# Contextualização

O principal objetivo deste projeto é proporcionar à máquina a capacidade de reconhecer batidas de bola e, de seguida produzir estatísticas relativas a essas batidas de bola. No fundo, o que se pretende, é realizar processamento *offline* sobre áudios com batidas de bola e identificar em que instantes temporais estas ocorrem. Para esse efeito, é necessário numa fase inicial, criar um *dataset.* O *dataset* em questão, corresponde a um conjunto de dados com padrões identificativos de batidas de bola.

Em geral, um *dataset* é um conjunto de dados representados como uma tabela composta por colunas e linhas. Cada linha constituinte do *dataset* designa-se por exemplo ou instância. Considerando ainda um determinado exemplo do *dataset*, uma das colunas representa a classe desse exemplo, ao passo que as restantes colunas representam as características (“*features*”).

Neste sentido, considerando um vetor de características, *X*, e o vetor de valores de classes correspondente, *y*, o problema de classificação em questão, tem como objetivo obter a função, *f*, tal que:

Equação - Fórmula representativa do modelo a desenvolver.

A função *f* corresponde a um modelo que deve receber o vetor de características e, classificar de maneira correta cada um dos exemplos desse mesmo vetor. Esse modelo é treinado com o *dataset*, que constitui uma forma de permitir que o primeiro seja capaz de reconhecer as batidas de bola.

# Características a considerar

Tal como referido anteriormente as características permitirão identificar os padrões identificativos de batidas de bola. Neste projeto, as características a considerar para identificar as batidas de bola são o *Onset* e o RMS.

A característica *Onset* divide-se em duas: *Onset detect* e *Onset strength* (*Spectral Flux*). O primeiro *onset* corresponde à identificação do início de um determinado áudio ou nota musical e pode ser utilizado na deteção de picos que ocorram numa determinada onda. Já o segundo *onset* corresponde à deteção de variações de energia entre *frames* consecutivas numa onda, o que pode ser útil na deteção de tipos de sons específicos, como é o caso de sons impulsivos. Os sons impulsivos são caracterizados por terem bastante energia no início. Ao longo do tempo, estes sons sofrem um decaimento em termos de energia. Um exemplo de um som impulsivo, é uma batida de bola.

(Colocar aqui uma imagem representativa dos onsets)

O RMS permite obter uma representação da energia de uma onda ao longo do tempo.

(Colocar aqui uma imagem representativa do RMS)

# Obtenção do vetor de características – X

A matriz X de características é constituída por M linhas ou exemplos. Por sua vez, cada exemplo tem N colunas ou características, pelo que o matriz X pode ser representada da seguinte forma:

Figura – Representação da matriz de características (considerando uma característica).

Simplificando a matriz acima, tem-se:

Figura - Simplificação da matriz de características.

onde cada uma das linhas ou exemplos, é um vetor que contém *N + 1* elementos.

Tal como referido anteriormente, será realizado processamento sobre vários áudios para obter as características associadas a batidas de bola. No decurso deste processo é realizado um varrimento sobre o áudio com uma janela fixa de meio segundo. A janela em questão tem esta dimensão, porque corresponde à duração média de uma batida de bola. A frequência de amostragem a que os áudios são obtidos é de 44100Hz, pelo que meio segundo corresponde a 22050 amostras. Ainda relativamente ao varrimento com a janela fixa, é importante referir que a janela desliza de 1024 em 1024 amostras como primeira abordagem.

O valor de N acima, corresponde ao número de características obtidas considerando 0.5 segundos (22050 amostras). Este valor pode ser dado pela seguinte expressão:

Equação - Fórmula de cálculo do valor de N.

Onde *duration* corresponde à duração da batida de bola (0,5 segundos), *samplingRate* é a frequência de amostragem a que os áudios são obtidos, e *hopLength* o número de grupos de amostras para perfazer a duração da batida de bola (0,5 segundos). Assim, tem-se a seguinte tabela de correspondência entre o grupo de amostras e o valor de N (para uma frequência de amostragem igual a 44100Hz):

Figura - Valores de N obtidos considerando um determinado hop.

|  |  |
| --- | --- |
| *Hop Length* | Valor de N |
| 256 | 86 |
| 512 | 43 |
| 1024 | 21 |
| 2048 | 10 |

No processo de deslizamento, as amostras abrangidas pela janela deslizante são utilizadas para calcular as características associadas. Para o cálculo das características será utilizada a biblioteca [*librosa*](https://librosa.org/doc/latest/index.html). Esta biblioteca divide a janela de amostras em segmentos de *hopLength*, e sobre os mesmos aplica uma FT (Transformada de *Fourier*). Neste sentido, considerando apenas uma iteração do varrimento sobre o áudio, um valor de *hopLength* igual a 1024, e que a janela abrange cerca de 22050 amostras, ao aplicar o algoritmo de cálculo de uma característica, obter-se-á um vetor de 21 valores. Pela mesma lógica, se se considerar um *hopLength* de 256 amostras, obter-se-á um vetor de 86 elementos e se se considerar *hopLength* de 512 amostras, obtém-se um vetor de 43 elementos.

Considerando ainda a figura 2, cada um dos elementos, *Xi*, da matriz é um vetor com três valores (F1, F2 e F3). Esses valores resultam do cálculo das características para um grupo de *hopLength* amostras. Pelo que, são calculados esses três valores sobre N grupos até perfazer as 22050 amostras Ou seja, numa determinada iteração, sobre as amostras abrangidas pela janela deslizante, são calculados os valores das características *RMS*, *Onset* e *Spectral flux*, N vezes Abaixo figura uma representação do varrimento sobre um determinado áudio:

Uma imagem com texto, relógio

Descrição gerada automaticamente

Figura - Processo de varrimento sobre um determinado áudio.

No processo de deslizamento ao longo do áudio acima, *F*1, *F*2 e *F*3 correspondem aos vetores das características obtidas para cada um dos grupos (*Hop Length*) da janela de amostras selecionada (numa determinada iteração). As características são os valores de *Onset*, *Spectral* *flux* e RMS obtidos para a janela deslizante. Note-se que o agrupamento destes três vetores de características permite obter cada um dos elementos, *X*i, do vetor X (nas figura 2). As iterações seguintes permitem obter os elementos restantes do vetor de características final (X).

# Obtenção do vetor de classes

A cada um dos exemplos, , da matriz de características é atribuída uma classe. Desta forma, o vetor de classes pode ser representado da seguinte forma:

Equação - Vetor de classes.

onde cada, , tem apenas um elemento correspondente à classe atribuída ao exemplo.

Numa fase inicial, considerar-se-á a existência de duas classes: Batida de bola e não batida de bola. Para obtenção deste vetor, começou-se por etiquetar os mesmos áudios utilizados para obter o vetor de características, X. O processo resume-se a ouvir cada um dos áudios e anotar num ficheiro .csv, em que instantes de tempo (em termos de amostras) ocorrem batidas de bolas.

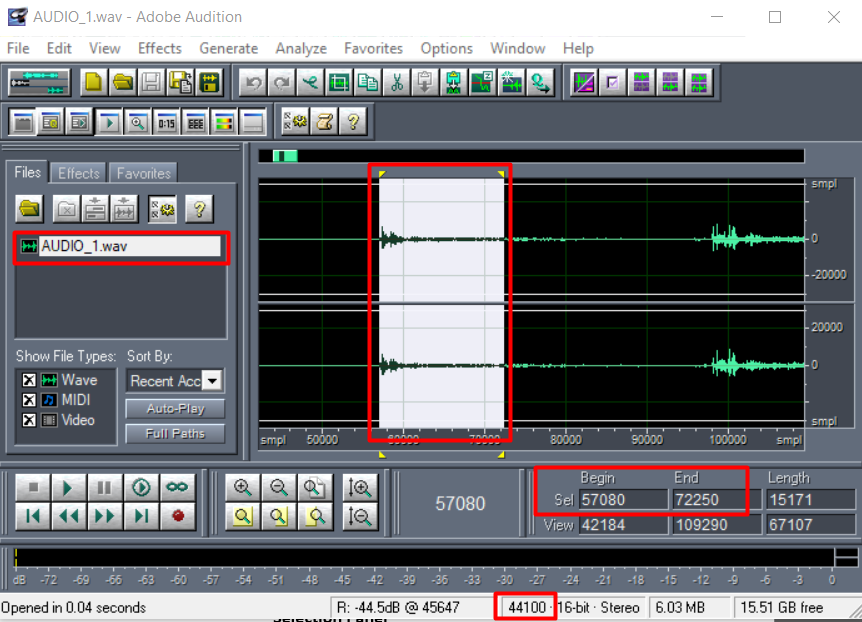


Figura - Etiquetação de um áudio no programa Adobe Audition.

Na figura acima, encontra-se uma seleção do áudio em termos de amostras. De acordo com o painel de seleção (no canto inferior direito), estão selecionadas as amostras de índices no intervalo [57080, 72250]. Esta seleção corresponde a uma batida de bola, pelo que no ficheiro é registada a seguinte linha da tabela abaixo:

|  |  |  |
| --- | --- | --- |
| *Ball hit* | *Begin Sample* | *End Sample* |
| Batida de bola | 57080 | 72250 |

Figura - Anotação de eventos correspondentes a batidas de bola.

Note-se que apenas eventos correspondentes a batidas de bola são registados no ficheiro. Desta maneira, admite-se que os eventos que não são registados, são automaticamente interpretados como não sendo batidas de bola.

Pelo que, cada um dos exemplos, , do vetor y terá um valor que depende da expressão:

Equação - Expressão para obtenção das classes.

# Obtenção do dataset

Tendo obtido a matriz de características e o vetor de valores das classes correspondente, o processo de obtenção do *dataset*, corresponde à junção destas duas componentes. Assim, considerando que a janela desliza ao longo do áudio, um conjunto de exemplos, *X*i, corresponderão a uma mesma classe, *y*i:

Figura - Processo de obtenção do dataset.

Continuando o mesmo raciocínio, um segundo conjunto de amostras corresponderá a outra classe, *y2*, e assim sucessivamente, até que se realize o varrimento sobre todo áudio. Note-se que entre linhas consecutivas existe um desfasamento de um elemento que representa o deslizamento da janela utilizada no processo de varrimento.

# Referências

[1] – <https://librosa.org/doc/latest/index.html>

[2] – <https://en.wikipedia.org/wiki/Onset_(audio)>

[3] – <https://en.wikipedia.org/wiki/Spectral_flux>