

# Análise de Métodos e seus Parâmetros aplicados à Etapa de Pós-processamento em Detecção de Onsets em Sinais de Áudio

Guilherme Ihlenffeldt , Leonardo G. Tavares (Orientador)

Departamento de Engenharia Elétrica  
Universidade Positivo

guilherme.ihlenffeldt@gmail.com, leonardo.tavares@up.com.br

**Resumo** – Métodos para detecção de onsets normalmente possuem 3 etapas: pré-processamento, redução e pós-processamento. Muitos dos esforços nos últimos anos têm focado na etapa de redução onde se objetiva desenvolver funções cujos máximos locais coincidam no tempo com onsets reais. No entanto, pouca atenção tem sido dada às etapas de pré e pós-processamento. Neste trabalho foram investigados os principais algoritmos citados na literatura para a etapa de pós-processamento e foram conduzidos experimentos para avaliar o impacto de diversos parâmetros de entrada, bem como, relatar os melhores valores encontrados.

**Palavras-chave:** onset, pós-processamento, transcrição automática.

## 1. Introdução

Em processamento de sinais musicais, onset é o termo dado para designar o instante de tempo em que ocorre o ataque de uma nota musical.

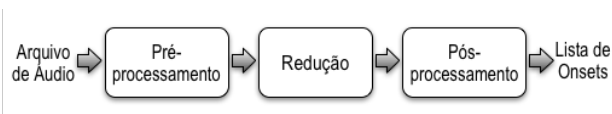
A identificação automática de onsets normalmente leva em consideração o fato de que no instante do ataque de uma nota musical ocorre uma mudança significativa em uma das propriedades psicoacústicas do som: intensidade, altura ou timbre [1, 5, 6].

A detecção de onsets é uma tarefa importante em diversas aplicações no processamento de sinais musicais como na transcrição musical automática, estimação de *beat* (pulsação da música) e outras.

A detecção de onsets também é útil em outras áreas e não somente no processamento de sinais de áudio. Na engenharia biomédica, por exemplo, a detecção de onsets é útil para detectar ataques epiléticos, contrações musculares involuntárias e distúrbios do sono.

## 2. Detecção de Onsets

A tarefa de detecção de onsets é composta de inúmeras etapas podendo ser resumidas de diversas maneiras. Uma das formas de representar a tarefa é resumindo-a em 3 etapas: pré-processamento, redução e pós-processamento (Figura 1).



**Figure 1. Etapas do processo de detecção de onsets**

### 2.1. Pré-processamento

O objetivo do pré-processamento é adequar o sinal de áudio original, acentuando ou atenuando alguns aspectos do sinal para facilitar as etapas seguintes.

É uma etapa opcional. Nessa são executadas estratégias como filtragens (para evidenciar ou atenuar determinadas frequências), normalizações (para reduzir o efeito negativo que diferentes dinâmicas em uma mesma música provocam no desempenho final) e outros tipos de manipulações matemáticas.

### 2.2. Redução

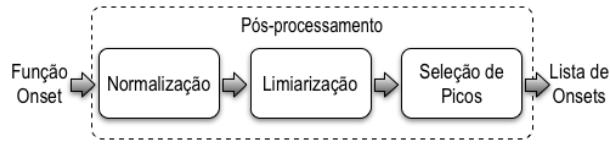
A etapa de redução é executada com auxílio de uma função normalmente chamada de Função de Detecção de Onset, ou simplesmente, Função Onset. O objetivo dessa função é fazer com que os onsets sejam identificados através de seus máximos locais.

Muitas das Funções de Redução são baseadas na ideia de detectar mudanças nas propriedades do sinal de áudio. Métodos baseados em energia, magnitude, fase e representações no domínio complexo são os mais comuns.

As Funções Onset utilizadas neste trabalho são: *Energy*, *High Frequency Content*, *Spectral Difference*, *Complex Domain* e *Phase Deviation*.

### 2.3. Pós-processamento

A etapa de pós-processamento visa a identificação correta dos máximos locais da função onset. Essa etapa pode ser resumida em normalização, limiarização e seleção de picos, como mostra a Figura 2.



**Figure 2. Etapas do pós-processamento**

### 2.3.1. Normalização

O objetivo da normalização é diminuir a diferença entre sinais com dinâmicas distintas podendo assim adotar um mesmo conjunto de parâmetros nas etapas seguintes sem prejudicar o desempenho dos mesmos.

Existem descritas na literatura algumas estratégias para a normalização da função onset. Neste trabalho o tipo de normalização adotada é feita em dois passos: primeiramente a função onset tem todos os seus valores subtraídos do valor médio. Em seguida, a função tem todos os seus valores divididos pelo valor máximo em módulo.

Desta forma, a função onset normalizada resultante possui valor médio igual a 0 e valor máximo, em módulo, igual a 1.

### 2.3.2. Limiarização

A estratégia mais simples para a tarefa de limiarização é a utilização de um limiar constante. Define-se um valor limiar  $\delta$  qualquer e considera-se que qualquer máximo local que esteja acima deste valor limiar é um provável onset. Neste contexto é possível denotar então uma função limiar  $\tilde{\delta}$  que, nesse caso, se resume a uma constante  $\delta$  (Eq. 1).

$$\tilde{\delta}(n) = \delta \quad (1)$$

Estratégias mais elaboradas fazem uso de funções de suavização da função onset no lugar de uma função constante. Essa função de suavização pode ser linear, como a baseada na função média [3] (Eq. 2), ou em estratégias não-lineares, como a baseada na função mediana [2, 7] (Eq. 3).

Nesse caso  $\delta$ ,  $\lambda$  e  $M$  são parâmetros ajustáveis.

$$\tilde{\delta}(n) = \delta + \lambda \cdot \text{mean}\{|d(n-M)|, \dots, |d(n+M)|\} \quad (2)$$

$$\tilde{\delta}(n) = \delta + \lambda \cdot \text{median}\{|d(n-M)|, \dots, |d(n+M)|\} \quad (3)$$

### 2.3.3. Seleção de Picos

Após o processo de normalização e limiarização da função de detecção, é necessário identificar quais máximos locais da função de detecção limiarizada correspondem de fato a um onset real.

É possível considerar que todo máximo local da função de detecção limiarizada é um onset, entretanto essa estratégia conduz a um número grande de falsos positivos. A segunda estratégia é considerar apenas o máximo local dentro de uma janela de largura fixa. Essa foi a estratégia adotada neste trabalho.

## 3. Metodologia

Nesta seção são apresentados o banco de dados e as medidas de desempenho utilizadas neste estudo.

### 3.1. Banco de Dados

Para investigação dos métodos e seus parâmetros foi adotado como referência um banco de dados chamado ODB (*Onset Detection Database*) compilado por Pertusa [4] em sua tese de doutorado.

Esse banco de dados é composto por 19 arquivos de áudio que cobrem uma grande variedade de instrumentos e estilos musicais e contém um total de 2155 onsets anotados.

Durante a construção deste banco de dados, as posições verdadeiras dos onsets foram marcadas e revisadas manualmente com auxílio de um software chamado SFS7, *Speech Filling System*. Esse banco de dados está disponível para download com a finalidade de incentivar a pesquisa.

### 3.2. Medidas de Desempenho

*F-measure* é a medida de desempenho mais utilizada para avaliação dos métodos de detecção de onsets. Ela basicamente consiste em uma média harmônica de outras duas medidas: *Precision* e *Recall*.

Essas duas medidas (*Precision* e *Recall*) possuem características distintas sendo muito comum acontecer de um determinado método possuir alta *Precision* e baixo *Recall*, ou vice-versa. O objetivo da *F-measure* é convergir em uma única medida essas características. A *F-measure* é calculada como mostra a Eq. 4.

$$FMeasure = \frac{2PR}{P + R} \quad (4)$$

*Precision*, por sua vez, expressa a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas (Eq. 5).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

*Recall* apresenta a porcentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas (Eq. 6).

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Ambas as medidas, *Precision* e *Recall*, são calculadas a partir de três possíveis resultados apresentados por um método classificador qualquer.

- TP: *true positive* (verdadeiro positivo) - número de casos positivos que foram corretamente classificados como positivos;
- FN: *false negative* (falso negativo) - número de casos positivos que foram equivocadamente classificados como negativos;
- FP: *false positive* (falso positivo) - número de ocorrências negativas que foram equivocadamente classificados como positivas;

## 4. Experimentos e resultados

Com o objetivo de identificar melhores valores para os parâmetros dos algoritmos de pós-processamento foram conduzidos alguns experimentos. Nesses foram executadas as etapas de normalização, limiarização e seleção de picos conforme descritas na seção 1.3.

A etapa de normalização foi executada em todos os experimentos. Essa foi realizada como descrito na seção 1.3.1.

O algoritmo de limiarização foi executado de três formas: limiarização constante (Eq. 1), limiarização dinâmica com suavização através da função média (Eq. 2) e limiarização dinâmica com suavização através da função mediana (Eq. 3).

O método de seleção de picos adotado é o de identificação de um máximo local dentro de uma janela móvel de largura  $W$ .

### 4.1. Experimento 1

O primeiro experimento realizado teve por objetivo identificar valores de limiar constante ( $\delta$ ) e largura de janela do processo de seleção de picos ( $W$ ) que apresentassem bons resultados.

Nesse experimento foram utilizados os procedimentos de normalização, limiarização fixa com valor de limiar  $\delta$  variando de -0.9 a 0.9, e valores de largura de janela  $W$  variando de 1 a 15.

A Figura 3 mostra o resultado deste experimento para o método *Spectral Difference*. Neste gráfico, o eixo horizontal representa o valor de limiar  $\delta$ , e o eixo vertical a largura da janela  $W$ . A área em roxo mostra a região onde o algoritmo apresentou melhores resultados ( $F\text{-Measure} > 0.6$ ). Essa região revela que os melhores resultados se encontram, em média, para  $\delta$  próximo a 0 e  $W$  em torno de 5. Para valores de  $\delta$  maiores que 0.6 o desempenho é bem baixo ( $F\text{-Measure} < 0.2$ ).

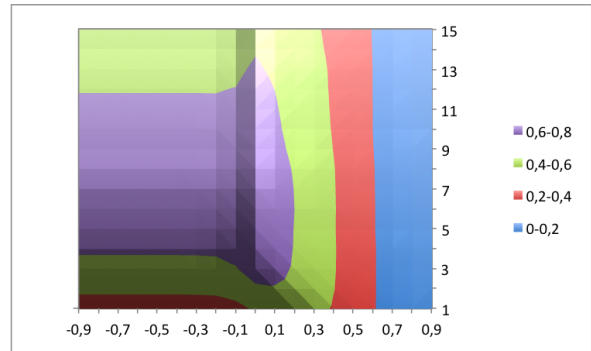


Figure 3. Limiarização Constante com *Spectral Difference*

### 4.2. Experimento 2

O segundo experimento teve por objetivo encontrar bons valores para o limiar  $\delta$  usando as funções de limiarização com suavização por função média e mediana. Nesse foram executados os procedimentos de normalização, limiarização dinâmica (média e mediana) com  $\delta$  variando de -0.5 a 0.5,  $\lambda$  variando de 0.2 a 1.8 e  $M$  de 1 a 18. A largura da janela de seleção de picos  $W$  variou de 1 a 15.

O resultado médio de todas as execuções organizados pelo valor de limiar  $\delta$  está apresentado nas Figuras 4 e 5. Nesses gráficos o eixo horizontal representa o valor de limiar  $\delta$ , e o eixo vertical a medida de desempenho  $F\text{-Measure}$ .

É possível observar através dos gráficos que bons resultados são obtidos com valores de limiar  $\delta$  entre

-0.1 e 0.2 aproximadamente.

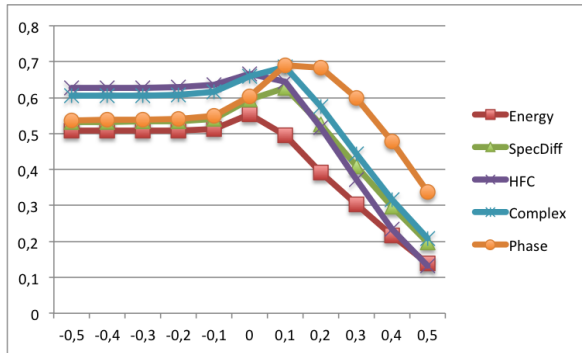


Figure 4. Resultado do Experimento 2 com Limiarização Dinâmica por Função Média

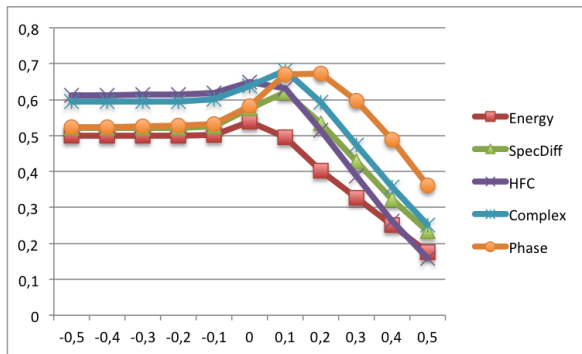


Figure 5. Resultado do Experimento 2 com Limiarização Dinâmica por Função Mediana

### 4.3. Experimento 3

O terceiro experimento teve por objetivo identificar o impacto do parâmetro  $\lambda$ . Nesse foram executados os procedimentos de normalização, limiarização dinâmica (média e mediana) com  $\delta$  variando de -0.1 a 0.2,  $\lambda$  variando de 0.2 a 1.8,  $M$  de 1 a 18 e  $W$  de 1 a 15.

O resultado médio de todas as execuções organizados pelo parâmetro  $\lambda$  está apresentado nas Figuras 6 e 7. Nesses gráficos o eixo horizontal representa o parâmetro  $\lambda$ , e o eixo vertical a medida de desempenho  $F$ -Measure. É possível identificar que valores de  $\lambda$  maiores que 0.9 devem ser evitados, pois o desempenho cai bastante.

### 4.4. Experimento 4

O quarto experimento teve por objetivo identificar o impacto do parâmetro  $M$  (a largura da janela da função de suavização por média ou mediana). Nesse foram executados os procedimentos de normalização, limiarização dinâmica (média e mediana) com

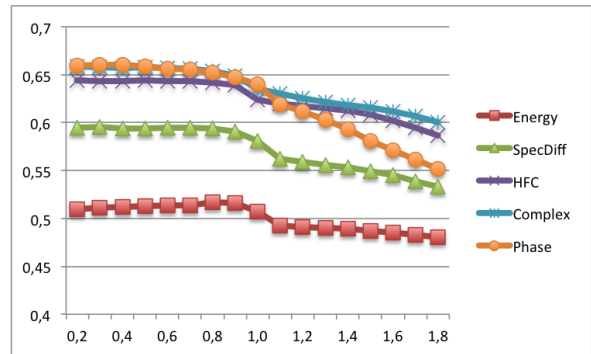


Figure 6. Resultado do Experimento 3 com Limiarização Dinâmica por Função Média

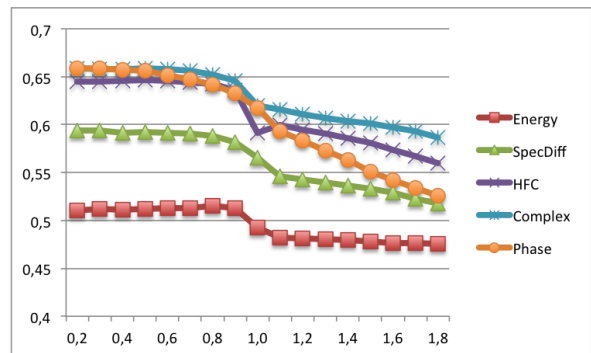


Figure 7. Resultado do Experimento 3 com Limiarização Dinâmica por Função Mediana

$\delta$  variando de -0.1 a 0.2,  $\lambda$  variando de 0.2 a 0.9,  $M$  de 1 a 16 e  $W$  de 1 a 15.

O resultado médio de todas as execuções organizados pelo parâmetro  $M$  está apresentado nas Figuras 8 e 9. Nesses gráficos o eixo horizontal representa o parâmetro  $M$ , e o eixo vertical a medida de desempenho  $F$ -Measure.

É possível identificar que inicialmente o desempenho aumenta com o aumento de  $M$ , mas estabiliza por volta de aproximadamente 8.

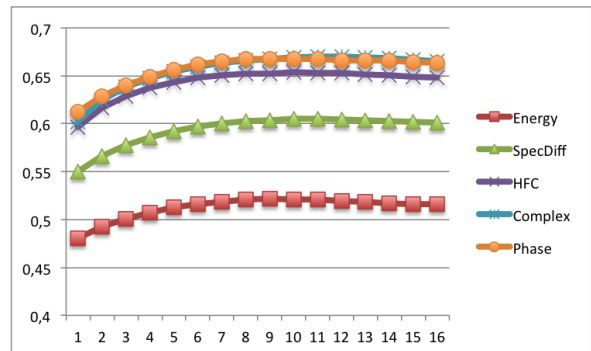


Figure 8. Resultado do Experimento 4 com Limiarização Dinâmica por Função Média

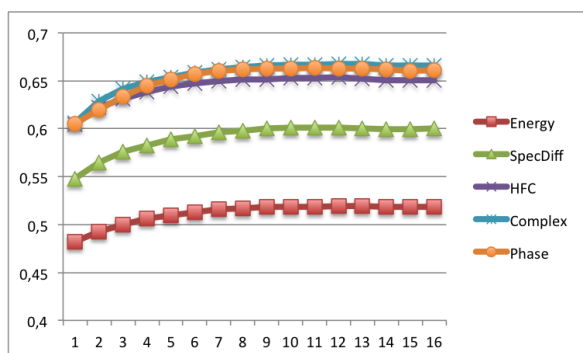


Figure 9. Resultado do Experimento 4 com Limiarização Dinâmica por Função Mediana

## 5. Conclusões

Neste trabalho foram descritos e investigados os principais métodos de pós-processamento em detecção de onsets e seus parâmetros.

Os experimentos realizados revelaram informações interessantes sobre os parâmetros utilizados e podem servir de guia para implementações futuras.

Em praticamente todos os casos o desempenho utilizando limiarização dinâmica, seja por média ou mediana, tem superioridade sobre a limiarização constante, fato este encontrado na literatura.

Através dos resultados encontrados é possível afirmar que não há evidências de que a limiarização pela mediana produz melhores resultados que a limiarização pela média, o que contraria estudos anteriores. Aliás, é possível afirmar que a escolha adequada dos parâmetros  $\delta$ ,  $\lambda$  e  $M$  é bem mais significativa que a própria escolha da função de limiarização.

Bons valores de  $\delta$  estão entre -0.1 e 0.2 aproximadamente, entretanto foi observado que o melhor resultado depende do método escolhido. Por exemplo, para os métodos *Energy* e *HFC* o melhor  $\delta$  é 0, para *Spectral Difference* e *Complex Domain* o melhor  $\delta$  é 0.1.

O valor do parâmetro  $\lambda$  não deve ultrapassar 0.9. Após esse valor o desempenho cai bastante.

Com relação ao parâmetro  $M$  é possível afirmar que o desempenho do método aumenta junto com o aumento do valor de  $M$ . Entretanto a partir de 8, o desempenho estabiliza.

Em geral, o melhor método é o *Complex Domain* e o pior é o *Energy*.

A superioridade do método *HFC* sobre o *Energy* descrita na literatura pode ser observada pelos resultados.

Como trabalhos futuros pode-se citar o estudo de métodos que não foram abordados neste trabalho como o *Weighted Phase Deviation* e o *Rectified Complex Domain*.

Sugere-se também o estudo de estratégias de pré-processamento como o *Adaptive Whitening*.

## References

- [1] Miguel A. Alonso, Gael Richard, and Bertrand David. Extracting note onsets from musical recordings. In *ICME*, pages 896–899. IEEE, 2005.
- [2] J. P. Bello, C. Duxbury, M. E. Davies, and M. B. Sandler. On the use of phase and energy for musical complex domain. In *IEEE Signal Processing Letters*, pages 533–556, 2004.
- [3] Simon Dixon. Onset detection revisited. In *Proceedings of the 9th international conference on digital audio effects*, pages 133–137, 2006.
- [4] Antonio Pertusa Ibanez. *Computationally Efficient Methods for Polyphonic Music Transcription*. PhD thesis, Universidad de Alicante, 2010.
- [5] Anssi Klapuri. Sound onset detection by applying psychoacoustic knowledge. In *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*, volume 6, pages 115–118, Washington, DC, USA, 1999.
- [6] Anssi Klapuri. *Signal Processing Methods for the Automatic Transcription of Music*. PhD thesis, Tampere University of Technology, Finland, March 2004.
- [7] Dan Stowell and Mark Plumbley. Adaptive whitening for improved real-time audio onset detection. In *Proceedings of the International Computer Music Conference (ICMC)*, 2007.