DOI 10.20504/opus2017c2309

De Montserrat às ressonâncias do piano: uma análise com descritores de áudio

Danilo Rossetti
Jônatas Manzolli
(Universidade Estadual de Campinas, Campinas-SP)

Resumo: Neste artigo, pertencente à pesquisa sobre a elaboração de um modelo para a análise da música eletroacústica mista a partir de descritores de áudio, realizamos a análise de *Ressonâncias* (2015) para piano e sons eletrônicos de Jônatas Manzolli. Como fundamentação teórica, apresentamos uma abordagem sobre as Séries e a Transformada de Fourier, operação que serve de base para os processamentos eletroacústicos empregados na composição e para as análises realizadas a partir dos descritores. As hipóteses a serem verificadas dizem respeito à maleabilidade dos processamentos eletroacústicos empregados na peça, que podem fornecer resultados distintos em interpretações diferentes, além da existência de uma tendência de maior similaridade espectral no processo composicional realizado, em que o nível de variação entre duas interpretações é maior nas seções iniciais e menor nas seções finais. Como metodologia, analisamos as características espectrais dos sons compostos em suporte fixo e comparamos o timbre produzido (amálgama dos sons instrumentais e eletroacústicos) em duas interpretações diferentes, a partir dos dados obtidos nas análises. Concluímos discutindo dois aspectos: (1) sobre o nível de maleabilidade sonora atingido pelo compositor na estruturação do processo criativo, e (2) a análise desenvolvida aponta para a impossibilidade da notação musical tradicional de representar a completude dos resultados sonoros emergentes, que decorrem da interação entre os sons instrumentais e eletroacústicos. Ou seja, o modelo de análise estudado justifica a sua aplicabilidade no contexto da música eletroacústica mista atual, principalmente quando foca a fusão dinâmica da parte instrumental com o suporte fixo e/ou com o tratamento em tempo real.

Palavras-chave: Música eletroacústica mista. Análise musical. Descritores de áudio. Transformada de Fourier. *Phase vocoder source-filter*.

From Montserrat to Piano Resonances: An Analysis with Audio Descriptors

Abstract: As part of a research study of an audio descriptors model for the analysis of live-electronic music, this article addresses *Ressonâncias* (2015) for piano and electronic sounds by Jônatas Manzolli. As a theoretical basis, we approach the Fourier Series and Transform, an operation that serves as a basis for the electroacoustic processes employed in the composition and for the analyses performed from descriptors. The hypotheses to be verified are related to the malleability of the electroacoustic processes employed in the piece, which can deliver specific results for different performances, besides the existence of a tendency of greater spectral similarity in the compositional process, where the level of variation between two performances is higher in the initial sections and lower in the final sections. As a methodology, we analyze the spectral features of the fixed tape sounds and compare the timbre produced (the amalgam of instrumental and electroacoustic sounds) in two different performances from the data obtained from the analysis. We conclude by discussing two aspects: (1) the level of sonic malleability reached by the composer in structuring his creative process, and (2) the analysis that emphasizes the impossibility of traditional music notation to represent the plentitude of the emerging sound resulting from the interaction between instrumental and electroacoustic sounds. In other words, the applicability of the analytical model studied is relevant in the context of current live-electronic music, mainly when it addresses the dynamic fusion of the instrumental part with tape sounds and/or real-time treatments.

Keywords: Live-electronic music; musical analysis; audio descriptors; Fourier transform; phase vocoder source-filter.

ROSSETTI, Danilo; MANZOLLI, Jônatas. De Montserrat às ressonâncias do piano: uma análise com descritores de áudio. *Opus*, v. 23, n. 3, p. 193-221, dez. 2017. http://dx.doi.org/10.20504/opus2017c2309

Submetido em 17/08/2017, aprovado em 02/11/2017.

dos arcos, pedras,
catedrais e esquinas,
nos corpos e nas mentes
aproximam sussurros dos
ventos,
das gentes.

No contexto da música eletroacústica mista, a análise musical a partir de aspectos sonoros apresenta-se como um grande desafio. Esse se instaura quando o projeto de análise foca o entendimento das diferentes formas de integração sonora, as quais, em última instância, ocorrem somente no momento da execução da obra. Para abarcar as muitas dimensões sonoras que emergem como consequência direta desse discurso musical, propomos um novo método analítico que amplia a capacidade de descrição e detalhamento circunscritos na representação digital e na espectrografia do som. Ou seja, refletimos sobre a ampliação da questão musicológica já postulada quando o objeto de análise é a obra eletroacústica para suporte fixo (MALT, 2012. ROSSETTI, 2017a, 2017b). A esses conceitos levantados nessas referências, somam-se outros aspectos importantes que são derivados da ampliação do repertório de técnicas de escrita (PADOVANI; FERRAZ, 2011), do tratamento sonoro com suporte computacional (GERZSO; BOULEZ, 1988) e da abertura da estrutura da obra mista a partir da dinâmica da performance (MANZOLLI, 2013).

Neste artigo, apresentamos um método que se ancora em três aspectos: (1) os elementos da obra que foram compostos em suporte fixo, (2) aqueles gerados por execução instrumental de técnicas estendidas, e (3) outros produzidos a partir da fusão dinâmica da parte instrumental com o suporte fixo e/ou com o tratamento em tempo real. Dos três pontos apontados nos parágrafos acima, este artigo trata dos aspectos (1) e (3). Uma abordagem específica sobre a interação entre sons instrumentais produzidos por técnicas estendidas (aspecto 2) e suporte eletroacústico é encontrada no trabalho de Simurra e Manzolli (2016a), Rossetti (2017a) e Rossetti, Teixeira e Manzolli (2017).

A nossa tese é que podemos avançar no entendimento desses pontos se utilizarmos medidas matemático-estatísticas extraídas dos componentes espectrais do som. Propomos uma abordagem que poderá gerar elementos quantitativos que, então, serão úteis ao escopo geral qualitativo da análise musical. Essas medidas são conhecidas e disponíveis na literatura do campo de pesquisa conhecido como *MIR* (*Music Information Retrieval*) (PEETERS, 2004. PEETERS et al., 2011). Ou seja, a possibilidade de recuperação de informação musical fomentou o desenvolvimento de descritores de áudio, os quais são justamente as medidas estatísticas que interessam ao nosso modelo.

Os descritores são algoritmos que extraem informações de gravações musicais digitalizadas de modo a estabelecer relações entre as características sonoras que desejamos analisar com dados numéricos ou gráficos. A utilização dessas ferramentas na análise musical a partir da gravação do áudio pode apresentar novas perspectivas, independentemente das técnicas de composição utilizadas. Tradicionalmente, os descritores são empregados na análise de gravações de obras instrumentais e vocais com o intuito de comparar questões sobre andamento, contrastes de dinâmica ou intensidade dos modos de ataque de alturas musicais entre diferentes gravações.

Assim, a partir dos dados extraídos pelos descritores, torna-se possível compreender questões a respeito da percepção sonora, como apresentado em Peeters (2004), categorizar estilos musicais em banco de dados (PEREIRA, 2009) e já surgem aplicações no campo da análise musical com suporte computacional (MALT; JOURDAN, 2008, 2009a, 2009b). O Núcleo Interdisciplinar de Comunicação Sonora da UNICAMP desenvolve pesquisa sobre o tema que resultou nos trabalhos de Monteiro (2012); Simurra e Manzolli (2016a, 2016b) e Rossetti, Teixeira e Manzolli (2017). Em suma, o objetivo geral de nossa pesquisa, da qual esse artigo faz um recorte, é construir e aplicar esse novo modelo para analisar obras recentes do repertório da música eletroacústica mista.

A utilização dos descritores na análise musical justifica-se por diferentes razões. Enumeramos três delas a seguir:

(1) Análises baseadas apenas na partitura não representam a complexidade sonora atingida na performance, mesmo em obras que utilizam técnicas instrumentais e composicionais tradicionais. Como afirmado acima, os descritores permitem comparar diferentes gravações de uma mesma obra, possibilitando realçar nuances interpretativas entre elas. Portanto, dentro dessa abordagem, o resultado sonoro atingido por uma interpretação ganharia um *status* equivalente ao de uma análise tradicional da mesma obra calcada fundamentalmente na partitura.

(2) No caso da música contemporânea, em que são empregadas técnicas instrumentais estendidas, uma análise a partir de descritores de áudio também é desejável. Essas técnicas produzem "novos sons" que são associados a características transientes e ruidosas que não fazem parte dos modelos vibratórios já estudados para os instrumentos musicais. Produzir um estalo no tampo do violoncelo não faz parte do repertório acústico e técnico já descrito para esse instrumento de corda. Por essa razão, uma análise baseada apenas na partitura não representa a totalidade do resultado sonoro atingido por obras dessa natureza.

(3) No caso da música eletroacústica mista, a importância em se utilizar descritores de áudio na análise se deve ao fato de que a interação entre os sons compostos em tempo diferido (ou suporte fixo) e a parte dinâmica processada em tempo real produz sonoridades emergentes. Essas novamente extrapolam a notação musical descrita na partitura. A contribuição da pesquisa, aqui reportada, apoia-se na tese de que a análise musical assessorada por descritores tem o potencial de detectar e descrever alguns desses processos emergentes que se estabelecem entre as diversas camadas sonoras que normalmente compõem as peças mistas.

Ressonâncias (2015) é uma obra de Jônatas Manzolli para piano, sons eletrônicos e imagens[1]. No processo composicional, Manzolli buscou uma interação imbricada entre a parte instrumental e os sons eletroacústicos difundidos. No total, durante a performance da peça, temos três camadas de sons de diferentes origens que são executadas simultaneamente: (1) o som acústico do piano; (2) a filtragem do som do piano captado por microfones pelos sons armazenados em um *buffer*[2] (processamento do *phase vocoder source-filter*, discutido e analisado numa seção específica do artigo); e (3) sons em suporte fixo compostos previamente em estúdio. Em relação às interações produzidas entre os sons instrumentais e eletroacústicos, ressaltamos o caso específico da segunda camada, em que o timbre do som captado do piano influencia em tempo real a qualidade do timbre dos sons eletroacústicos gerados pela filtragem do *phase vocoder source-filter*. Outro fato que é recorrente dentro da prática da música mista e que se encaixa nesse contexto interativo acontece

[1] Em nossa análise, não trataremos de questões sobre as imagens projetadas durante a performance da obra.
[2] Memória computacional física usada para armazenar temporariamente dados em geral.

em relação ao tempo musical conduzido pelo intérprete do instrumento, que é diretamente influenciado pela escuta do tempo de ressonância dos sons eletroacústicos.

Outro ponto importante a respeito da composição de *Ressonâncias* se refere à operação conhecida como Transformada de Fourier, que, segundo o compositor, se relaciona com a obra de duas maneiras: a primeira como a principal operação na qual se baseiam seus processamentos eletroacústicos (*phase vocoder source-filter*), e a segunda de forma metafórica e imagética. Ressaltamos ainda que a Transformada também é uma das operações de base para o cálculo dos dados fornecidos pelos descritores de áudio, procedimento que abordaremos em detalhes em seção específica do artigo.

A Transformada não deixa de ser um processo de ressonância no qual o espectrograma apresenta uma imagem dos componentes espectrais. Na Fig. 1a vê-se uma foto de Monserrat, maciço rochoso situado na região da Catalunha, local que inspirou Manzolli a criar a obra e o poema em epígrafe no início do artigo; na Fig. 1b vê-se a imagem tridimensional de um espectro sonoro. Comparando as duas imagens, é possível observar semelhanças nas suas formas e, assim, interpretá-las a partir da intuição poética do compositor: o maciço rochoso se apresenta como uma grande ressonância que tem um perfil semelhante a um sonograma tridimensional. Citando a teoria de Gilbert Simondon sobre o modo de existência dos objetos técnicos, evocamos sua afirmação de que na medida em que um objeto técnico é empregado num determinado sistema, torna-se portador de forma e promove um acoplamento e pertencimento mútuo entre matéria e forma (2012 [1958]: 236). Assim, pode-se estabelecer um paralelo entre o objeto técnico, nesse caso a Transformada, e a forma musical resultante do acoplamento entre a matéria sonora e o processamento computacional.

(a)



(b)

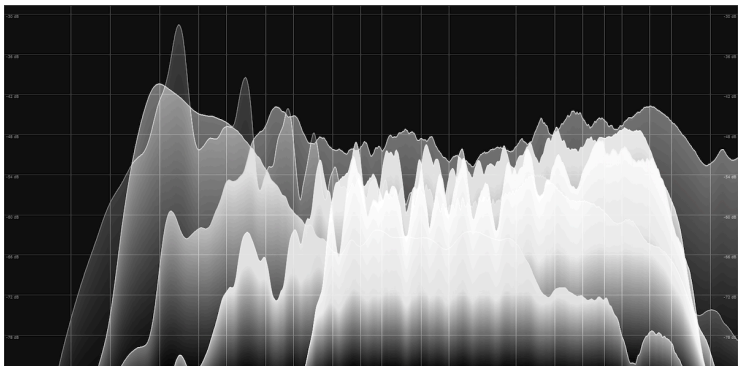


Fig. 1: Maciço rochoso de Montserrat (a) e representação tridimensional de um espectro sonoro (b).

Na próxima seção, desenvolvemos uma introdução teórica sobre a Transformada de Fourier. Nas seções seguintes analisamos o processamento do *phase vocoder source-filter* e o modelo computacional criado para a obra, no qual o autor implementou um patch em *PureData*. Estabelecidas as premissas teóricas e a modelização computacional, as próximas seções focam a análise da obra.

Premissas teóricas

A operação matemática que possibilita a decomposição de um sinal periódico em parciais de frequência e amplitude é denominada Transformada de Fourier. Essa ferramenta matemática cuja utilização primordial é a solução de equações diferenciais parciais tem vasta gama de aplicações em Física, teoria dos sinais, processamento de sinal e, particularmente, nos interessa o processamento de sinais musicais. A seguir, apresentamos uma breve introdução teórica às séries e à Transformada de Fourier. Para mais informações sobre suas propriedades e aplicações, cf. Oppenheim, Schafer, Buck (1999).

As chamadas séries de Fourier foram criadas pelo matemático e físico francês Jean-Baptiste Joseph Fourier (1768-1830) com o intuito de representar e analisar fenômenos periódicos tais como o fluxo de calor entre moléculas adjacentes, o movimento das marés, ou os componentes de um som. Num segundo momento, essa análise foi estendida para fenômenos não periódicos, que puderam ser descritos pela Transformada de Fourier. Esse tipo de representação e análise foi desenvolvida a partir da constatação de que cada sinal periódico possui um espectro de frequências (que determina suas características) e pode ser analisado tanto nos domínios temporal como frequencial (OSGOOD, 2007: 1-2. TEMPELAARS, 1996: 118-120).

As séries de Fourier mais simples são representadas por funções periódicas, as quais exibem um padrão regular que se repete, seja no fenômeno em si, seja em relação a seus componentes[3]. O período (*T*) de um fenômeno periódico é a sua variável temporal, medida em unidades de tempo (segundos, minutos, horas, dias etc.), já as suas periodicidades internas são medidas por sua frequência, cujas dimensões são normalmente indicadas em ciclos por segundo (Hertz). Sua periodicidade no espaço é dada pelo comprimento de onda (λ), sendo o seu inverso (1/ λ) o valor frequencial em Hz. A partir dessa relação, chega-se às seguintes relações primordiais: quanto maior a frequência, menor o tamanho da onda; por outro lado, quanto menor a frequência, maior o seu comprimento de onda (OSGOOD, 2007: 3-4).

No âmbito sonoro, relaciona-se a propriedade física das frequências com a percepção das alturas musicais. No caso das funções periódicas, as frequências correspondem aos múltiplos inteiros de uma frequência fundamental (os harmônicos no caso da representação de um som). Outro ponto importante, a partir da observação do comportamento de sinais diversos, é possível extrair uma importantíssima relação recíproca entre os domínios temporal e frequencial: quanto maior for o período no tempo, menor será o espaçamento de seu espectro, ou, ainda, quanto menor for o período no tempo, maior será o espaçamento do seu espectro (OSGOOD, 2007: 8-16).

[3] As séries de Fourier periódicas são definidas pela expressão: $f(t + nT) = f(t)$, onde a função $f(t)$ é periódica em relação ao período T quando $T > 0$, para todos os valores de t.

A Transformada (ou análise) de Fourier é uma operação que consiste na divisão de uma determinada função em inúmeros componentes senoidais. O inverso, a adição (síntese aditiva[4]) de componentes senoidais formando uma função (onda) "complexa" (não senoidal) é chamada de Transformada Inversa de Fourier[5]. No caso de um sinal completamente periódico[6], a análise de um único período já é suficiente para determinar as características do sinal, pois seus componentes senoidais (harmônicos ou parciais) têm frequências que são múltiplos da frequência fundamental ($f_0 = 1/T$) e se repetem a cada período (TEMPELAARS, 1996: 128-129). No caso da análise de Fourier aplicada a sinais não periódicos, um impulso não periódico é aproximado por um impulso periódico com período de duração infinita[7] (TEMPELAARS, 1996: 139).

De maneira prática, a fim de analisarmos sinais não periódicos com janelas de duração menores, é utilizada a operação de "janelamento" (*windowning*). Nela, são suprimidos componentes não desejados do som, facilitando o cálculo da Transformada Inversa de Fourier (processo de síntese). Para essa operação, são comumente utilizadas janelas com envelope senoidal, tais como a *Hanning Window*, sendo também possível a utilização de janelas como *Hamming*, *Kaiser* e *Gauss* (TEMPELAARS, 1996: 146-147). Na Fig. 2 mostramos graficamente um sinal senoidal periódico (à esquerda), o envelope de uma janela *hanning* (ao centro) e o resultado do efeito de janelamento (à direita).

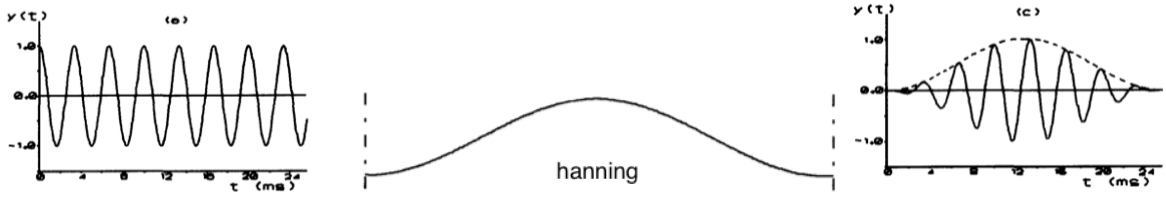


Fig. 2: Modulação de um sinal de áudio e envelope de uma janela *hanning* (TEMPELAARS, 1996: 146-147).

[4] A síntese aditiva consiste na formação de um agregado sonoro através da superposição (soma) de componentes periódicos que são organizados a partir de um som fundamental (ROSSETTI, 2016: 64).

[5] Um ponto a ser destacado é que a análise de Fourier não determina qual deve ser o período (*T*) utilizado na análise, ou seja, o que é chamado de "*Janela de Análise*". Nesse sentido, a escolha do período depende do que sabemos do sinal estudado. Se ele for completamente periódico, *T* deve ser igual à duração do período, constituindo uma análise de altura sincrônica (*pitch synchronous analysis*). Se o sinal é aperiódico, *T* deve ter uma duração muito longa ou infinita, dessa feita a duração total do sinal constitui um único período.

[6] A descrição de uma função periódica pela Transformada de Fourier consiste numa lista de amplitudes e ângulos de fase iniciais (relativos aos parciais), ou seja, as amplitudes e as fases do espectro da forma de onda. A equação que representa a análise de um sinal periódico, tendo a_n e b_n como valores dados em números reais e discretos, é a seguinte:

$$y\,(t) = \sum_{n\text{-}1}^{N} (a_n \cos(2\pi nt) + b_n sen(2\pi nt))$$

[7] No caso de sinais não periódicos, na definição da série de Fourier, a soma dos componentes espectrais é substituída por uma integral. Os coeficientes discretos a_n e b_n da fórmula anterior são substituídos por funções frequenciais contínuas a(*f*) e b (*f*), consistindo na seguinte equação:

$$y\,(t) = \int_0^{\infty} \big(a(f)\cos 2\pi\, ft + b(f) sen\, 2\pi ft\big) df$$

Na Fig. 2, temos à esquerda um sinal senoidal completamente periódico, de uma determinada frequência (*f*) com amplitude constante e totalizando 24ms de duração, que é modulado em termos de amplitude por uma janela de envelope senoidal do tipo *Hanning*. No gráfico da direita temos o resultado dessa modulação, com o sinal periódico moldado de acordo com o envelope de amplitude da janela *Hanning*. No seu início, a amplitude é 0, crescendo até o valor 1 (por volta de 12ms), e voltando a decrescer gradualmente até 0 no final da janela. Através dessa operação, evita-se quebras em locais não desejados (de amplitude superior a 0), que resultam em deformações na forma de onda do sinal, proporcionando uma "suavização" de intensidade nos locais dos cortes ou segmentações utilizadas no processo de análise.

Uma maneira otimizada e mais rápida de realizar a Transformada de Fourier foi implementada por Cooley e Tukey (1965). Esse novo método foi denominado Transformada Rápida de Fourier (*Fast Fourier Transform"* (*FFT*)), que reduz significativamente a quantidade de cálculos para a determinação dos valores de amplitude e de fase dos componentes dos sinais analisados[8]. Na figura a seguir, apresentamos um diagrama simplificado de como o processo de *FFT* é realizado (baseado na representação de Curtis Roads, 1996) e aplicado em ressíntese sonora. Temos, no sentido da esquerda para a direita, um sinal de entrada, sua segmentação em janelas de tempo e a operação da Transformada Rápida de Fourier aplicada a cada janela individualmente, resultando nos valores de magnitude e fase do espectro. A magnitude indica a amplitude de cada parcial frequencial analisado e a informação da fase corresponde ao seu valor inicial para cada parcial. A partir desses valores, por exemplo, o *phase vocoder* pode realizar as transformações espectrais e temporais desejadas que culminam na ressíntese de um novo som com o novo timbre.

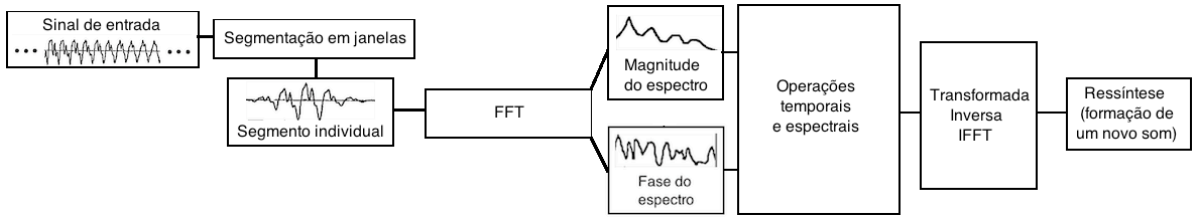


Fig. 3: Diagrama de blocos da Transformada Rápida de Fourier a partir do modelo de Roads (1996: 551).

No caso da utilização de descritores de áudio como ferramenta de análise musical, a *FFT* é empregada dentro do processo de extração dos valores fornecidos pelos descritores. O artigo de Peeters (2004: 2) nos fornece um diagrama de blocos (Fig. 4), especificando como o sinal sonoro é analisado nesse processo, de maneira similar à figura anterior. A partir de um sinal de entrada, temos sua segmentação em janelas de tempo, determinando os segmentos individuais. Nesse ponto é possível extrair descritores temporais, tais como a detecção do início e fim de ataques executados por instrumentos musicais, ou o trecho do regime de sustentação de determinados sons. De maneira geral, realizada a *FFT*, podemos extrair

[8] Para um detalhamento dos procedimentos para redução dos cálculos empregados na *FFT*, cf. Tempelaars (1996: 150-154) e Osgood (2007: 279-287).

informações a respeito dos descritores espectrais, como o Centroide e o Fluxo Espectral. Após a *FFT* opera-se sobre os componentes espectrais, através do modelo harmônico-senoidal, a fim de obtermos valores numéricos que representam as curvas desses descritores. Se o interesse da análise são aspectos harmônicos, pode-se realizar operações para detectar a Frequência Fundamental, descrever a Harmonicidade ou Inarmonicidade, entre outros. Outro exemplo é modular as amplitudes do espectro a partir de modelos psicoacústicos, como é o caso do *Loudness*. Em nossa análise utilizaremos o Centroide, o Fluxo Espectral e o *Loudness*, como será apresentado mais adiante.

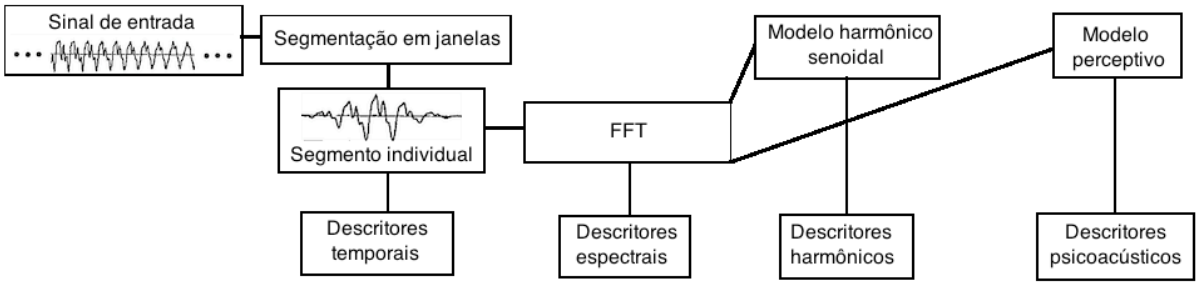


Fig. 4: Detalhes do processo de extração de informações de descritores temporais, espectrais, harmônicos e psicoacústicos, baseado em Peeters (2004: 2).

Feita essa introdução geral à Transformada, recomendamos ao leitor interessado a leitura de Tempelaars (1996). Trata-se de uma referência fundamental dedicada ao ensino de processamento de sinais para músicos. Apresentaremos, a seguir, o método de processamento denominado "*phase vocoder*", largamente utilizado na prática da música eletroacústica, que se baseia no processamento de sinais sonoros a partir da Transformada Rápida de Fourier (*FFT*). Discutimos também a conceito de "*source-filter*", que, por sua vez, é um tipo de síntese cruzada baseada na multiplicação de dois sinais de áudio.

O *phase vocoder* e seu modelo *source-filter*

Introduzimos nessa seção os fundamentos do processamento sonoro utilizado na obra analisada, o *phase vocoder source-filter*. Em linhas gerais, o tratamento desenvolvido envolve dois sons diferentes, nos quais o envelope temporal de um deles é estampado ou copiado no espectro frequencial do outro. Assim, cria-se um terceiro som que contém as características temporais e frequenciais dos dois sons de entrada. Esse processamento é um tipo de convolução espectral que opera a partir dos dados obtidos através da operação de *FFT* dos dois sons envolvidos. Temos, dessa forma, a multiplicação dos dados relativos ao domínio temporal de um som pelos dados do domínio frequencial do outro, correlacionando as informações e gerando um novo som a partir da Transformada Inversa (*IFFT*) ou ressíntese (ROSSETTI; FERRAZ, 2016: 83-84).

O *phase vocoder source-filter*, especificamente, pode fornecer como resultado um sinal com características idênticas ao sinal de entrada, ou promover modificações em seu envelope temporal (*time-stretching*) ou no seu espectro de frequências (*pitch-shifting*). As modificações temporais se referem à contração ou dilatação da duração do som de entrada e as

modificações frequenciais se referem à transposição do espectro inicial para regiões mais agudas ou mais graves (MOORER, 1978. DOLSON, 1986).

O chamado "*source-filter*" é um tipo de *phase vocoder* que opera com dois sinais de entrada, sendo o primeiro um sinal fonte (*source*), que é analisado espectralmente com a Transformada, e o segundo um sinal que funciona como filtro (*filter*), também analisado via *FFT*, que modula o sinal fonte com seu envelope espectro-temporal. Como mencionamos, esse processo é um tipo de convolução espectral ou síntese cruzada entre dois sinais de áudio, também denominado *timbre stamp*, já que a marca do envelope espectro-temporal do filtro é estampada no sinal fonte, sendo possível identificá-lo auditivamente. Juntando os dois processamentos, afirma-se que o "*phase vocoder source-filter*", resumidamente, funciona com a multiplicação dos valores obtidos pela Transformada do sinal Fonte (amplitude e fase) pelos valores do sinal Filtro, obtidos com a mesma operação (ROEBEL; LIUNI, 2013: 77-78. PUCKETTE, 2006: 278-279).

A sequência de passos envolvidos é a seguinte: (1) o *source-filter* inicia-se com dois sinais sonoros de entrada; (2) esses sinais são segmentados em janelas de curta duração (512, 1024, 2048, 4096 *samples*[9]), ou seja, realiza-se a multiplicação do sinal de entrada por uma janela do tipo *Hanning*; (3) aplica-se no sinal janelado a *FFT* e, desta forma, obtém-se os valores de amplitude e de fase dos componentes do sinal; (4) multiplicação dos valores frequenciais do som fonte pelos valores temporais do som filtro; (5) a Transformada Inversa de Fourier (*IFFT*) é aplicada, configurando-se na síntese que gera um novo sinal de áudio; (6) realiza-se novamente o janelamento do novo sinal; (7) é feita a justaposição temporal das janelas (*overlapping*) na geração do som de saída. Na Fig. 5 há a representação gráfica dos passos desse processamento, e, mais à frente, apresentamos esses mesmos passos num *patch* implementado em *PureData*, vide Fig. 8.

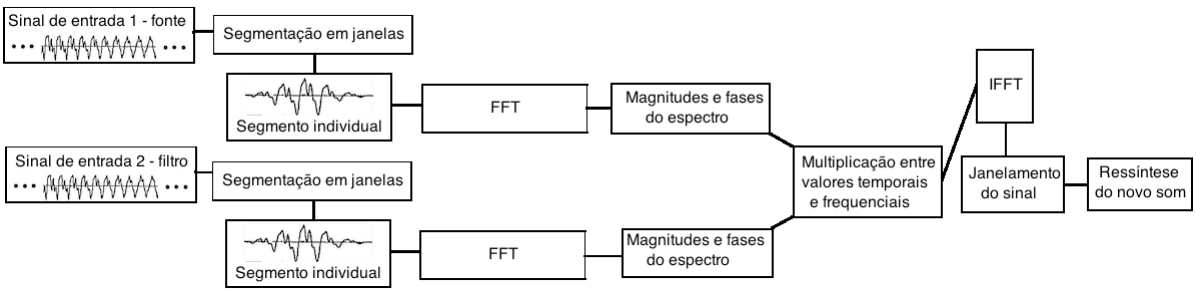


Fig. 5: Diagrama de blocos das etapas de um processo de convolução espectral ou *source-filter*.

Esse tipo de filtragem faz com que o som do piano "ressoe" ou "adquira as características de ressonância" contidas nos parciais do espectro do som armazenado. Aqui, temos uma referência à noção de modulação, no sentido de "moldar a forma", proposta por Gilbert Simondon, em sua tese sobre a individuação das formas (2005 [1958]: 47-48). A matéria, enquanto suporte de energia potencial, toma a sua forma quando se encontra em um

[9] Os *samples* são as amostras de áudio tomadas no momento da sua conversão analógico-digital. Por exemplo, com uma taxa de amostragem de 16 bit/44.100Hz (padrão do CD), temos 44.100 amostras do sinal de áudio em um segundo com uma definição de 2^{16} em termos do espectro de frequências, totalizando 65.536 valores.

estado de ressonância interna completa. Como exemplo, apresentamos um processamento de *phase vocoder source-filter* realizado comparando as formas de onda de um som fonte, o trecho de *Ressonâncias* entre os compassos 31 e 34 (Fig. 6), que é modulado pelo som *SussurroDrone2* (*tape* em suporte fixo), o som filtro. Como resultado, temos a geração de um terceiro som proveniente da filtragem ou convolução espectral do primeiro pelo segundo. Na Fig. 7 observa-se as diferenças entre as formas de onda e os sonogramas de cada som.

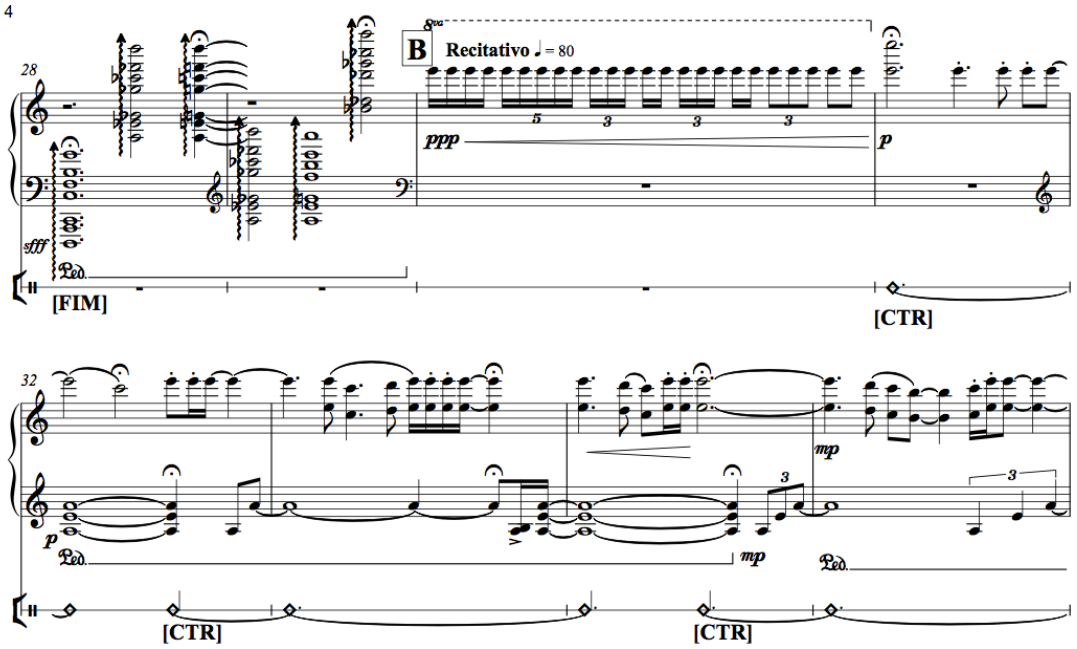


Fig. 6: Trecho da partitura referente ao som fonte analisado (c. 31-34).
Faixa de áudio: <http://git.nics.unicamp.br/danilo.rossetti/ressonancias/blob/master/áudio/1.%20Fig.%206-c.28-34-gravação2.mp3>

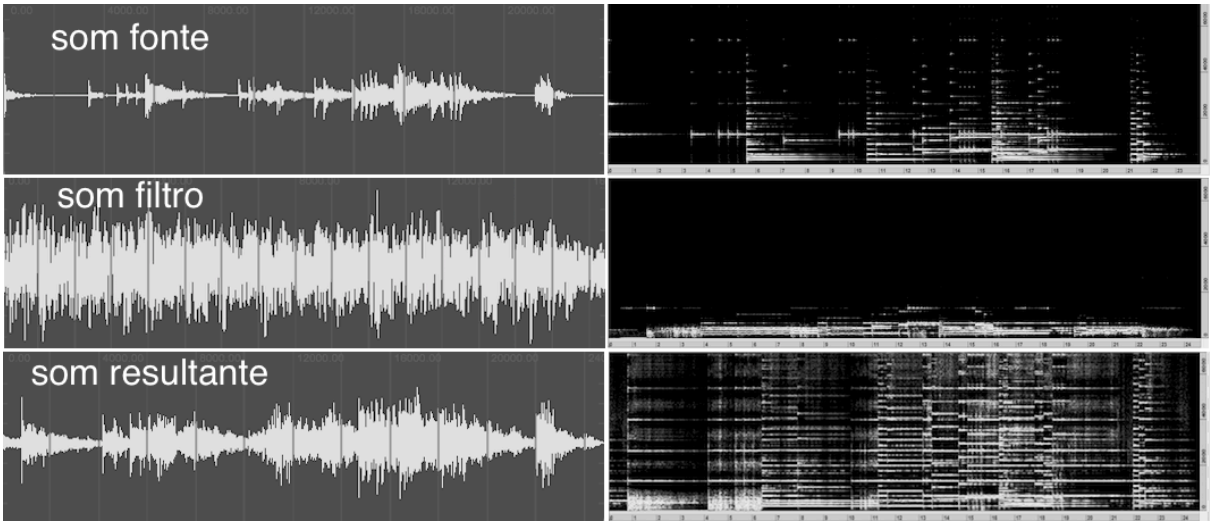


Fig. 7: Processamento do *phase vocoder source-filter:* formas de onda (esq.) e sonogramas (dir.) do som fonte, som filtro e o som resultante do proceso.

Introduzidos esses fundamentos, apresentamos a seguir o processamento do tipo "*source-filter*" implementado pelo compositor em *PureData* para *Ressonâncias*. Além desse modelo computacional, discutimos também a visão geral do conjunto de tratamentos eletroacústicos e dos *patches* implementados pelo compositor.

O modelo computacional de *Ressonâncias*

Em *Ressonâncias*, o *phase vocoder source-filter* realiza processos de contração e dilatação temporal do som do piano captado ao vivo (fonte), que é modulado por quatro sons compostos previamente em estúdio (filtro), denominados *SussurroDrone* (1 a 4). Esses sons são atribuídos a *presets* dentro do *patch* da peça e são disparados ao longo da performance pelo intérprete da parte eletrônica. A Fig. 8 apresenta o *patch*[10] implementado pelo compositor em *PureData*, que é baseado no modelo *timbre stamp* de Puckette (2006: 291).

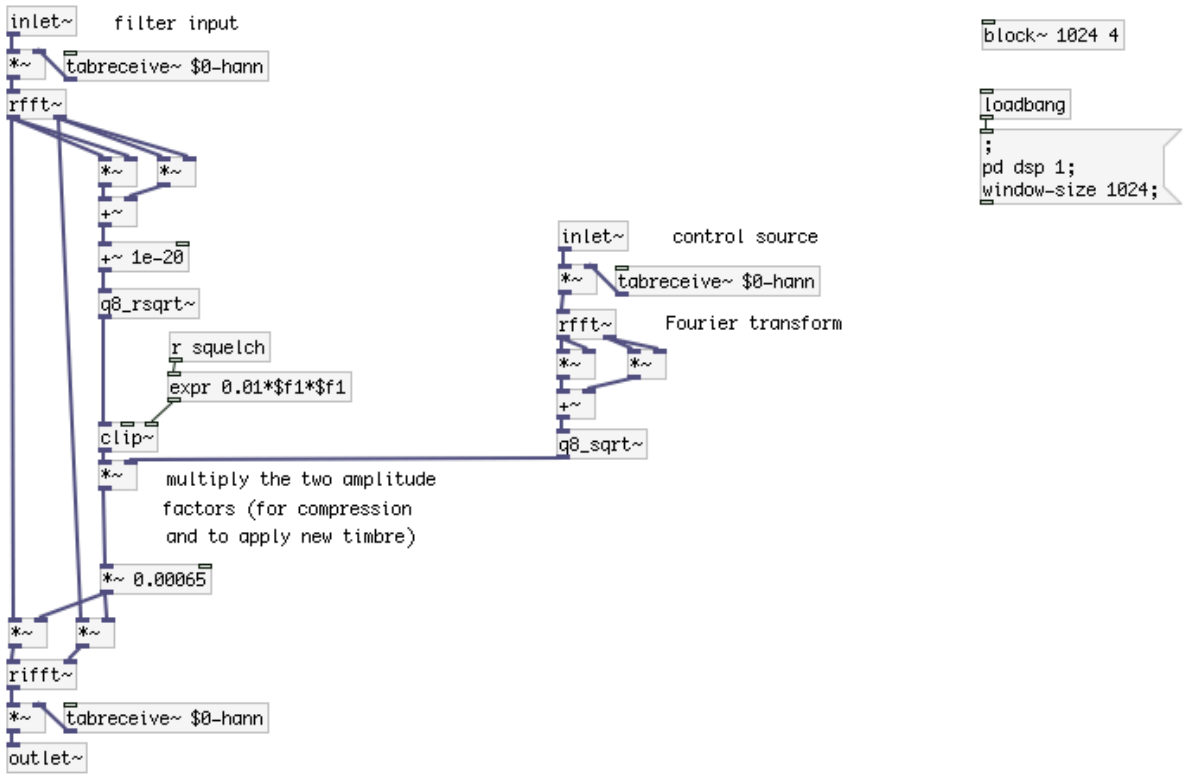


Fig. 8: *Phase vocoder source-filter* do *patch* de *Ressonâncias*.

Na Fig. 8 à direita, temos o tamanho janela de análise (*window size*), que é de 1024 *samples*. Há duas entradas de som (*inlets~*) referentes ao filtro (*filter input*) e ao som fonte (*control source*). Os dois sinais são modulados por janelas *hanning* (*tabreceive~ 0$-hann*) e submetidos à Transformada de Fourier pelo objeto *rfft~*. Os valores reais e imaginários (a saída da

[10] O *patch* elaborado por Manzolli que realiza os processamentos sonoros em tempo real e suas abstrações estão disponíveis no endereço eletrônico <http://git.nics.unicamp.br/danilo.rossetti/ressonancias>.

Transformada de Fourier fornece pares de números) são multiplicados entre si depois de somados e têm sua raiz quadrada extraída de amplitude para se obter o valor do módulo ou magnitude das parciais do espectro, ou seja, obtém-se o valor da magnitude de cada componente de frequência do espectro. Posteriormente, esses valores são submetidos a outra operação para seu nivelamento ao valor máximo pelo objeto *clip~*, obtendo-se valores mapeados entre 0 e 100. Verifica-se que essa operação é aplicada aos dois sinais. A seguir, os valores de amplitude espectral são multiplicados entre si, efetuando-se, assim, a síntese cruzada. Finalmente, o resultado da operação é submetido à Transformada Inversa de Fourier pelo objeto *rifft~*, à qual se segue o janelamento que é realizado antes do novo sinal ser enviado à saída do sistema (*outlet~*). Enfatizamos que os passos implementados no *patch* são os mesmos que já foram descritos na recapitulação teórica.

É importante ressaltar que o *phase vocoder source-filter* é apenas um dos processamentos realizados em *Ressonâncias*. Há também processos de granulação dos sons chamados de *SussurroDrone*, a extração da frequência fundamental do som do piano pelo objeto *fiddle~* (detecção de altura fundamental), cujos valores controlam o tamanho do grão da granulação, e a mixagem final do som direto do piano amplificado, do som gerado pelo *source-filter* e dos sons compostos em tempo diferido, denominados *Mix* (1 a 6). Na Fig. 9, temos o diagrama de blocos de todos os processamentos da obra.

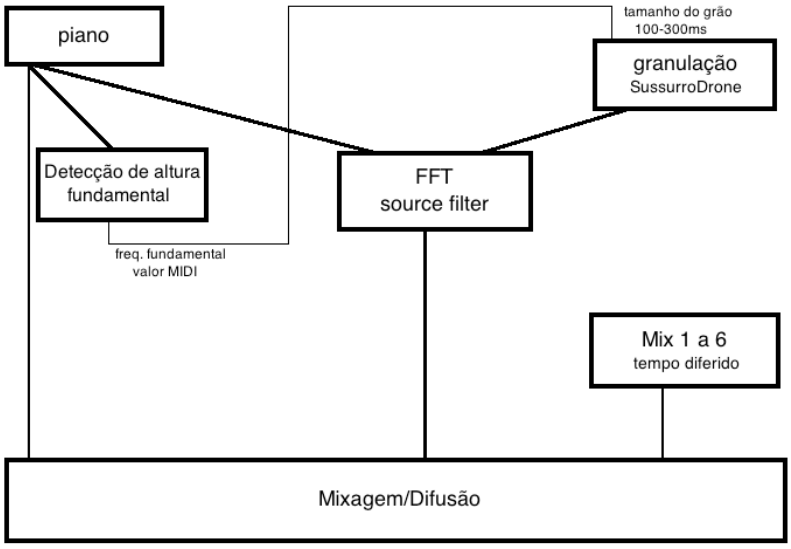


Fig. 9: Diagrama de blocos de todos os processamentos implementados para a obra *Ressonâncias*.

Ao observarmos a Fig. 9, percebemos que há ao menos três camadas de sons sobrepostas, e essas são manipuladas em tempo real pelo intérprete eletroacústico e mixadas antes da difusão: o som amplificado do piano, o som oriundo do *phase vocoder source-filter* e os sons compostos em tempo diferido (*Mix* 1 a 6), estes últimos disparados ao longo da performance. O que merece ser destacado em relação ao diagrama da Fig. 9 é a configuração de uma rede de objetos, em que valores dinâmicos são extraídos do som captado do piano para realizarem operações em tempo real sobre os parâmetros da granulação, e, consequentemente, sobre o resultado do som processado no *source-filter*.

O objeto *fiddle~*, ao qual o som captado ao vivo do piano é endereçado, estima a altura (em valores MIDI[11]) e a intensidade desse som, gerando uma lista de valores numéricos. Utilizando-se de um cálculo proporcional, o compositor mapeia os valores da tabela MIDI em valores entre 100 e 300. Esses últimos determinam o tamanho dos grãos (em milissegundos) do processo de granulação implementado. A granulação, por sua vez, é realizada sobre os quatro sons denominados *SussurroDrone*, criados previamente em estúdio e que são acionados individualmente durante a performance, de acordo com o *preset* disparado pelo intérprete eletroacústico. Como mencionamos, o *phase vocoder source-filter* tem como entrada o som do piano, e as alturas executadas pelo pianista influenciam na granulação. Ou seja, o intérprete instrumental interage nas "duas pontas" do processo (sinal Fonte e Filtro); portanto, influencia consideravelmente o resultado da síntese cruzada ou convolução espectral.

Por fim, há a difusão dos sons denominados *Mix* que são disparados e mixados junto com os outros dois sons processados durante a performance, função realizada pelo intérprete eletroacústico. Os seis sons *Mix* também dependem do *preset* escolhido por esse intérprete. Na Fig. 10, mostramos a interface do *patch* de *Ressonâncias*. Nela visualizam-se os controles sonoros operados durante a performance.



Fig. 10: Interface visual do *patch* de *Ressonâncias*.

Numa breve explicação dos controles, o intérprete eletroacústico controla a intensidade sonora de três fontes: o som captado pelo microfone (*input_mic*), som difundido (*output_difusão*) e efeito do *phase vocoder source-filter* (*output_effect*). Os controles numéricos de 0 a 15 devem ser disparados em pontos determinados na partitura, produzindo o disparo dos efeitos programados e associados a cada botão.

Encerrada a descrição do modelo computacional e dos processos criados para *Ressonâncias*, na próxima seção iniciamos a análise da peça junto com a apresentação do modelo analítico proposto. Nesse modelo, utilizamos conjunto de quatro descritores de áudio denominados *Centroide, Fluxo Espectral, Loudness* e *BStD*.

[11] O protocolo MIDI (*Musical Instrument Digital Interface*) trabalha com 128 valores (0 a 127). O valor 60 é atribuído ao Dó central (261Hz).

Ferramentas e metodologia de análise

Para a análise das obras eletroacústicas mistas que desenvolvemos ao longo de nossa pesquisa, implementamos um modelo analítico que se baseia na utilização de diferentes tipos de descritores paramétricos de áudio, os quais são selecionados tendo em vista peculiaridades ou características relacionadas à hipótese sobre a natureza sonora da obra analisada. Cabe aqui ressaltar que o processo de escolha de descritores vem da reflexão qualitativa musical do analista, da sua experiência com o repertório e da projeção ou hipótese de análise que cabe aos interesses musicológicos em questão. Desta forma, trata-se de um processo de análise assistida pelo computador no qual há espaço para especulações e aberturas. Por fim, cabe dizer que a partir do método desenvolvido não se busca um automatismo, mas, sim, uma convivência simbiótica com outros fatores musicais. Esses podem ser enriquecidos com a aplicação dos descritores que nos fornecem informações temporais, espectrais e psicoacústicas (relacionadas às características da percepção do som).

Na análise de *Ressonâncias*, nosso conjunto analítico constitui-se por quatro descritores, a saber, *Centroide Espectral, Fluxo Espectral, Loudness* e BStD (*Brightness, Standard Deviation*[12]). Cabe ressaltar que todos eles são calculados a partir das parciais de amplitude e frequência geradas através da Transformada de Fourier do sinal musical analisado. Para o emprego desses descritores, foi utilizada a biblioteca *PDescriptors*, implementada no software *PureData* pelos pesquisadores do NICS Adriano Monteiro (MONTEIRO, 2012) e revisada por Gabriel Rimoldi. Também foram utilizados descritores de baixo nível (*low level features*) disponíveis no *software Sonic Visualiser*, desenvolvido pela *Queen University of London*[13]. A seguir, apresentamos as definições desses descritores utilizados em nosso modelo analítico.

O Centroide Espectral $C_{(i)}$, de acordo com Agostini, Longari e Pollastri (2003: 7), é o baricentro da distribuição de energia de um determinado som, sendo medido em *Hertz* (Hz). É a média ponderada das frequências presentes no sinal sonoro analisado, sendo que as magnitudes do espectro calculadas pela *FFT* funcionam como os pesos da média, onde $X_i(k)$ são as magnitudes extraídas pela *FFT* da janela *i*, e *k* é a metade dos componentes espectrais utilizados na Transformada. Em termos psicoacústicos, o Centroide Espectral relaciona-se à percepção do brilho sonoro em uma determinada região frequencial. Ou seja, valores altos do Centroide caracterizam a predominância ou o peso maior de frequências perceptivamente mais agudas, e valores mais baixos caracterizam a presença de frequências perceptivamente mais graves. Como o valor do Centroide é uma média, e a região frequencial em que ele se encontra caracteriza diferentes tipos de percepção (aspecto a ser elucidado mais adiante), utilizaremos na análise os seus valores numéricos absolutos. O $C_{(i)}$ é calculado pela seguinte expressão:

$$C_{(i)} = \frac{\sum_{k=1}^{k} k^{*}|X_i(k)|^2}{\sum_{k=1}^{k}|X_i(k)|^2}$$

Equação 1

Onde Xi(k) são as amplitudes das frequências da janela de análise.

[12] Brilho (Centroide Espectral) e Desvio Padrão Espectral.

[13] No *Sonic Visualiser*, o *plug-in* do fluxo espectral foi desenvolvido pela BBC, e os *plug-ins* do centroide espectral e *loudness* foram desenvolvidos por Chris Cannam e Jamie Bullock.

De acordo com Pereira (2009: 17) e Monteiro (2012: 27-29), o Fluxo Espectral $F_{(i)}$ determina a velocidade de mudança da energia do espectro, calculada a partir da diferença de magnitude entre duas janelas de análise consecutivas (X_i e X_{i-1}). Esse descritor fornece valores numéricos normalizados próximos de 0 quando as magnitudes permanecem relativamente invariáveis, e fornece valores normalizados próximos de 1 quando há grandes variações entre janelas sucessivas. O Fluxo é calculado pela Distância Euclidiana entre dois espectros normalizados, assim não é dependente da energia absoluta nem da relação de fase, já que apenas as magnitudes são comparadas. O $F_{(i)}$ é calculado pela expressão abaixo:

$$F_{(i)} = \sum_{k=1}^{k} \{log_{10}[X_i(k)] - log_{10}[X_{i-1}(k)]\}^2$$

Equação 2

Onde Xi(k) e X(i-1)(k) são as amplitudes das frequências de duas janelas de análise sucessivas.

O descritor *Loudness* $L_{(i)}$, segundo Pereira (2009: 19-20) e Monteiro (2012: 138-139), é uma medida psicoacústica relacionada à percepção da intensidade sonora, que varia em função de bandas de frequência e descreve a sensação auditiva de variação de amplitude do som analisado. A relação entre a percepção da intensidade e diferentes bandas de frequência é descrita pelas chamadas Curvas de Audibilidade ou de *Iso-Loudness,* de Fletcher e Munson (1933). Os valores do *Loudness* são dados por unidades de escala denominadas *phon* ou *sone* (as curvas de Fletcher e Munson fornecem medidas em *phon*). Em nossa análise, utilizaremos valores normalizados para este descritor. O *Loudness* $L_{(i)}$ de uma janela de análise espectral é dado pela Equação 3, sendo que as curvas de Fletcher e Munson são incluídas na Equação 4 pelo fator *W(k)*, que modula os valores $X_i(k)$.

$$L_i = \sum_{k=1}^{k-1} |X_i(k)|^2 10^{w(k)/20}$$

Equação 3

Onde Xi(k) são as amplitudes de frequência das janelas de análise.

$$W(k) = -0{,}6 * 3{,}64 * f(k)^{-0{,}8} - 6{,}5 * e^{-0{,}6*(f(k)-3{,}3)^2} + 10^{-3} * f(k)^{3{,}6}$$

Equação 4

Onde a frequência f(k) é medida em kHz, descrita como f(k) = k.d, e d é a diferença entre duas linhas espectrais consecutivas em kHz.

Os descritores acima mencionados são utilizados para extrair características sonoras da junção dos sons oriundos do piano com os processados pelo computador. Podemos agora esclarecer melhor como juntamos os pontos de análise com as ferramentas disponíveis. Como mencionamos, esses três descritores ajudam-nos a descrever certas características do timbre perceptível a partir de informações sobre o brilho (Centroide), nível de variação de eventos (Fluxo) e intensidade sonora perceptível (*Loudness*). Ao analisar essas medidas, o nosso objetivo é verificar quanto das características sonoras produzidas são decorrentes do comportamento do som instrumental, executado pelo pianista (e reflexo da escrita do compositor na partitura), e quanto se relaciona com o processamento eletroacústico implementado no *patch* (e reflexo do modelo computacional criado pelo compositor). Já o descritor BStD, definido a seguir, traz informações sobre da evolução no tempo do Volume Espectral em função do Centroide e do Desvio Padrão Espectral.

O BStD (MALT; JOURDAN, 2009a. MALT, 2012. BONDUKI; MONTEIRO, 2012), representação gráfica da evolução da textura sonora analisada, é composto pelas curvas do Centroide e do Desvio Padrão Espectral (raiz quadrada da variância espectral deste mesmo descritor). Malt e Jourdan (2009a: 3) afirmam que a distância visual entre essas duas curvas relaciona-se ao parâmetro espectral chamado de "Volume". Esse, por sua vez, pode ser associado ao nível de intensidade que é indicado pela variação da cor cinza (mais claro ou escuro) na implementação gráfica em *PureData* (BONDUKI; MONTEIRO, 2012: 55). O descritor BStD será empregado exclusivamente para comparar as duas gravações das interpretações de *Ressonâncias*, referente às hipóteses de análise detalhadas a seguir.

A primeira hipótese considera que o processamento do *phase vocoder source-filter* é bastante maleável em relação ao som resultante, já que a fonte sonora (o piano) varia devido a vários fatores interpretativos. Dentre eles, enfocamos três possibilidades: (1) aspectos interpretativos realizados pelo pianista em função da partitura; (2) influência de aspectos técnicos como a posição, nível de pré-amplificação e modelo dos microfones utilizados para captar o som do piano; e (3) a qualidade acústica da sala e do instrumento nos quais foi gravada a performance. Dessa forma, imaginamos que o resultado sonoro da peça seja derivado tanto de elementos invariáveis ou fixos, tais como os sons compostos em tempo diferido (*Mix* 1 a 6) como outros passíveis de variação, como a interpretação instrumental e o processamento do computador: o processamento do *phase vocoder source-filter*, o tamanho dos grãos da granulação dos sons *SussurroDrone* (1 a 4), os quais dependem da análise *FFT* feita pelo detector de altura fundamental (objeto *fiddle~*).

A segunda hipótese considera que a peça apresenta uma "tendência de maior similaridade espectral" na medida em que caminha ao fim, ou seja, o nível de variação entre diferentes interpretações tende a diminuir ao longo da peça. Isto significa que as seções iniciais de *Ressonâncias*, a saber, "Rubato" e "Recitativo" (considerando-se os *presets* de 1 a 11 do *patch*), que utilizam escolhas randômicas dos sons *Mix* e *SussurroDrone* quando acionados. Esses refletem diferentes escolhas dos processamentos eletroacústicos, pois todos os sons fixos são acionados, o que varia é a sua ordem. Por outro lado, nas seções "Decidido" e "Leve e Sutil" (*presets* 12 a 15), os sons *Mix* e *SussurroDrone* não variam, diminuindo a diferença dessas duas seções entre performances.

Análise de *Ressonâncias*

A seguir, o modelo com os quatro descritores (adicionando-se o BStD) é aplicado a duas gravações da obra, que serão denominadas a partir de agora "Gravação 1" e "Gravação 2". Associamos então as informações obtidas de todos os descritores em ambas as gravações com elementos da escrita da partitura da peça, considerando a própria segmentação formal estabelecida pelo compositor. Para extrair dados e verificar as hipóteses colocadas, aplicamos primeiramente o Centroide, o Fluxo e o *Loudness* aos sons compostos em tempo diferido e, assim, desejamos compreender os seus comportamentos espectrais no tempo. Dado que se trata de suporte fixo, esse comportamento é fonte de referência sonora para toda a obra.

Apresentamos na Tab. 1 a segmentação formal da peça definida *a priori* pelo compositor. Essa é intercalada por sequências de acordes arpejados que aparecem em três momentos (Acordes 1 a 3) e variam entre si na quantidade de notas e nos seus intervalos. Apresentamos também o número de compassos de cada seção, a quantidade de *presets* da parte eletrônica e a duração cronométrica das seções de cada uma das gravações, e a sua duração absoluta. Na Gravação 1, o tempo total da peça é de 8'22", e, na Gravação 2, de 7'55".

Seção	Compassos	*Presets* (eletrônica)	Gravação 1	Gravação 2
Acordes 1	1-5	Não há	0 - 43" (43")	0 - 37" (37")
Rubato	6-27	1 a 6	43" - 3'45" (3'02")	37" - 3'20" (2'43")
Acordes 2	28-29	Não há	3'45" - 4'15" (30")	3'20" - 4'03" (43")
Recitativo	30-47	7 a 11	4'15" - 6'57" (2'42")	4'03" - 6'28" (2'25")
Decidido	48-52	12 e 13	6'57" - 7'28" (31")	6'28" - 6'56" (28")
Acordes 3	53	Não há	7'28" - 7'40" (12")	6'56" - 7'05" (9")
Leve e Sutil	54-57	14 e 15	7'40" - 8'22" (42")	7'05" - 7'55" (40")

Tab. 1: Proposta de segmentação formal em *Ressonâncias*. As colunas Gravação 1 e 2 apresentam o tempo cronométrico inicial e final de cada seção gravada e a duração delas. Ou seja, a duração absoluta que é a diferença entre o tempo final e inicial.

Análise dos sons em suporte fixo compostos em tempo diferido. Os seis sons denominados *Mix*, que são difundidos sem transformações, tal como mostrado no diagrama de blocos dos processamentos do *patch* (Fig. 9), são mixados em conjunto com o som amplificado do piano e o som resultante do processamento do *phase vocoder source-filter*. Aplicamos os três descritores (Centroide, Fluxo Espectral e *Loudness*), a fim de extrair características relacionadas à composição do material fixo e a percepção do timbre de cada um deles. Trazemos, na Fig. 11, gráficos correspondentes aos valores absolutos do Centroide (média aritmética, desvio padrão, valor máximo e valor mínimo), além dos valores normalizados do Fluxo espectral e do *Loudness* (média aritmética e desvio padrão), com o intuito de comparar o comportamento dos descritores de cada som.

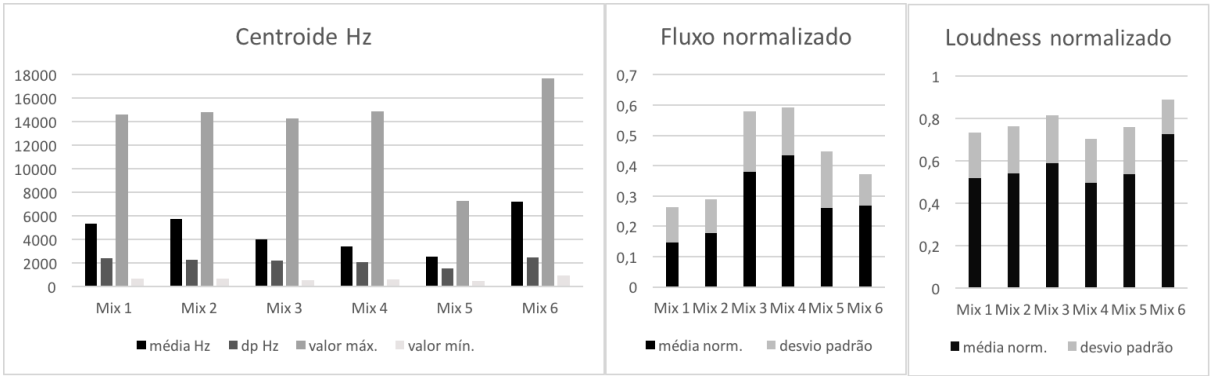


Fig. 11: Gráfico dos descritores Centroide, Fluxo e *Loudness* aplicados aos seis sons "*Mix*". O gráfico de barras do Centroide é organizado de modo diferente dos gráficos do Fluxo e *Loudness* (vide legenda na parte inferior da imagem).

Em relação ao Centroide, observa-se que os sons *Mix* 1 a 4 têm comportamentos similares, mas com algumas peculiaridades. Entre esses quatro sons, os *Mix* 1 e 2 são muito semelhantes em relação às quatro barras do gráfico. *Mix* 1 possui média de 5.355 Hz com desvio padrão de 2.406 Hz, valor máximo de 14.614 Hz e valor mínimo de 657 Hz. *Mix* 2 possui uma média de 5.772 Hz com desvio padrão de 2.279 Hz, valor máximo de 14.805 Hz e valor mínimo de 711 Hz. Os sons *Mix* 3 e 4, por sua vez, também são muito semelhantes entre si. *Mix* 3 possui média de 4.008 Hz com desvio padrão de 2.191 Hz, valor máximo de 14.825 Hz e valor mínimo de 585 Hz. *Mix* 4 possui média de 3.399 Hz com desvio padrão de 2.078 Hz, valor máximo de 14.891 Hz e valor mínimo de 643 Hz. *Mix* 5 é o som que mais se diferencia dos outros em relação ao Centroide: possui média de 2.536 Hz, desvio padrão de 1.544 Hz, valor máximo de 7.266 Hz e valor mínimo de 472 Hz. *Mix* 6 possui média de 7.211 Hz, desvio padrão de 2.482 Hz, valor máximo de 17.701 Hz e valor mínimo de 921 Hz.

Quanto à percepção do brilho, podemos afirmar que, na média, os sons *Mix* 1 a 4 têm características similares, sendo *Mix* 1 e 2 ligeiramente mais agudos que *Mix* 3 e 4. Os valores frequenciais máximo e mínimo encontrados são também bastante similares para esses sons. *Mix* 5 é o som dessa categoria que possui a sua média incidindo em 2.500 Hz (uma região perceptivamente médio-aguda), com valores máximo e mínimo consideravelmente inferiores aos sons acima abordados. *Mix* 6 é o som que possui os valores de média, máximo e mínimo mais elevados, além de maior amplitude entre os valores frequenciais mínimo e máximo, fazendo com que a percepção de seu brilho se situe na região mais aguda dos seis sons analisados.

É importante ressaltar também que o som *Mix* 6 é sempre disparado no mesmo ponto da peça pelo *preset* de mesmo número, no primeiro tempo do compasso 25 da partitura. Os compassos 25 a 27 podem ser entendidos como um trecho "solista" da parte eletroacústica. Nos compassos 25 e 26, temos a ressonância de um acorde sustentado executado pelo piano (formado pelas alturas[14] Lá 3, Mi 3, Lá 4, Fá 6) que foi atacado no compasso anterior. O compasso 27 tem apenas a presença dos sons eletroacústicos, o que reforça o caráter solista desse trecho. A Fig. 12 apresenta a partitura da peça, entre os compassos 24 e 27.

[14] Nesse artigo consideramos o Lá 4 como relativo à frequência 440 Hz.

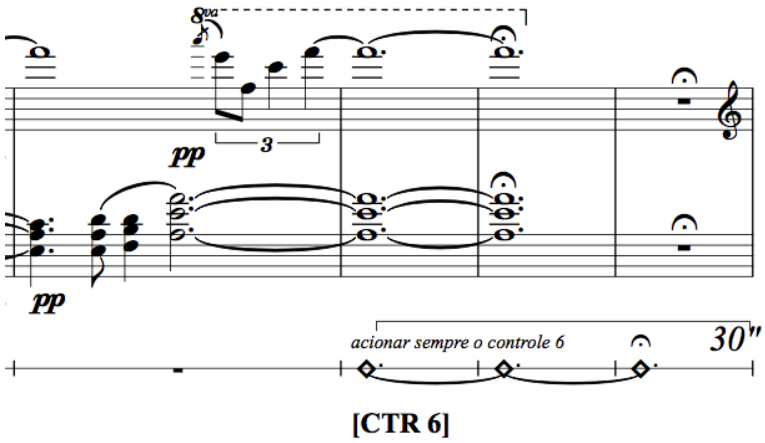


Fig. 12: Partitura de *Ressonâncias*, compassos 24 a 27.
Faixa de áudio: <http://git.nics.unicamp.br/danilo.rossetti/ressonancias/blob/master/áudio/2.%20Fig.%2012-Mix6.mp3> (som Mix 6)

Como utilizamos valores normalizados para os descritores de Fluxo Espectral e *Loudness*, seus gráficos incluem somente a média aritmética e o desvio padrão. Em relação ao Fluxo, encontramos semelhanças de comportamento entre os grupos de sons *Mix* 1 e 2, *Mix* 3 e 4, e *Mix* 5 e 6. *Mix* 1 tem fluxo médio de 0.15 e desvio padrão de 0.12, e *Mix* 2 tem fluxo médio de 0.18 e desvio padrão de 0.11. Os valores de *Mix* 3 e 4 são superiores, indicando que possuem maior variação ou diferença da magnitude entre janelas sucessivas. Em outras palavras, mais atividade ou mudança espectral. *Mix* 3 possui fluxo médio de 0.38 e desvio padrão de 0.2, enquanto *Mix* 4 possui fluxo médio de 0.43 e desvio padrão de 0.16. Os valores correspondentes aos sons *Mix* 5 e 6 situam-se em um nível intermediário entre os dois grupos anteriores. *Mix* 5 possui fluxo médio de 0.26 e desvio padrão de 0.19; *Mix* 6 possui média de 0.27 e desvio padrão de 0.1.

Dentro dos grupos propostos, sons com valores de Fluxo médio semelhante apresentam diferenças no valor do desvio padrão. Isso significa que, quando o valor dessa medida estatística for maior, a probabilidade de variação numérica das amostras é maior em relação à média, ao passo que, se esse valor for menor, a probabilidade de variação é menor. Consequentemente, um som que possui Fluxo médio e desvio padrão com valores maiores apresenta maior descontinuidade e maiores variações de magnitude entre janelas de análise sucessivas. No caso de *Mix* 5 e 6, apesar de terem um Fluxo médio muito próximo, o desvio padrão de *Mix* 5 é superior, acarretando numa maior percepção de variação de magnitude do espectro. Em linhas gerais, com Fluxo foi possível comparar como o compositor introduziu variedade na atividade espectral.

O comportamento do *Loudness* dos seis sons é bastante similar. *Mix* 1 a 5 possuem valores próximos de sua média normalizada (respectivamente 0.52, 0.54, 0.59, 0.5 e 0.54). O valor superior se dá em *Mix* 6 (0.73), o que nos faz ressaltar novamente o caráter solista dos sons eletroacústicos nesse trecho da música. Ou seja, *Mix* 6 tem uma curva de dinâmica destacada dos outros sons. Os valores do desvio padrão desse descritor também são semelhantes: 0.21, 0.22, 0.23, 0.2, 0.22 e 0.16. Percebe-se, novamente, que em *Mix* 6 há um valor que se destaca da média (0.16). Assim, podemos concluir que o nível da intensidade perceptível desse som varia menos, em relação à sua média, que os outros cinco sons analisados. Ou seja, mantém um nível de intensidade maior e com menos variações ao longo de seu desenvolvimento. Essas comparações

podem ser interpretadas como uma estratégia do compositor de criar um padrão ou nivelamento de amplitude e assim criar uma constância para que outros aspectos da interpretação da peça se destaquem. Como uma cortina sonora que abre espaço para o desenvolvimento da cena interpretativa.

A seguir, aplicamos o modelo dos descritores aos quatro sons *SussurroDrone*, a fim de conhecermos algumas de suas características espectro-temporais. Lembramos que esses sons não são simplesmente difundidos durante a performance, mas primeiramente granulados. Reiteramos que o tamanho do grão, atrelado à variação das alturas executadas pelo pianista, são extraídos em tempo real pelo objeto *fiddle~* (Fig. 9). Assim, quanto maior o intervalo entre duas alturas sucessivas, maior será o grão, este último escalonado no intervalo entre 100 e 300 ms.

Nesse ponto, faremos um comentário a respeito da percepção granular. Uma nuvem de sons granulada com grãos situados nos valores estabelecidos para a obra é percebida como granulação sutil. O fatiamento de grãos entre 100 e 300 ms preserva a presença dos transientes de ataque e uma boa parte dos regimes estacionários. Este procedimento possibilita a identificação da altura desses sons. Uma nuvem granular de caráter mais ruidoso pode ser concebida, em média, com grãos de tamanho inferior a 80 ms, com a predominância dos transientes de ataque dos sons fatiados pela granulação (ROSSETTI, 2016: 284-285). Então, o som resultante da granulação é enviado para o processamento do *phase vocoder source-filter*, o qual funciona como "filtro" do som captado ao vivo do piano (som fonte). O resultado desse processamento é mixado e difundido em conjunto com o som amplificado do piano e os sons *Mix*, já analisados. Na Fig. 13, temos os gráficos com a média aritmética dos valores dos descritores Centroide, Fluxo Espectral e *Loudness*, aplicados aos quatro sons *SussurroDrone*.

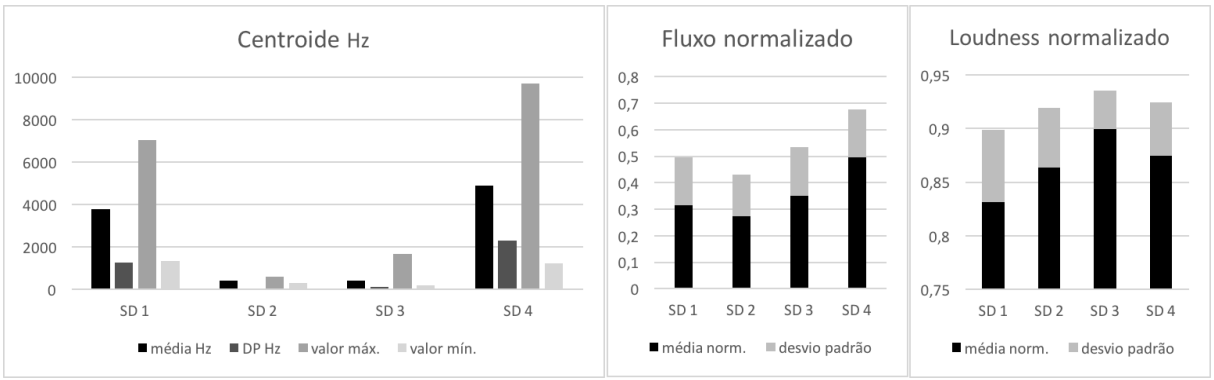


Fig. 13: Gráficos dos valores dos descritores Centroide, Fluxo e *Loudness* aplicados aos sons "*SussurroDrone*".

Em relação ao Centroide, observando seus gráficos, os quatro sons *SussurroDrone* têm valores muito distintos. No entanto, visualmente pode-se inferir certa semelhança entre os sons *SussurroDrone* 2 e 3. *SD* 1 possui média de 3.803 Hz com desvio padrão de 1.273 Hz, valor máximo de 7.052 Hz e valor mínimo de 1.345 Hz. *SD* 2 possui média de 421 Hz com desvio padrão de 42 Hz, valor máximo de 603 Hz e valor mínimo de 289 Hz. *SD* 3 possui média de 398 Hz com desvio padrão de 126 Hz, valor máximo de 1.664 Hz e valor mínimo de 205 Hz. *SD* 4 possui média de 4.913 Hz com desvio padrão de 2.289 Hz, valor máximo de 9.689 Hz e valor mínimo de 1.245 Hz.

A possível semelhança entre *SD* 2 e 3 não se confirma quando observamos seus dados detalhadamente. Apesar de uma média bastante próxima (421 e 398 Hz), seus valores frequenciais máximos são bastante distintos (603 e 1.665 Hz), além de seus desvios padrão também divergirem (42 e 126 Hz). Nesse sentido, os valores das amostras analisadas variam mais em relação à média em *SD* 3, o que nos permite verificar que a percepção do brilho se situa numa região ligeiramente mais aguda. *SD* 1 e *SD* 4 têm a percepção do brilho em regiões consideravelmente agudas (3.803 e 4.913 Hz), apesar de apresentarem uma amplitude bastante grande em relação aos seus valores frequenciais máximo e mínimo. Destacamos a agudeza de *SD* 4, que, além de uma média elevada, possui também um desvio padrão elevado, de 2.289 Hz, que demonstra que o valor médio pode variar ainda mais de 2.000 Hz para o agudo.

Em relação ao Fluxo Espectral, há valores bastante similares entre *SD* 1 e *SD* 3: média de 0.32 e desvio padrão de 0.18 para ambos os sons. *SD* 2 possui valores ligeiramente inferiores (respectivamente 0.27 e 0.16), enquanto *SD* 4 apresenta valor superior para a sua média (0.5) e valor similar para seu desvio padrão 0.18. Em relação a *SD* 1 e *SD* 3, observamos que, apesar de terem Fluxo similar, seus valores do centroide são bastante distintos. Aqui temos um exemplo da importância da utilização de mais de um descritor de áudio para termos uma visão mais abrangente do comportamento sonoro. Se houvéssemos utilizado apenas o primeiro descritor, poderíamos concluir erroneamente que estes dois sons são bastante similares ou bastante diferentes, ao passo que, na realidade, eles se assemelham em algumas características e se distinguem em outras. Em *SD* 4, observa-se um Fluxo médio mais elevado devido à presença de uma modulação de amplitude que permanece constante durante toda a evolução do sinal, com isso a diferença de magnitude entre janelas sucessivas aumenta. O desvio padrão não muito elevado e similar aos outros sons dessa categoria atesta que essa modulação não varia no tempo, permanecendo constante.

O *Loudness* dos quatro sons *SD* não apresenta grandes variações e se situam entre 0.8 e 0.9, demonstrando que todos possuem um nível de sonoridade elevado. Um desvio padrão extremamente baixo como o encontrado para esses sinais (entre 0.04 e 0.06) confirma que há muito pouca variação em relação à média, comprovando seu comportamento constante. De maneira similar ao *Loudness* dos sons *Mix*, supomos que a estratégia do compositor em manter o *Loudness* desses sons destacada e constante deve-se ao fato de que esses sons são granulados. Ou seja, ao fatiarmos os sons em ínfimas partes, transformando-as em uma nuvem sonora descontínua, a tendência é que o nível de intensidade média perceptível diminua quanto mais rarefeita for a nuvem sonora gerada.

A seguir, apresentaremos a análise comparativa das duas gravações da obra. Nosso objetivo é detectar suas singularidades que envolvem a interpretação pianística e suas dependências com as transformações eletroacústicas aplicadas em tempo real.

Análise das gravações da obra. Como já mencionamos, as durações totais das Gravações 1 e 2 são diferentes (8'22"e 7'55"). Ademais, as durações internas das sete seções da peça variaram também, como pode ser observado na Tab. 1. Na Fig. 14, apresentamos os gráficos delimitados pela segmentação formal proposta, os quais se referem às duas gravações. Neles apresentamos os Sonogramas e as curvas do Centroide, Fluxo e *Loudness*, produzidos pelo *software Sonic Visualiser*. Soma-se a essas representações a curva do BStD, obtida com a implementação em *PureData* de Monteiro (BONDUKI; MONTEIRO, 2012). Assim, observamos se a evolução dos descritores é similar nas duas gravações, se há diferenças entre as seções da

obra e se o disparo randômico dos sons *Mix* e *SussurroDrone,* nas seções Rubato e Recitativo, influencia a sonoridade resultante. Verificamos também se há diferenças entre a mesma seção em gravações diferentes levando-se em conta a evolução temporal do amálgama tímbrico formado. Convém ressaltar que, apesar das figuras referentes às Gravações 1 e 2 terem tamanho igual, suas durações absolutas são diferentes. Consideramos que, em termos comparativos, a diferença total de 27 segundos não acarreta grande distorção na escala visual do gráfico.

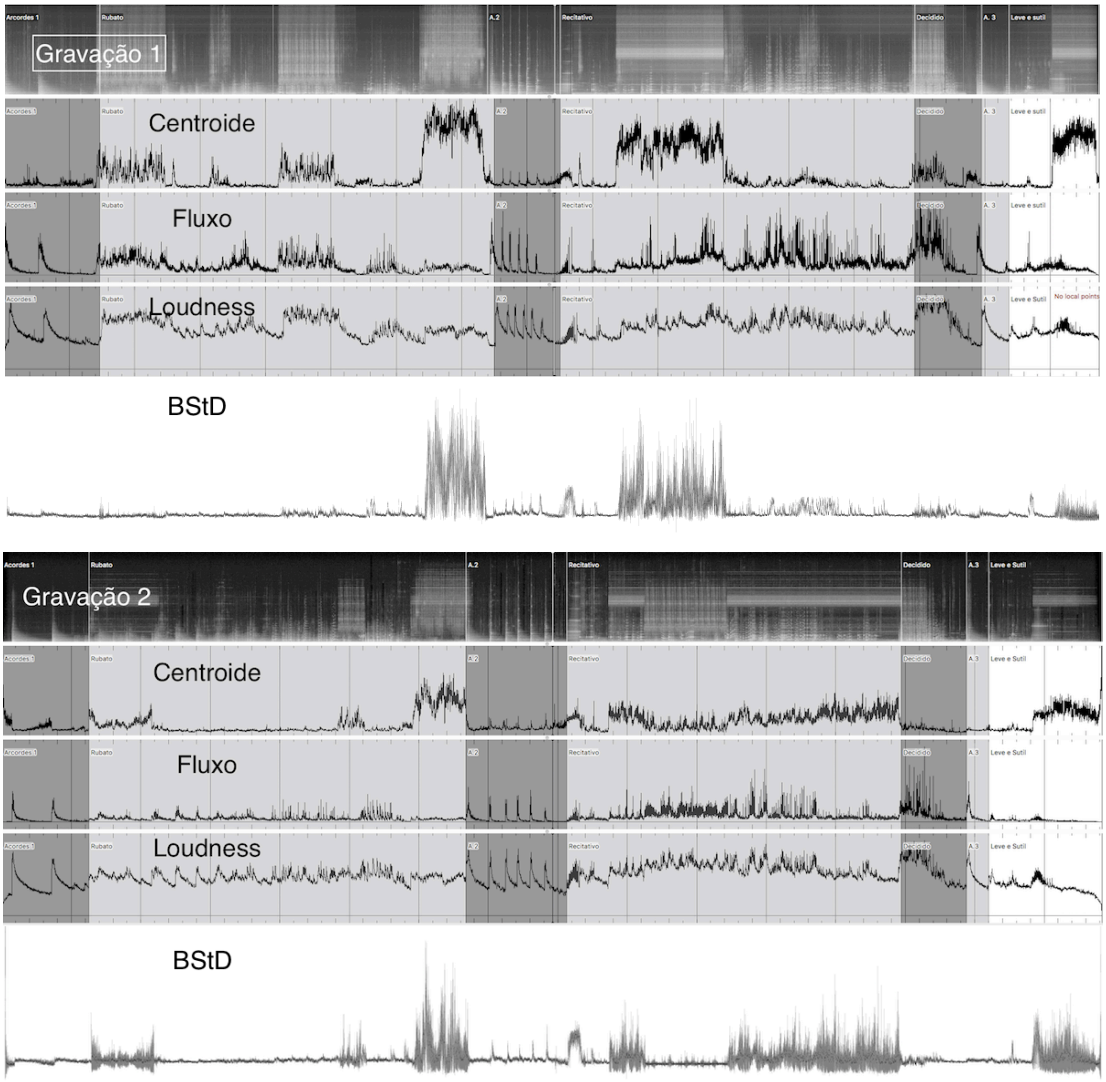


Fig. 14: Sonograma, curvas de Centroide, Fluxo e *Loudness* e representação BStD nas duas gravações analisadas.

Faremos referência nessa análise às seções que apresentam interação dos sons eletroacústicos com os sons instrumentais (Rubato, Recitativo, Decidido, e Leve e Sutil). As seções referentes aos acordes do piano (Acordes 1, 2 e 3), por serem puramente instrumentais, não serão mencionadas. Numa visão geral, a Gravação 1 apresenta maiores variações espectrais dentro das seções formais, enquanto a Gravação 2 apresenta maior constância em relação aos valores dos descritores utilizados. Além disso, o brilho do espectro total (instrumento e sons eletroacústicos) dado pelo Centroide é principalmente influenciado pelo brilho dos sons

eletroacústicos. O Fluxo Espectral, por outro lado, é influenciado, primordialmente, pela linha executada no piano, sendo que seus maiores valores correspondem aos ataques com muita energia realizados pelo intérprete desse instrumento. A curva do *Loudness* nos fornece a evolução do nível intensidade percebida ao longo da peça. Por essa razão, ela sofre influência nos diferentes momentos da peça tanto dos sons instrumentais acústicos como dos sons eletroacústicos.

As seções Rubato e Recitativo (números 2 e 4) são as mais divergentes em relação ao amálgama tímbrico produzido. Nelas incide a escolha randômica do disparo dos sons eletroacústicos *Mix* e *SussuroDrone*. Portanto, nesses trechos, a informação musical que se mantém com maior constância é a produzida pela partitura do piano, que, por outro lado, está sujeita somente às variações interpretativas dos dois pianistas. Na seção Rubato da Gravação 1, há maior contraste interno do timbre (variações do Centroide e do Fluxo), o que nos leva a considerar maior diversidade nos sons disparados na parte eletroacústica (*presets* 1 a 5) e maior contraste interpretativo do piano, principalmente em relação ao manejo das dinâmicas. Na mesma seção pertencente à Gravação 2, há uma maior constância do amálgama tímbrico produzido, que se concentra principalmente na região grave (observável pelas curvas do Centroide e Fluxo), exceto no seu final. Consideramos, no caso dessa gravação, que os sons disparados pelos *presets* da parte eletroacústica possuíam morfologias espectrais semelhantes.

Na seção Recitativo, a divergência entre as duas gravações se dá novamente em relação ao Centroide. Há, nas duas gravações, contrastes internos dentro dessa seção, muito provavelmente devido à escolha randômica dos sons disparados. A curva do Fluxo Espectral, no geral, tem valores mais altos e maior amplitude nos gráficos das duas gravações, já que a escrita pianística apresenta maior densidade de notas e contrastes de dinâmica. A curva do *Loudness* também tem valores mais altos nas duas gravações, mostrando que o nível de intensidade percebido dessa seção é maior.

As outras duas seções finais, Decidido e Leve e Sutil, não apresentam muita divergência em relação ao disparo dos sons eletroacústicos, fazendo com que as curvas dos descritores apresentem maior convergência nas duas gravações analisadas. Mesmo assim, ainda é possível encontrarmos diferenças que podem ser atribuídas às interpretações dos pianistas, à equalização entre o som do piano e da parte eletroacústica, e à gravação e mixagem realizada no momento da performance. Na seção Decidido, que apresenta uma escrita pianística com grande densidade de notas que variam entre *f* e *ff*, temos, para a Gravação 1, o Centroide com valores médio-baixos, o Fluxo com valores altos e o *Loudness* também com valores altos. Na Gravação 2, temos a curva de Centroide com valores baixos, e Fluxo e *Loudness* com valores altos. Na seção Leve e Sutil temos, na Gravação 1, Centroide com valores altos, Fluxo com valores baixos e o *Loudness* com valores na região média. Na Gravação 2, o Centroide possui valores médios, o Fluxo com valores baixos e o *Loudness* valores médios.

Na Fig. 15, trazemos a partitura desse trecho, a fim de observar as diferenças da escrita pianística nessas seções; assim, podemos observar, entre outras informações, o caráter *fff* da seção C (Decidido), com ênfase no fraseado da mão esquerda, em contraste com a seção D (Leve e Sutil), que se inicia com um acorde em *sfff* e depois apresentando figurações em *p* e *pp*. Na escrita pianística há um ganho de energia e densidade com os eventos da seção C, e uma espécie de *fade-out* gradual à medida em que a peça caminha para o final, com eventos mais rarefeitos e com menor dinâmica que convergem para a região aguda do instrumento.

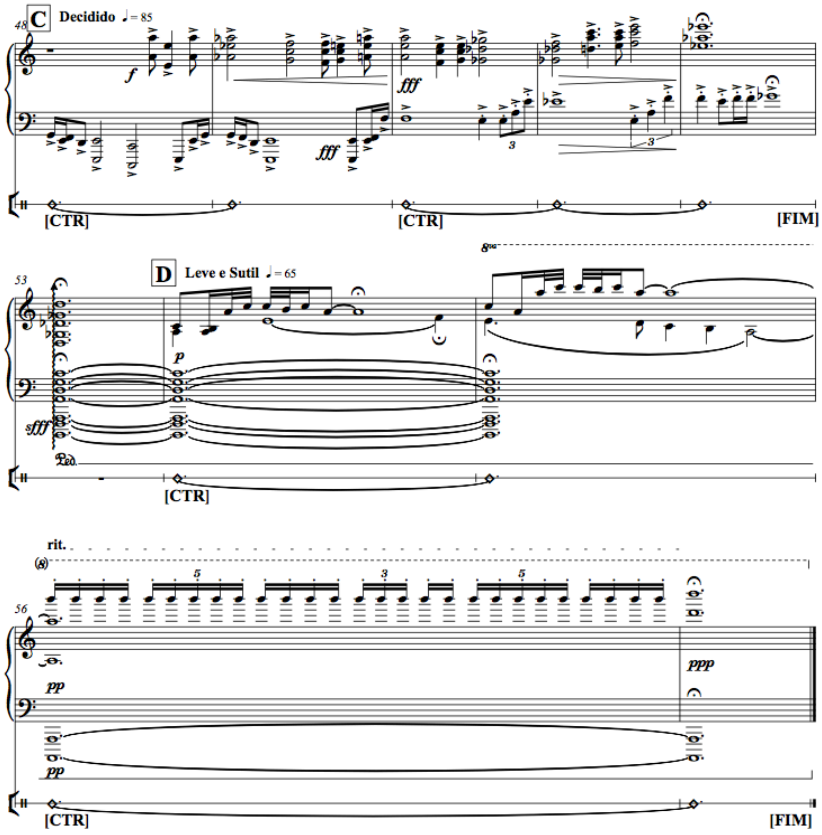


Fig. 15: Partitura de *Ressonâncias*, seções Decidido e Leve e Sutil.
Faixa de áudio: <http://git.nics.unicamp.br/danilo.rossetti/ressonancias/blob/master/áudio/3.%20Fig.15-c.%2048-57-Gravação1.mp3>

O gráfico do descritor BStD referente às duas gravações nos mostra o Volume do timbre analisado, como comentado anteriormente. De maneira geral, num determinado ponto temporal das curvas, quanto maior é o espaço entre elas, maior é a variação em Hz dos dados utilizados para o cálculo da média do Centroide. Como observado, nesses momentos em que há maior espaço, a percepção do timbre é mais densa e, por outro lado, nos momentos em que a BStD se assemelha a uma linha, há uma percepção mais sutil do timbre. Na Fig. 16, utilizamos o BStD para comparar as Gravações 1 e 2.

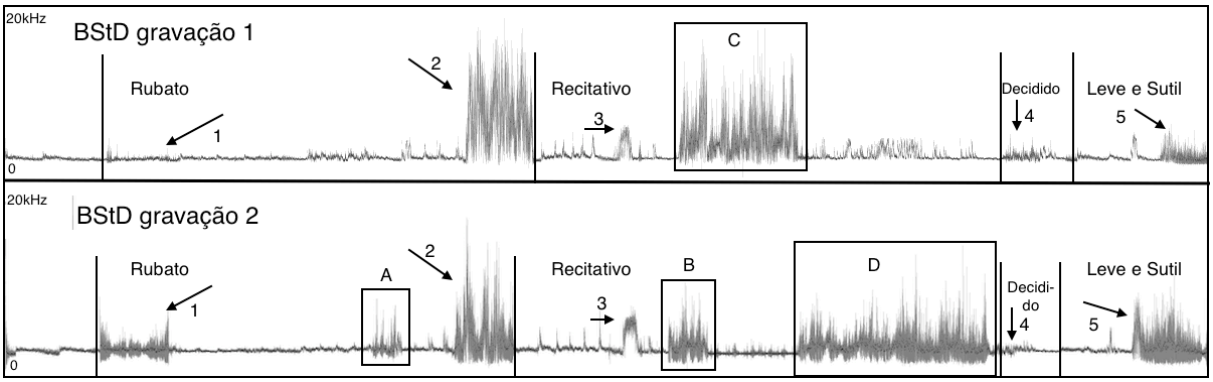


Fig. 16: Curvas do descritor BStD das Gravações 1 e 2 de *Ressonâncias*.

Na Fig. 16, as semelhanças espectrais das curvas de BStD relativas aos momentos estruturais da peça são apontadas pelas setas, numeradas de 1 a 5. É interessante notar que, de maneira geral, a Gravação 2 apresenta um Volume maior, acarretando na percepção de maior densidade. Vide a maior espessura dos pontos apontados no gráfico pelas setas numeradas e as letras de A a D. Corroborando essa afirmação, temos também uma tonalidade de cinza mais escura para os trechos da Gravação 2, enquanto os trechos destacados da Gravação 1 mostram-se mais claros. A variação dos tons de cinza está associada à variação quadrática média da energia do som. Ou seja, quanto mais cinza, mais energia foi verificada.

No geral, os trechos de maior volume da BStD se referem a trechos com mais presença perceptiva dos sons eletroacústicos de caráter agudo. Por outro lado, trechos com BStD mais lineares se referem a sons eletroacústicos mais graves. Ressalta-se novamente que o Volume é interpretado como a distância visual no gráfico que está associada à distância entre o espectro frequencial dos sons eletroacústicos e instrumentais. As setas 1 e 2 da seção Rubato apontam para uma semelhança espectral referente ao início dos sons eletroacústicos (seta 1) e ao disparo do *preset* 6, trecho solista da parte eletroacústica que encerra essa seção. No retângulo A da Gravação 2, temos uma divergência espectral notável que demonstra que a escolha randômica do disparo dos sons em suporte fixo acarretou grandes diferenças no resultado sonoro obtido.

Na seção Recitativo, as diferenças espectrais se intensificam, destacadas na figura pelos retângulos B, C e D, confirmando que este é outro momento de divergência estrutural da peça, distância esta ainda maior que na seção anterior analisada. Destacamos também que o Volume da curva de BStD referente aos retângulos B e D da Gravação 2 é significativamente mais intenso (coloração mais escura de cinza) que o do retângulo C da Gravação 1, ainda que este tenha um espaço frequencial mais extenso que o retângulo B da Gravação 2. Mencionamos ainda nessa seção, indicado pela seta número 3, um momento em que o piano, ao executar ataques rápidos da nota Mi 7, localizada numa região extremamente aguda, produz um Volume destacado por esse descritor. Na Fig. 17, reproduzimos os compassos 30 e 31, referentes a esse trecho da partitura.

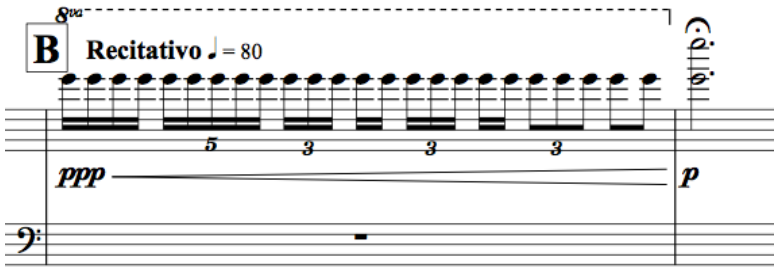


Fig. 17: Compassos 30 e 31 da partitura: repetição da altura Mi 7.

Por fim, as setas 4 e 5, localizadas nas seções Decidido e Leve e Sutil, apontam para semelhanças estruturais nesses dois trechos e confirmam sua convergência. Assim, podemos extrapolar que diferentes interpretações dessas seções têm menor possibilidade de variar que as seções anteriores. Isto devido, principalmente, à invariância dos disparos dos sons eletroacústicos programados. As diferenças interpretativas dos pianistas podem ser colocadas como a principal fonte de variação nesses trechos. Ainda assim, notamos que o trecho apontado pela seta 5 na Gravação 2 tem um volume espectral mais denso que o mesmo trecho da Gravação 1. Daí constatamos a presença de um maior Volume dos sons eletroacústicos nesse trecho da gravação.

Os momentos de maior volume e densidade espectral da peça apontados pelo descritor BStD, sejam eles convergentes ou divergentes em relação às duas gravações, nos descrevem a existência de pregnâncias sonoras, no sentido empregado por René Thom (1988: 20-22). As pregnâncias são tipos de formas estacionárias e contínuas, de média e longa duração, uma espécie de fluido que se espalha em nosso campo perceptivo a partir de algum tipo de descontinuidade que se instala num momento anterior. O rompimento dessas pregnâncias, por sua vez, vai ocorrer com o surgimento de novas descontinuidades ou rupturas.

Considerações finais

O modelo de análise musical para a música eletroacústica mista proposto nesse artigo não pretende ser um modelo para substituir as outras formas de análise que já estão estabelecidas nesse contexto de estudo musicológico. Nossa intenção é propor uma nova possibilidade complementária às outras abordagens. O ganho advindo da utilização dos descritores de áudio é a possibilidade de captar e representar variações sonoras que muitas vezes são pouco perceptíveis e difíceis de serem representadas por análises ou notações musicais tradicionais, muitas delas calcadas somente na informação da partitura. Como já mencionamos na introdução, a interação entre os instrumentos e os sons eletroacústicos produzem resultados sonoros emergentes que fogem à capacidade representativa da notação musical com os meios e ferramentas que possuímos atualmente.

Através do método de análise proposto, pudemos verificar a veracidade das hipóteses colocadas, as quais estavam majoritariamente ligadas à variedade produzida nas interpretações da obra. Teoricamente, sabe-se que o resultado do processamento *phase vocoder source-filter* é dependente da interação dos dois sons que são injetados no sistema de processamento. Porém, o fato de que o som a ser filtrado é um som captado ao vivo (no caso de *Ressonâncias*, o som do piano) produz um nível de variação maior como resultado desse processo, principalmente por causa das diferentes interpretações. A existência de uma tendência de maior similaridade espectral ou invariância[15] na peça também pôde ser verificada, na medida em que foi mostrado que as seções Rubato e Recitativo são consideravelmente mais díspares que as seções Decidido e Leve e Sutil nas duas gravações analisadas.

Desta forma, o imbricamento estrutural dos processos é a principal fonte de variedade sonora, e é a partir dele que emergem as sonoridades derivadas do aqui-agora da interpretação da obra. Assim, o intérprete eletroacústico e o pianista dialogam não somente no nível das interações prescritas na partitura, mas, notadamente, num jogo invisível, implícito, que foi fomentado pelo compositor a partir do pensamento musical transcrito no modelo implementado no computador. Esse opera como projeção da proposta poética do compositor e da apropriação do objeto técnico, como já discutimos anteriormente. Tal como a epígrafe do poema inicial, esse diálogo é o que aproximaria "nos corpos e nas mentes os sussurros das ressonâncias dos ventos e das gentes".

Assim, ressaltamos também que o nível de interação entre sons instrumentais e eletroacústicos foi projetado pelo compositor no modo como criou o *patch*. Temos, como já foi dito, a interação entre sons captados em tempo real e sons compostos em tempo diferido através do *phase vocoder source-filter*, mas também a extração de informações captadas e geradas através da análise da Transformada de Fourier em tempo real. Esse material, por sua vez, influencia o tamanho

[15] A invariância corresponde a uma ou mais propriedades de um sistema emergente que permanece invariável dentro de alguma transformação.

do grão e a densidade das nuvens de sons produzidas nos processos de granulação dos sons compostos em tempo diferido e armazenados em suporte fixo. Todas essas interações formam uma rede de conexões complexa entre os objetos do *patch*, em que o nível de automação presente é amplamente reduzido. O microfone que capta o som do piano é a interface do ambiente digital com o meio, provocando outro nível de abertura criado no sistema, o que poderíamos descrever de forma conceitual, como a "utilização do som como a própria interface do som". Esse nível, por sua vez, é amplificado devido às inter-relações criadas entre os objetos e valores empregados nos processamentos (ASHBY, 1970 [1956]: 49-84).

No caso de *Ressonâncias*, há através do *patch* uma modulação, no sentido de Simondon, de moldar a forma (2005 [1958]: 47-48) dos materiais sonoros processados, ou seja, os sons do piano e os sons eletroacústicos compostos em tempo diferido, produzindo novas morfologias sonoras resultantes da interação entre esses materiais. Essa visão também corrobora a intenção poética inicial da composição, na medida em que as amplitudes do espectro de Fourier reproduzem a amplitude da ressonância de cada parcial que foram considerados pelo compositor como representações imagéticas das morfologias de algum fenômeno.

Diferentes ressonâncias podem se modular entre si, produzindo novas morfologias com as características herdadas das morfologias iniciais.

Referências

AGOSTINI, Giulio; LONGARI, Maurizio; POLLASTRI, Emanuele. Musical Instrument Timbre Classification with Spectral Features. In: *EURASIP Journal of Applied Signal Processing*. Springer International Publishing, 2003. p. 5-14.

ASHBY, William Ross. *Introdução à cibernética* [1956]. São Paulo: Perspectiva, 1970.

BONDUKI, Said; MONTEIRO, Adriano. Compositional Influences in *Jonchaies* from *La Légende d'Éer*. In: SOLOMOS, Makis (Ed.). *Iannis Xenakis, la musique électroacoustique*. Paris: L'Harmattan, 2012. p. 53-67.

COOLEY, James; TUKEY, John. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation,* v. 19, n. 90, p. 297-301, 1965.

DOLSON, Mark. The Phase Vocoder: A Tutorial. *Computer Music Journal,* v. 10, n. 4, p. 14-27, 1986.

FLETCHER, Harvey; MUNSON, Wilden. Loudness, its Definition, Measurement and Calculation. *Journal of the Acoustic Society of America,* v. 5, p. 82-108, 1933.

GERZSO, Andrew; BOULEZ, Pierre. Computers in Music. *Scientific American,* v. 258, n. 4, p. 44-51, 1988

MALT, Mikhail. Une proposition pour l'analyse des musiques électroacoustiques de Xenakis à partir de l'utilisation de descripteurs audio. In: SOLOMOS, Makis (Ed.). *Iannis Xenakis, la musique électroacoustique*. Paris: L'Harmattan, 2012. p. 159-198.

MALT, Mikhail; JOURDAN, Emmanuel. Zsa.Descriptors: A Library for Real-Time Descriptors Analysis. In: SOUND AND MUSIC COMPUTING CONFERENCE, 8., 2008, Berlim. *Proceedings...* Berlim, 2008. Disponível em: <http://smcnetwork.org/node/1246>. Acesso em: 2 ago. 2017.

_____. Le "BStD" – Une représentation graphique de la brillance et de l'écart type spectral, comme possible representation de l'évolution du timbre sonore. In: L'ANALYSE MUSICALE AUJOURD'HUI, CRISE OU (R)ÉVOLUTION, 2009, Strasbourg. *Proceedings..,* Strasbourg, 2009b.

Disponível em: <https://www.researchgate.net/publication/233389199_Le_BSTD_-_Une_representation_graphique_de_la_brillance_et_de_l%27ecart_type_spectral_comme_possible_representation_de_l%27evolution_du_timbre_sonore>. Acesso em: 19 jul. 2017.

_____. Real-Time Uses of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa. Descriptors Library. In: BRAZILIAN SYMPOSIUM ON COMPUTER MUSIC, 12., 2009, Recife. *Proceedings...* São Paulo, 2009b.

MANZOLLI, Jônatas. Interpretação mediada: pontos de referência, modelos e processos criativos. *Revista Musica Hodie,* v. 13, n. 1, p. 48-63, 2013.

MONTEIRO, Adriano. *Criação e performance musical no contexto dos instrumentos musicais digitais*. Dissertação (Mestrado em Música). Instituto de Artes, Universidade Estadual de Campinas, Campinas, 2012.

MOORER, James. The Use of the Phase Vocoder in Computer Music Applications. *Journal of Audio Engineering Society,* v. 26, n. 1, p. 42-45, 1978.

OPPENHEIM, Alan; SCHAFER, Ronald; BUCK, John. *Discrete-Time Signal Processing*. 2ª ed. Londres: Pearson, 1999.

OSGOOD, Brad. *Lecture Notes for The Fourier Transform and its Applications*. Stanford: Stanford University, Electrical Engineering Department, 2007.

PADOVANI, José Henrique; FERRAZ, Silvio. Proto-história, evolução e situação atual das técnicas estendidas na criação musical e na performance. *Revista Música Hodie,* v. 11, n. 2, p. 11-25, 2011.

PEETERS, Geoffroy. *A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project*. Paris: Institut de Recherche et de Coordination Acoustique-Musique (IRCAM), 2004. Disponível em: <http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf>. Acesso em: 22 jun. 2017.

PEETERS, Geoffroy et al. The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals. *Journal of Acoustic Society of America,* Melville, v. 130, n. 5, p. 2902-2916, Nov. 2011.

PEREIRA, Erica. *Estudos sobre uma ferramenta de classificação musical*. Dissertação (Mestrado em Engenharia Elétrica e Computação). Faculdade de Energia Elétrica e Computação, Universidade Estadual de Campinas, Campinas, 2009.

PUCKETTE, Miller. *The Theory and Technique of Electronic Music*. World Scientific Publishing, 2006. Disponível em: <http://msp.ucsd.edu/techniques.htm>. Acesso em: 7 out. 2016.

ROADS, Curtis. *The Computer Music Tutorial*. Cambridge: The MIT Press, 1996.

ROEBEL, Alex; LIUNI, Marco. Phase Vocoder and Beyond. *Musica/Tecnologia,* v. 7, p. 73-89, 2013.

ROSSETTI, Danilo. *Processos microtemporais de criação sonora, percepção e modulação da forma*: uma abordagem analítica e composicional. Tese (Doutorado em Música). Instituto de Artes, Universidade Estadual de Campinas, Campinas, 2016.

_____. Temporalidades, morfologias e granulações em *Schall* de Horacio Vaggione. In: ENCONTRO INTERNACIONAL DE TEORIA E ANÁLISE MUSICAL, 4., 2017, São Paulo. *Anais...* São Paulo, 2017a. p. 325-336.

_____. The Qualities of the Perceived Sound Forms: A Morphological Approach to Timbre Composition. In: ARAMAKI, M. et al. (Eds.). *Bridging People and Sound: 12th International Symposium, CMMR 2016, São Paulo, Brazil, July 5-8 2016, Revised Selected Papers*. Cham: Springer, 2017b. p. 259-283.

ROSSETTI, Danilo; TEIXEIRA, William; MANZOLLI, Jônatas. An Analysis of *Desdobramentos do Contínuo Performad by Audio Descriptors*. In: BRAZILIAN SYMPOSIUM ON COMPUTER MUSIC, 16., 2017, São Paulo. *Proceedings...*, São Paulo, 2017. p. 123-130.

ROSSETTI, Danilo; FERRAZ, Silvio. Forma musical como um processo: do isomorfismo ao heteromorfismo. *Opus,* v. 22, n. 1, p. 59-96, 2016.

SIMONDON, Gilbert. *Du mode d'éxistence des objets techniques* [1958]. Paris: Aubier, 2012.

_____. *L'individuation à la lumière des notions de forme et d'information* [1958]. Grenoble: Jérome Millon, 2005.

SIMURRA, Ivan; MANZOLLI, Jônatas. O azeite, a lua e o rio: o segundo diário de bordo de uma composição a partir de descritores de áudio. *Revista Música Hodie,* v. 16, n. 1, p. 101-123, 2016a.

_____ . Sound Shizuku Composition: a Computer-Aided Composition System for Extended Music Techniques. *MusMat Brazilian Journal of Music and Mathematics,* v. 1, n. 1, p. 86-101, 2016b.

TEMPELAARS, Stan. *Signal Processing, Speech and Music*. Lisse: Sweets & Zeitlinger, 1996.

THOM, René. *Esquisse d'une sémiophisique*. Paris: InterEditions, 1988.

Agradecimentos

Danilo Rossetti tem sua pesquisa de pós-doutorado financiada pela FAPESP (processo 2016/23433-8), e Jônatas Manzolli é financiado por uma bolsa Pq do CNPq (processo 305065/2014-9).

Danilo Rossetti é Doutor em Composição Musical pela Unicamp, com período sanduíche no Centre de recherche Informatique et Création Musicale da Université Paris 8. Estudou composição com José Manuel López López, Silvio Ferraz e Flo Menezes, e técnicas de música eletroacústica mista com Alain Bonardi e Anne Sèdes. Atualmente é pós-doutorando no Núcleo Interdisciplinar de Comunicação Sonora da UNICAMP sob supervisão de Jônatas Manzolli, com apoio FAPESP, e professor da pós-graduação do Centro Universitário Senac SP, nas áreas de comunicação e artes. Foi um dos premiados do Prêmio Funarte de Música Clássica 2016, na categoria solos, música acusmática e mista. Atua principalmente nas seguintes áreas: composição, teoria e análise musical. É autor de *Modelos de composição e percepção em Xenakis: Concret PH e o Pavilhão Philips* (*Opus*, v. 18 n. 1) e Forma musical como um processo: do isomorfismo ao heteromorfismo (*Opus,* v. 22, n. 1). danilo.rossetti@nics.unicamp.br

Jônatas Manzolli é compositor e matemático, pesquisa a interação entre Música e Ciência em criação musical e ciências cognitivas. Professor Titular do Instituto de Artes da UNICAMP e Coordenador Associado do Núcleo Interdisciplinar de Comunicação Sonora (NICS), pesquisador 1B do CNPq. Suas publicações focam, principalmente, os seguintes temas: ambientes interativos para composição, modelos matemáticos e computacionais no contexto da criação e análise da música atual. Sua produção artística relaciona música mista e instrumental, obras multimodais para dança e instalações sonoras. jonatas@nics.unicamp.br