

Segmentação automática de sinais musicais monofônicos para análise da expressividade

Tairone Magalhães

Rodrigo Borges

Thiago Campolina

Davi Mota

Maurício Alves Loureiro

Universidade Federal de Minas Gerais

mauricio@musica.ufmg.br

Sumário:

A detecção dos instantes de início e final de uma nota é um processo essencial para a análise da expressividade de uma performance. Neste artigo propomos dois métodos para a detecção destes instantes, um método que utiliza o envelope da energia RMS com limiar adaptativo conjuntamente com a estimação da frequência fundamental, e outro a partir do fluxo espectral. Os métodos foram aplicados em amostras de trechos extraídos de execuções solo na clarineta. O desempenho do método foi avaliado comparando os dados da segmentação automática com os dados obtidos em uma segmentação feita manualmente.

Palavras-Chave: expressividade, performance musical, extração e processamento de sinais musicais, segmentação.

Introdução

Métodos de segmentação automática de notas em sinais musicais visam extrair informações sobre pontos ou intervalos importantes na evolução de uma nota. A detecção precisa dos instantes de início e fim de uma nota não é um problema trivial, mesmo quando se trata de sinais musicais monofônicos, já que a subjetividade na discriminação destes instantes não deve ser desconsiderada.

Estudos sobre expressividade musical têm demonstrado que músicos utilizam pequenas variações de duração, articulação, intensidade, altura e timbre para comunicar ao ouvinte, aspectos da música que eles interpretam (Gabrielsson, 1995) (Juslin 1997; 2000). Tais variações podem ser observadas em performances de diferentes intérpretes e mesmo entre duas interpretações feitas pelo mesmo músico. Estes desvios são percebidos com uma clareza surpreendente, mesmo para ouvintes não especializados. Recentemente um grande número de pesquisas quantitativas sobre diferentes aspectos da expressividade musical vem sendo realizado, com base em medições de parâmetros como tempo, dinâmica e timbre, procurando-se descrever e reconhecer classes de padrões da evolução destes parâmetros que possam elucidar sua influência na expressividade percebida. Portanto, a análise da expressividade em performances musicais depende do desenvolvimento de ferramentas capazes de extrair e medir tais parâmetros, que por sua vez dependem da detecção precisa dos instantes de início e fim dos eventos a serem analisados, tais como notas musicais, grupos de notas, ou regiões específicas no âmbito de uma mesma nota. A precisão e robustez do processo de segmentação do sinal são fundamentais para a definição e estimação adequada dos descritores de expressividade. Além da detecção dos instantes de início e fim das notas que segmentam o sinal em notas individuais, é necessário detectar também instantes intra-notas que segmentam o envelope da nota em ataque, sustentação e decaimento.

Metodologia de segmentação das notas

Envelope da energia RMS

A estimação do envelope de energia é mais comumente feita através da medida RMS - *root mean square* - (raiz da média dos quadrados), que está relacionada com a potência média de curta duração do sinal. O valor RMS de um sinal, x é dado por:

$$RMS(n) = \sqrt{\frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x[n-m]^2 w[m]}$$

onde n é o número da amostra (instante de tempo) e N o tamanho da janela w de cálculo da média, o qual determina a resolução temporal do envelope.

Detecção de início e final de nota pela derivada da energia RMS

O início de uma nota é normalmente acompanhado por um aumento súbito na amplitude do sinal. Identificando os trechos em que o envelope do sinal possui maior inclinação, é possível detectar um potencial ponto de início de uma nota musical. Da mesma forma, um ponto de inclinação mínima (negativa) corresponde ao fim de uma nota.

A derivada da energia RMS de um sinal nos fornece informações sobre a variação da inclinação da curva do RMS ao longo do tempo. Portanto, os picos positivos da curva da derivada (pontos de inclinação máxima do RMS) correspondem a inícios de notas e os picos negativos por sua vez correspondem a finais de notas.

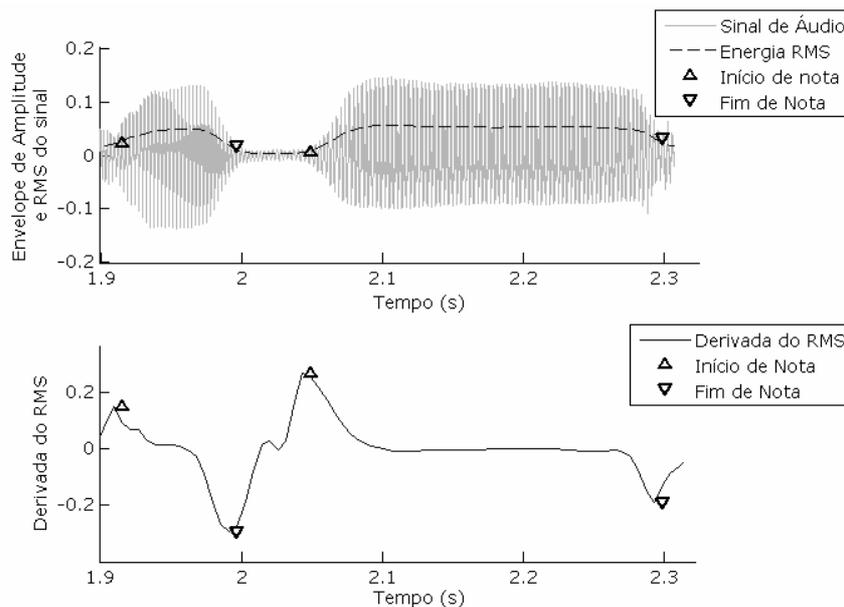


Figura 1: Detecção de inícios e finais de notas através dos picos na curva de derivada da energia RMS do sinal

Detecção de início e final de nota pela energia RMS com limiar adaptativo

De Poli (2005) sugeriu um método em que inícios de notas são calculados a partir de uma curva média de energia RMS calculada em um intervalo de duração bem maior do que a do quadro de tempo de cálculo da energia RMS. Esta curva estabelece um limiar adaptativo, que intercepta a curva de energia RMS nas regiões de transição entre notas consecutivas, nos pontos P_1 e P_2 da Figura 2, que mostra a curva da energia RMS (curva contínua) e a curva média de RMS de longa duração (curva pontilhada). Os pontos de início de nota podem ser encontrados a partir do valor mínimo de energia na região abaixo do limiar adaptativo.

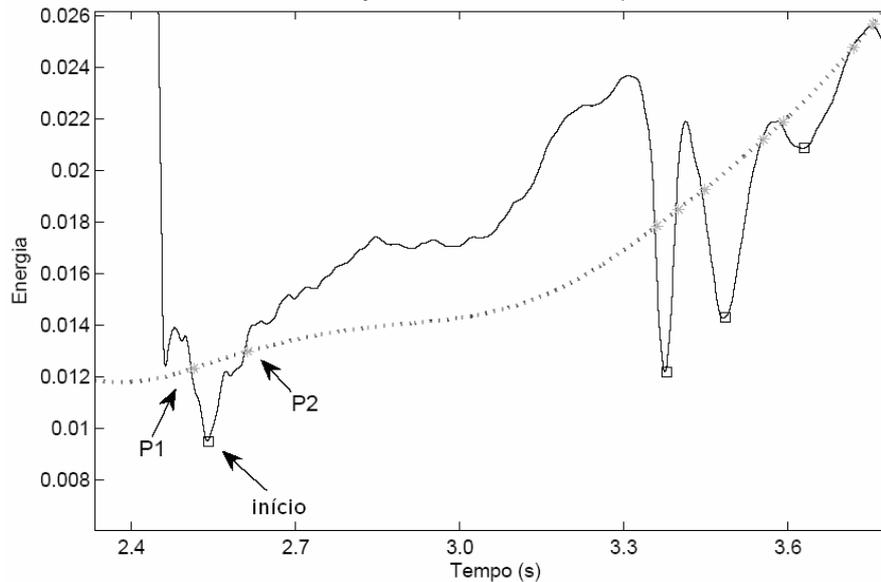


Figura 2: Detecção de inícios de notas a partir de um limiar adaptativo (média da energia RMS de longa duração)

Este método estima apenas um ponto entre duas notas consecutivas, mesmo no caso de duas notas separadas por pausa. A estimação dos instantes de início e final de nota foi possível a partir de uma alteração simples do método do De Poli, que consiste em considerar apenas um percentual do RMS médio de longa duração, a partir de um fator multiplicador ao limiar adaptativo, que considera o início da nota quando a energia do sinal ultrapassa esse limiar e seu fim quando o ultrapassa no sentido contrário (Figura 2).

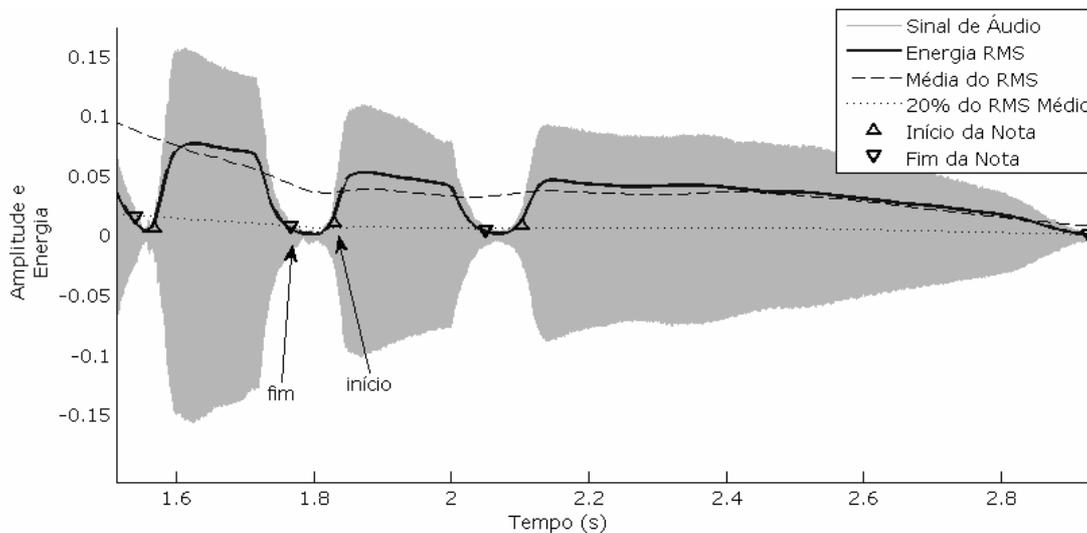


Figura 3: Detecção de inícios de notas a partir de um limiar adaptativo (20% da média da energia RMS de longa duração)

Detecção de início e final de nota a partir da frequência fundamental

A segmentação de notas consecutivas pode não ser viável apenas a partir de níveis de amplitude, como por exemplo, em casos de notas ligadas, em que o instante de final de nota pode coincidir com o instante de início da nota seguinte. Nesses casos a estimação da frequência fundamental pode auxiliar na detecção destes instantes.

Na análise de sinais musicais monofônicos, que não utilizam técnicas contemporâneas de execução, a altura (frequência fundamental) permanece muito estável durante a execução de uma nota. Desenvolvemos um algoritmo para detecção do início e fim da nota, correspondentes aos instantes de início e fim de uma estabilização da variação da frequência fundamental, permitida dentro de um limiar considerado de 5 %, ou

seja, pouco menor que um semitom, como mostra a Figura 4. A detecção de frequência fundamental foi feita a partir da estimação dos 3 valores máximos de magnitude do espectro, para cada quadro. A frequência fundamental é escolhida como aquele candidato que apresentar o maior valor de energia somada às energias de seus 2 primeiros componentes harmônicos.

Este método pode falhar no caso de duas notas de mesma altura consecutivas, entretanto, nestas situações o método do RMS é mais eficaz na detecção de início da nota seguinte, já que há uma articulação entre elas. A combinação entre estes dois métodos, estabilização da frequência fundamental e limiar de energia RMS, resulta num método mais robusto para a detecção de início e fim de nota.

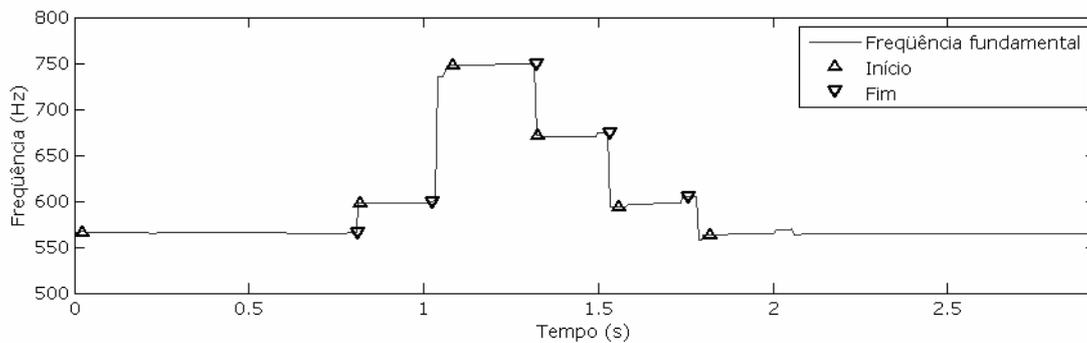


Figura 4: Detecção de inícios de notas a partir da frequência fundamental

Detecção de início e final de nota por fluxo espectral

Durante a transição entre duas notas acontecem quedas abruptas no valor do *fluxo espectral* devido à grande presença de transientes no sinal. Portanto, o fluxo pode ser utilizado para estimar pontos de início e final de notas, ou até mesmo de final de ataque e início do decaimento

O fluxo espectral tende a estabilizar quando existe uma harmonicidade entre as componentes espectrais de maior magnitude. Portanto, a estabilização do fluxo espectral está geralmente associada ao início de uma nova nota. Porém, o que se constatou é que a estabilidade total do fluxo espectral (correlação máxima entre quadros adjacentes) só é atingida após o ponto do início perceptivo na nota, o que pode ser contornado estabelecendo-se um limiar, a partir do qual o parâmetro seria considerado como estável e um limiar temporal mínimo correspondente ao valor mínimo de duração do evento (nota) a ser detectado pelo método.

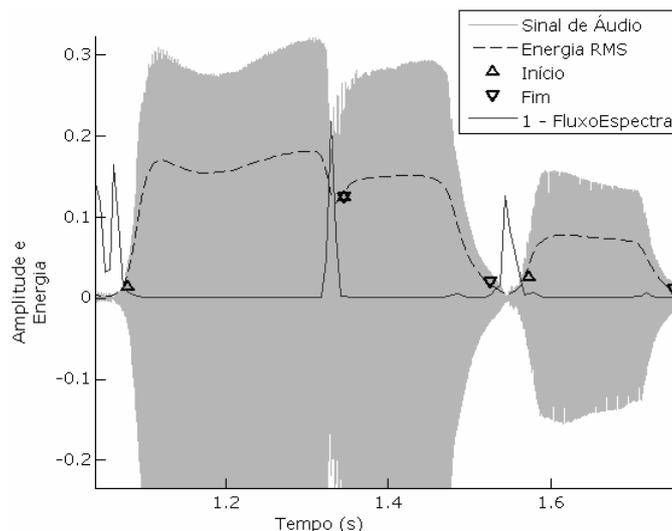


Figura 5: Detecção de inícios de notas a partir do fluxo espectral

A detecção do final da nota através do fluxo espectral não é tão trivial, pois o fluxo tende a sofrer uma queda significativa próximo ao início do decaimento da nota, e a partir dessa queda seu comportamento

é muito imprevisível (devido à grande presença de transientes nessa região). Portanto a combinação do método do fluxo espectral com o método do RMS pode gerar resultados mais satisfatórios. A detecção final de início e fim de nota elaborado neste trabalho consiste na utilização conjunta de três dos métodos descritos acima: detecção por derivada da energia RMS, por energia RMS com limiar adaptativo e por frequência fundamental.

Deteção de final de ataque e início do decaimento a partir da energia RMS

Não existe na literatura um método de medição que possa descrever inequivocamente o ataque (Park, 2004). Em muitas situações, o instante de final de ataque refere-se ao instante de amplitude máxima da nota, mas sua detecção pode demandar procedimentos mais complexos, como, por exemplo em instrumentos não percussivos ou não pinçados, tais como sopros, cordas e voz, nos quais a amplitude máxima poder ser atingida ao longo da parte sustentada da nota, bem depois do ataque. Pode-se contornar o problema estabelecendo-se um valor adequado de limiar máximo de amplitude, mas que pode não apresentar resultados consistentes para notas mais longas tocadas nestes instrumentos.

A detecção dos instantes de final do ataque e o início do decaimento pode ser feita a partir da estimação dos máximos das taxas de variação de energia dentro da nota. O método é baseado na análise do contorno do envelope de energia da nota. Uma vez que os inícios e os finais de notas foram detectados, observa-se a curvatura do envelope da nota, com a finalidade de identificar os três segmentos distintos: ataque, sustentação e decaimento. Considerando o envelope de energia como uma função diferenciável contínua no tempo, os pontos de máxima curvatura são considerados variações máximas locais da derivada primeira do sinal, e conseqüentemente, máximos e mínimos da derivada segunda. Os pontos correspondentes aos valores mínimos da derivada segunda (curva superior da Figura 5) determinam os pontos candidatos para os instantes de final de ataque e início de decaimento na curva de energia RMS (curva inferior da Figura 5). O candidato escolhido para o final de ataque será aquele que apresentar maior inclinação positiva medida entre ele e o início da nota, enquanto que a maior inclinação negativa com o final da nota indicará o início de decaimento.

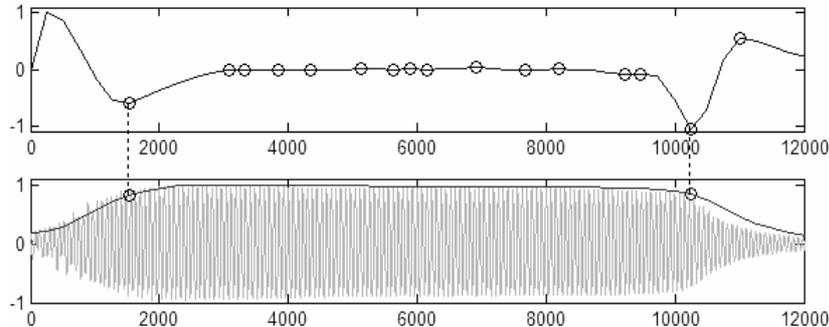


Figura 6: Deteção de final de ataque e início de decaimento a partir da energia RMS

O ataque é definido como o intervalo de tempo entre os instantes de início da nota e o final de ataque, enquanto o decaimento é definido como o intervalo de tempo entre os instantes de início de decaimento e final da nota. Quando o final de ataque coincide com o início do decaimento, a nota é considerada como sem sustentação.

Resultados

Os valores nos histograma abaixo correspondem à diferença em milissegundos entre a detecção automática e a média entre os instantes obtidos por detecção manual realizada por dois sujeitos.

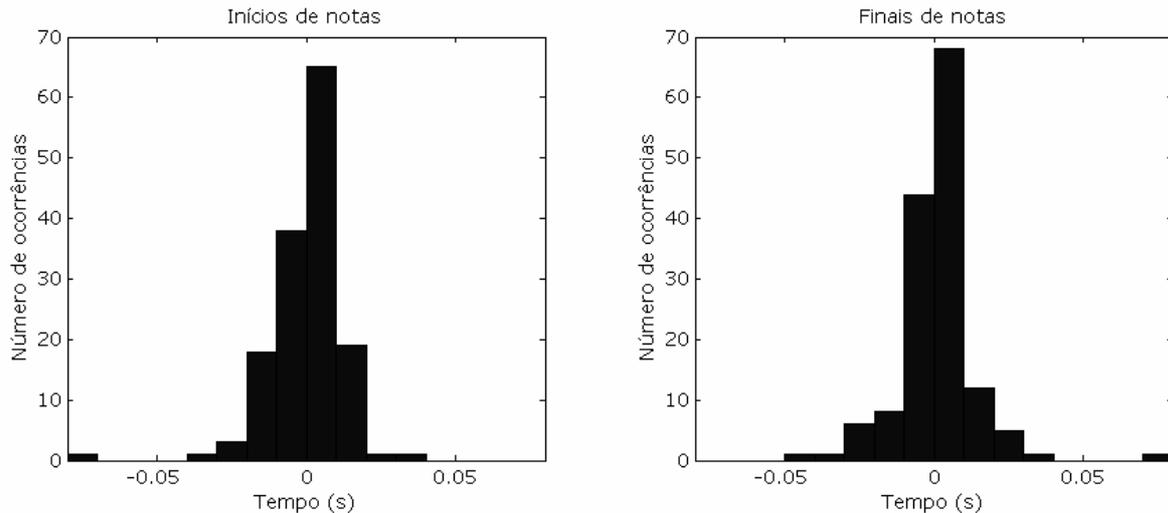


Figura 7: Diferença entre a detecção automática e a média entre duas detecções manuais

Como esperado, a segmentação a partir do fluxo espectral não mostrou precisão suficiente para a detecção do início e fim de nota, já que este parâmetro mede a sincronia da evolução da distribuição espectral ao longo do sinal, a qual é atingida somente após a região de transientes que ocorre antes do final de ataque. No entanto, acreditamos que o fluxo espectral pode ser associado ao instante de final de ataque, podendo ser utilizado na definição de descritores relacionados ao que se percebe no ataque (*descriptor de ataque*) ou nos intervalos de tempo situados entre notas consecutivas (*descriptor de transição, de articulação, de legato*).

A Figura 8 mostra a análise de um trecho do tema principal do primeiro movimento do Concerto para Clarineta e Orquestra em Lá Maior Kv 622 de Mozart. Percebe-se uma variação do fluxo espectral mais acentuada nas transições entre as notas ligadas lá-sol do que entre notas articuladas sol-fá, fá-mi e mi-mi do segundo compasso. Nota-se também a pequena variação de fluxo na transição mi-mi, devido à superposição do mesmo conteúdo espectral proveniente da reverberação da primeira nota mi, que não permite uma queda maior do fluxo. Percebe-se também uma maior variação do fluxo nas duas transições fá-ré do terceiro compasso do que nos intervalos referentes às pausas que antecedem e precedem as notas destas transições.

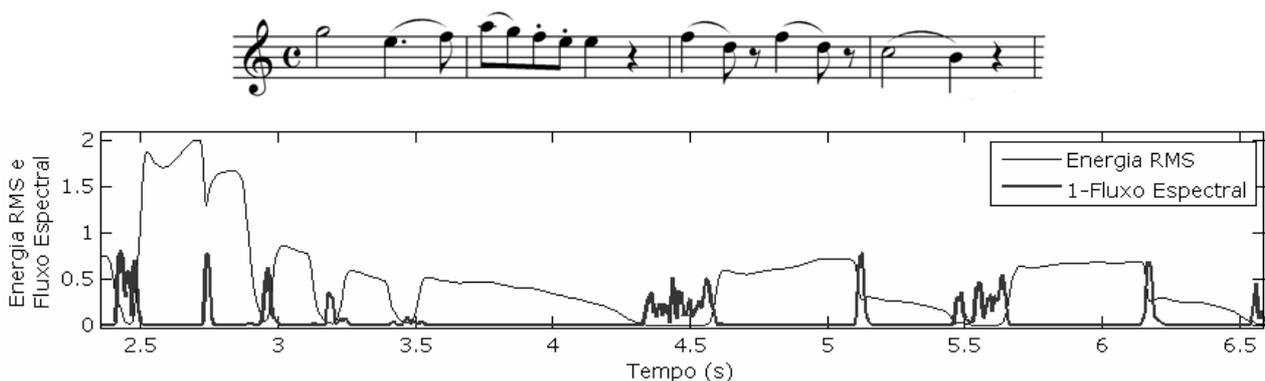


Figura 9: Variação do fluxo espectral nas regiões de transição entre notas consecutivas no Tema principal do primeiro movimento do Concerto para Clarineta e Orquestra em Lá Maior Kv 622 de Mozart

Conclusão

O método de segmentação descrito apresentou uma robustez suficiente para a os objetivos propostos neste estudo. Uma validação mais consistente está sendo elaborada a partir de testes subjetivos de detecção de inícios e finais de notas, cujo protocolo já foi testado com êxito em testes pilotos. O teste solicitará a sujeitos músicos profissionais e estudantes, determinar os instantes de início e fim de notas a

partir de um clique similar a uma batida de metrônomo, que pode ser deslizado ao longo do sinal de teste por meio de uma régua interativa.

Referências Bibliográficas

- Bello, J.P.; Daudet, L.; Abdallah, S. *et al.* (2005) Tutorial on Onset Detection in Music Signals. *IEEE Transactions on Speech and Audio Processing*, N° 5, Vol. 13, p. 1035 - 1047.
- De Poli, Giovanni.; Canazza, Sergio; Drioli, Carlo *et al.* (2005) Modeling and control of expressiveness in music performance. *Proceedings of the IEEE*. N° 4. Vol. 92. p. 686 -701.
- Gabrielsson, A. (1995). Expressive Intention and Performance. *Music, Mind and Machine*. New York: Springer. p. 35- 47, 1995.
- _____ (1999). Music Performance. *Psychology of music*. New York: Academic Press, p.506-602.
- _____ (2003). Music Performance Research at the Millenium. *Psychology of Music*. N° 3. Vol. 31, p. 221 - 272.
- Juslin, P. N. (1997). Emotional Communication in music performance: a functionalist perspective and some data. *Music Perception*. No 4. Vol.14, p.383-418.
- _____ (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human perception and performance*. N° 6. Vol. 26. p. 1797 - 1813.
- Park, T. H. *Towards Automatic Musical Instrument Timbre Recognition*. (Ph. D. Dissertation). Department of Music, Princeton University, 2004.
- Seashore, C. E. *Psychology of Music*. New York: McGraw-Hill, 1938.