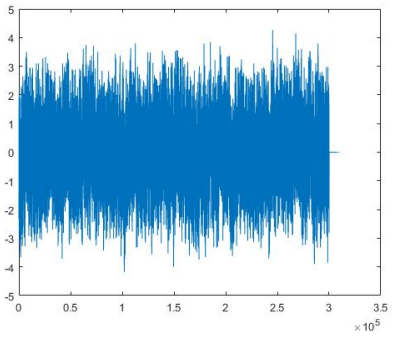
(a) Sinal à Saída 1



(b) Sinal à Saída 2



(c) Sinal à Saída 3



(d) Sinal à Saída 4

Figura 4.9: Separação por *fastICA* - Misturas Virtuais - 4 eventos - Topologia Ambisonic - eventos "disparo" , "sirene" , "ladrar" e "ruído de fundo"

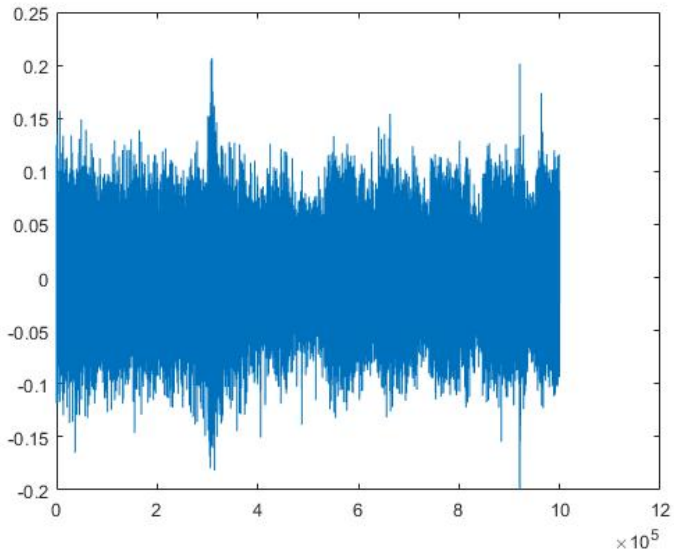
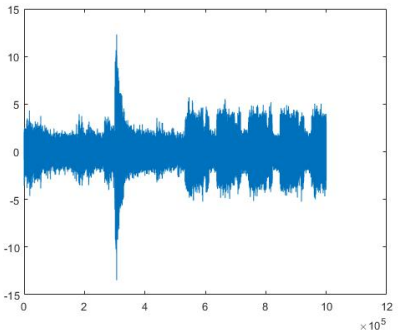
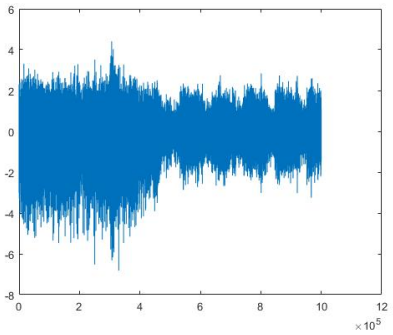


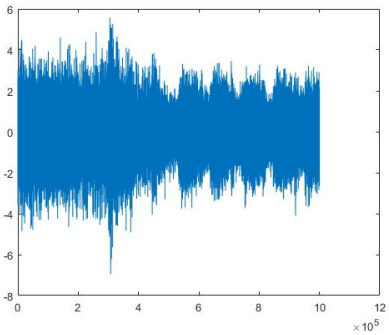
Figura 4.10: Sinal Áudio Original com 3 eventos - eventos "ladrar" , "buzina" e "ruído de fundo"



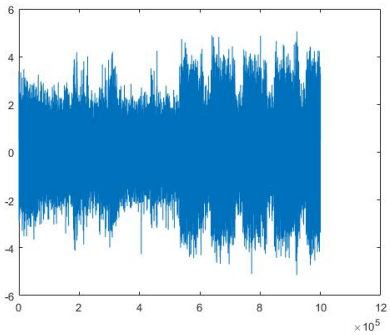
(a) Sinal à Saída 1



(b) Sinal à Saída 2



(c) Sinal à Saída 3



(d) Sinal à Saída 4

Figura 4.11: Separação por *fastICA* - Misturas Reais - 3 eventos - Topologia Ambisonic

Capítulo 5

Conclusões e Trabalho Futuro

5.1 Conclusões

Após a conclusão deste trabalho posso verificar que a topologia Ambisonic é bastante mais eficaz para a localização de eventos. Se a isso somarmos o facto de que esta topologia pode localizar um evento sonoro num espaço 3D, calculando não só o ângulo horizontal (azimute) mas também o vertical (elevação), posso dizer que a topologia Ambisonic é vantajosa em relação à de triângulo.

O algoritmo para o referido Ambisonic é também mais fácil de implementar e bastante mais rápido de executar. Quanto a esta topologia, pode-se concluir também que a fórmula baseada na soma de amplitudes é mais eficiente do que o histograma (ver equações 3.16).

Quanto à topologia de triângulo é perceptível que o algoritmo usado, *GCC-PHAT* não conseguiu obter os resultados desejados, ficando aquém do algoritmo e topologia Ambisonic. No trabalho futuro pretendo desenvolver o algoritmo de localização, baseado em tempos de atraso, que se baseia no conceito dos algoritmos *ICA* . Conceito explorado neste projeto para a separação de eventos, seria neste caso aplicado ao cálculo dos tempos de atraso.

Quanto ao algoritmo de detecção de eventos posso concluir que é bastante eficaz a eliminar momentos de silêncio (apenas ruído). Assim posso dizer que o algoritmo de *Spectral Flux* é bastante preciso, sendo que, obviamente, funciona tão melhor quanto menos ruído houver. Para o algoritmo de detecção de eventos é vantajoso que haja filtros de ruído. O mesmo não se verifica obrigatoriamente para os algoritmos de localização pois esta pode ser afetada

pela perda de informação relevante. Esta poderá ser recolhida sem querer no processamento dos filtros de ruído.

Desta forma acabei por efetuar a grande maioria dos meus testes de localização e separação de eventos sem recorrer a nenhum algoritmo de redução de ruído. Sendo que os resultados para essa mesma localização e separação de eventos ilustrada no capítulo anterior demonstram, exclusivamente, resultados para ficheiros de áudio aos quais não tinham sido aplicadas remoções de ruído.

5.2 Trabalho Futuro

No futuro pretendo prosseguir este projeto para desenvolver um algoritmo que identifique de forma mais eficiente e exata a localização dos eventos.

Um outro objetivo muito importante é a classificação de eventos. Esta secção já irá entrar na área de aprendizagem automática ou *Machine Learning* e será talvez a parte mais desafiante num projeto final que englobe esta área e a de processamento de sinal (DSP).

Para esta identificação de eventos seria essencial uma boa separação dos mesmos, abordada neste projeto através do algoritmo *fastICA*.

A classificação de eventos permitirá enviar avisos para centrais de bombeiros ou polícia caso o evento sonoro captado seja relevante para alguma destas entidades.

Haveria outros algoritmos de separação de eventos que poderiam ter sido estudados como por exemplo o algoritmo *DUET* . Desenvolver este algoritmo poderá passar pelo trabalho a ser desenvolvido no futuro.

Em trabalho futuro, espero também, desenvolver, como referido na secção anterior, o algoritmo de localização *TDOA - Time Difference of Arrival* - com base nos fundamentos do algoritmo *ICA*. Pela pesquisa que fiz e pelo documento que obtive acerca destes algoritmos (3 estudados) posso presumir que os resultados serão melhores podendo, eventualmente, superar os obtidos para o Ambisonic.

Os objetivos finais são interessantes e úteis no contexto real pelo que seria cativante cumpri-los.

Bibliografia

[1] Marc Karam, Hasan F. Khazaal, Heshmat Aglan, Cliston Cole *Noise Removal in Speech Processing Using Spectral Subtraction* 2014

[2] Hasan Khaddour, Jiri Schimmel, and Michal Trzos *Estimation of Direction of Arrival of Multiple Sound Sources in 3D Space using B-Format* 2013

[3] Simon Dixon Austrian Research Institute for Artificial Intelligence *Onset Detection* 2006

[4] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B. Sandler, Senior Member, IEEE *A Tutorial on Onset Detection in Music Signals* 2003

[5] Dimoulas C. A., Kalliris G. M., Avdelidis K. A. and Papanikolaou G. V. *Improved localization of sound sources using multi-band processing of ambisonic components* 2009

[6] J. Wierzbicki, P. Maecki and J. Wiciak *Localization of the Sound Source with the Use of the First-order Ambisconic Microphone* 2013

[7] Microphone Arrays and Time Delay Estimation

[8] Akash Kashyap, Mayank Prasad*Audio Noise Cancellation using Wiener Filter based LMS Algorithm using LabVIEW* 2013

[9] Christian Schorkhuber, Markus Zaunschirm, Franz Zotter, Alois Sontacchi *Localization of Multiple Acoustic Sources with a Distributed Array of Unsynchronized First- Order Ambisonics Microphones* 2014

[10] Augmented Multi-party Interaction with Distance Access *Localization and tracking of multiple interlocutors with multiple sensors* 2006

[11] R. Bullen *A system for autmotically detecting the direction and level noise sources* 2010

[12] Charles Blandin, Alexey Ozerov, Emmanuel Vincent *Multi-source TDOA estimation in reverberant audio using angular spectra and clustering*

[13] O. Bunting, D. Chesmore *Time frequency source separation and direction of arrival estimation in a 3D soundscape environment* 2013

[14] Anthony Lombard, Yuanhang Zheng, Herbert Buchner, Member, IEEE, and Walter Kellermann, Fellow, IEEE *TDOA Estimation for Multiple Sound Sources in Noisy and Reverberant Environments Using Broadband Independent Component Analysis* 2011

[15] Yushi1 Zhang and Waleed2 H. Abdulla *A Comparative Study of Time-Delay Estimation Techniques Using Microphone Arrays* 2005