



Escuta pela máquina e máquinas de escuta: a machine listening e o solfejo

MODALIDADE: COMUNICAÇÃO

José Henrique Padovani

UFPB – padovani.ufpb@gmail.com

Apoio: CNPq – Edital Universal 2014

Resumo: O texto procura contextualizar os recursos de *machine listening* disponíveis às práticas musicais/sonoras da atualidade. Após identificar na escuta automatizada e na escuta mediada por processos computacionais uma situação análoga àquela da composição com algoritmos a partir da década de 1950, são colocadas questões relacionadas à interdisciplinaridade e à incorporação dessas técnicas em atividades criativas, educacionais e investigativas. Ao final, propõe-se uma reavaliação da concepção de *solfejo* – já expandida, anteriormente, por Pierre Schaeffer.

Palavras-chave: *machine listening*; solfejo; escuta; máquina.

Title of the Paper in English: Listening through the machine and listening machines: the machine listening and the solfège

Abstract: The paper intends to contextualize the machine listening resources currently available to contemporary sound/musical practices. After identifying in the machine mediated and in the machine automated listening an analogous situation to that of composition with algorithms from the 1950s, questions are placed regarding the interdisciplinary and the incorporation of these techniques in creative, educational and investigative activities. In the end, it proposes a reevaluation of the concept of *solfège* – already expanded previously by Pierre Schaeffer.

Keywords: *machine-listening*; *solfège*; listening; machines.

1. Introdução

A noção de *solfejo* foi profundamente reelaborada na música do pós-guerra, seja a partir de sua reconsideração teórica e prática por Pierre Schaeffer (1966: 490-508), seja a partir do desenvolvimento de técnicas e objetos técnicos voltados à análise e processamento de sinais que, posteriormente, repercutiram em práticas e processos de escuta e análise mediados por recursos tecnológicos. A despeito da prevalência de uma acepção ainda tradicional do termo no campo da música (o que, evidentemente, é acompanhado de práticas e concepções pedagógicas não menos arcaicas), o *solfejo*, enquanto processo essencial para a escuta, a prática criativa e a investigação, defronta-se, hoje, com questões análogas àquelas que a composição mediada por recursos algorítmicos viveu em seus primórdios.

Se a computação musical maturou-se paralelamente ao desenvolvimento das práticas eletroacústicas a partir de processos de composição com algoritmos mais ou menos automatizados, hoje os processos de solfejo mediados por recursos computacionais também são transformados a partir de mecanismos em que a própria escuta pode ser assistida ou modelada computacionalmente a partir de diversas técnicas de *machine listening*. Se esse advento tecnológico pode parecer irrelevante a pesquisadores e artistas não diretamente envolvidos no seu estudo, seus efeitos são cada vez mais onipresentes no panorama atual,

englobando não apenas sua utilização criativa nas artes sonoras em geral mas, também, em aplicações as mais diversas: do emprego dessas técnicas na busca/identificação de músicas¹ até sua aplicação para identificar automaticamente anomalias em equipamentos ou eventos sonoros em ambientes urbanos (como a detecção de ruídos de tiros e gritos, por exemplo)².

No presente texto buscarei esboçar uma discussão inicial a respeito da questão da escuta mediada e automatizada por recursos computacionais, retomando, para isso, certas concepções oriundas das primeiras teorizações da música eletroacústica. Parte de uma pesquisa que prevê a pesquisa e a exploração criativa dos recursos técnicos envolvidos, o texto procura situar conceitualmente a questão da escuta por máquinas e através de máquinas, o que repercute, ainda, na própria definição de *solfejo* – já ampliada por Pierre Schaeffer.

2. Escuta e solfejo no contexto da abordagem interdisciplinar

Na elaboração de seu projeto de um programa para a pesquisa musical, Pierre Schaeffer delinea os desafios de estudos interdisciplinares em geral e, mais especificamente, de abordagens desse tipo no campo da música. Buscando superar aquilo que considera como uma “cautela de especialistas” receosos em sair de sua zona de atuação, Schaeffer reconhece a música como campo privilegiado para a interdisciplinaridade, relevando, porém, que tal aproximação apresenta desafios. “A música”, diz ele, “é uma montanha na qual cada um cava seu túnel, e as galerias se entrecruzam sem se encontrar” (SCHAEFFER, 1966: p. 29).

A abordagem interdisciplinar, assim, não deve ser concebida apenas a partir de uma abordagem positiva ou otimista de unidade e cooperação de disciplinas, mas, antes – como resume, a partir de Schaeffer, Michel Chion – em esclarecer,

desde o início, as *dificuldades de passagem* de uma disciplina a outra: é a que se emprega, por exemplo, o estudo das *correlações* entre sinal físico e música, que é também o estudo de suas não-correlações (dessemelhanças, diferenças de interpretação, termos com sentidos duplos, mal-entendidos quanto a noções e, finalmente, erros materiais sobre os fatos e nos enunciados). (CHION, 1983: 89-90)

É de significativo, assim, que a questão da correlação entre sinal físico, música e escuta encontre-se no cerne das reflexões sobre interdisciplinaridade, já nas primeiras reflexões voltadas à música eletroacústica. De fato, artistas e pesquisadores envolvidos com processos criativos e estudos do som mediados por recursos tecnológicos dificilmente podem prescindir de uma delimitação explícita ou implícita dessas assimetrias e *dificuldades de passagem* – sob o risco de atualizar certa atitude cientificista que marcou de maneira um tanto caricata o pensamento composicional pós-vanguardista, sobretudo na década de 1950.

No contexto mais amplo da pesquisa que o presente texto integra será necessário, de maneira análoga ao que Schaeffer faz nos capítulos VII e XIV do seu *Traité*, averiguar as potenciais contribuições e assimetrias entre as técnicas de *machine listening* e processos de escuta/solfejo que elas emulam e/ou mediam. Tal delimitação – circunstancial, de toda maneira – ultrapassa o escopo deste texto: buscarei, tão somente, esboçar o panorama atual da questão ao contextualizar alguns recursos técnicos e algumas práticas relacionadas.

3. A confluência de práticas de escuta e criação no computador

O pós-guerra é marcado por dois desenvolvimentos consideravelmente importantes do ponto de vista da transformação das práticas criativas no campo da música. Nesse período se desenvolvem as primeiras *músicas eletroacústicas* e as práticas criativas a elas associadas e, também, se tem o desenvolvimento inicial da computação musical.

Enquanto as primeiras pesquisas repercutiram imediatamente em uma rápida transformação da criação musical/sonora, o desenvolvimento e a disseminação da computação musical dependeram do aperfeiçoamento, invenção e transformação de elementos e objetos técnicos que só viriam a atingir maior maturidade e disponibilidade a partir da década de 1980. Das pesquisas iniciais na década de 1950, inicialmente voltadas ao cálculo de parâmetros de escrita composicional (HILLER, 1959) e aos primeiros esforços voltados à síntese e ao processamento digital do som, resultaram a sensível transformação do artesanato eletroacústico e da composição instrumental. Se no primeiro caso pode-se detectar uma absorção e uma transformação crescentes de práticas de estúdio a partir de sua implementação no computador – o que o estabelece cada vez mais como ferramenta privilegiada para a mediação de práticas de criação e escuta que antes se davam a partir do conjunto técnico do estúdio –, no campo da composição instrumental surgiram práticas como a *composição algorítmica* e a *composição assistida por computador*.

Em 1977 William Buxton já diferenciava esses termos e as músicas criadas com o computador: de um lado, os “programas que compõem” (*composing programs*) e, do outro, a abordagem que denominaria como “composição assistida por computador” (BUXTON, 1977, p. 64). Já Otto Laske definiu a “composição assistida por computador” como abordagem intermediária entre a “composição manual” e as “máquinas que compõem” (*composing machines*), ressaltando que o contínuo entre uma prática e outra se daria em função do “grau de controle”, por parte do compositor, do processo composicional (LASKE, 1981: 55).

No contexto de uma tendência de exploração conjunta de possibilidades instrumentais e eletroacústicas que já se manifestava nos primeiros experimentos com live-electronics do grupo de Colônia, das composições mistas para instrumentos e fita da *tape*

music norte-americana e de abordagens tecnomórficas diversas de composição instrumental, o computador integrou ferramentas e abordagens tradicionalmente associadas à música eletroacústica àquelas da composição instrumental. Se tal tendência já se prenunciava na própria concepção de linguagens voltadas à computação musical da família MUSIC V e Csound (estruturadas, metaforicamente, a partir de arquivos denominados “orquestra” e “partitura”³) o desenvolvimento dos computadores favoreceu o convívio de processos e estratégias de criação que antes davam-se a partir de artesanatos e práticas criativas consideravelmente distintas – o que pode ser mensurado pela sensível diferença de suas técnicas e seus suportes tradicionais: a saber, a fita magnética e o pentagrama.

Instrumento polivalente e sujeito a ser reconfigurado e adaptado a diversas finalidades, o computador não apenas pode ter suas funções radicalmente reprogramadas como, também, seus mecanismos de *registro e escrita* são igualmente adaptáveis a partir da multiplicidade de formatos de arquivos/*streaming* e da atomização e homogeneização da *escrita* em bits e bytes. Com isso, portanto, a própria noção de *suporte* perde a relevância já que os bits podem ser armazenados em diferentes “suportes” (CD, disco rígido, etc.) sem alteração dos dados registrados. Além disso, os processos de manipulação desses dados não são essencialmente diferentes: em última análise, independentemente de arquiteturas específicas, algoritmos digitais operam a partir da permutação de bits segundo instruções dadas. Do ponto de vista da técnica, enfim, não importa se o que representam os bits são informações relacionadas à cálculo e à escrita musical ou se são amostras de um áudio digital.

Portanto, a própria “mecânica” de funcionamento do computador enquanto objeto técnico propicia uma continuidade (ou certas “continuidades”) entre práticas criativas distintas como a criação eletroacústica, a criação instrumental (incluindo a improvisação), a composição escrita/simbólica e práticas mais recentes voltadas ao vídeo e a sistemas de interação física. Tanto é possível, por exemplo, compor música instrumental a partir da escuta e transcrição mediada por computador de modelos sonoros como, reciprocamente, torna-se possível também criar peças eletroacústicas a partir de uma abordagem voltada à determinação paramétrica de processos sonoros – seja isso realizado a partir de práticas de música generativa, seja, como já entrevia Schaeffer ser um uso mais interessante das “máquinas de cálculo”, na ressíntese a partir de dados previamente obtidos mediante análise.

Nós pensamos que é necessário fornecer ao sintetizador não alguns dados brutos e brutais, como o fez Stockhausen, mas uma grande quantidade de dados e, sem dúvida, dados paramétricos instantâneos para constituir cada um dos sons. A síntese se apresentaria então não como um ato gratuito de inventor ou um capricho “rigoroso” de autor,

mas como o recíproco de análises prévias, como foi sempre praticado nas tecnologias. A análise que nós preconizamos nessa obra consiste em observar os sons naturais, ou ainda, se eles são artificiais e inauditos, a observar *critérios perceptivos*, cujas combinações originais são inumeráveis. Os sons naturais consistiriam então *modelos* cujas propriedades poderiam ser retraduzidas ou desenvolvidas pela máquina, suscetível de “nutrir o som”, a cada instante, como o próprio executor. Esse nos parece ser não apenas o possível, mas o desejável, senão o rentável. (SCHAEFFER, 1966, p. 632–632, t.n)

4. *Machine listening*: escuta assistida, mediada e automatizada

Nick Collins define *machine listening* como a “capacidade das máquinas em simular as habilidades aurais e musicais humanas” (COLLINS, 2011: 439). Se essa possibilidade está na própria gênese dos instrumentos de transdução que possibilitaram a prática eletroacústica⁴, é apenas a partir do desenvolvimento de técnicas mais elaboradas que certas particularidades da escuta humana tornaram-se passíveis de servir como modelo estrutural para processos mais elaborados de análise e extração de informação de sinais de áudio.

Esse desenvolvimento está relacionado, evidentemente, ao surgimento de novos instrumentos tecnológicos. É o caso dos primeiros espectrógrafos baseados em circuitos heteródinos, desenvolvidos ainda na década de 1940 (KOENIG et al, 1946), e dos analisadores de espectro em tempo real implementados a partir do final da década de 1950 – inicialmente valendo-se, também, de circuitos heteródinos e, em seguida, empregando as primeiras implementações otimizadas da Transformada Discreta de Fourier (DEERY, 2007).

Ao tornar possível o exame detalhado dos componentes frequenciais de sinais de áudio, tais ferramentas logo influenciaram o *solfejo* composicional. O já célebre espectrograma da nota mi de um trombone (GRISEY, 2008: 131) utilizado como modelo para a composição das peças *Périodes* (1974) e *Partiels* (1975) é, antes que uma ferramenta de escrita/composição assistida, uma ferramenta de *escuta assistida*. Nesse contexto, o espectrograma é empregado antes de mais nada para fazer com que gestos visuais tecnograficamente extraídos do áudio e transcritos automaticamente pela máquina informem visualmente a escuta e o subsequente processo de escrita/transcrição do compositor.

O desenvolvimento ulterior de técnicas de análise de som se deu apenas partir da concomitante maturação de ambientes mais avançados de síntese/processamento do som (o que viria a ocorrer apenas a partir da década de 1990). Assim, enquanto trabalhos iniciais voltados à *machine listening* baseiam-se prioritariamente em streamings de dados simbólicos

(MIDI) – como em ROWE (1991) –, trabalhos seguintes passariam a empregar recursos de análise de sinal mais sofisticados para implementar computacionalmente processos de *escuta*, composição e transcrição (JEHAN, 2005; KLAPURI e DAVY, 2006).

No campo da computação musical, foi um marco, nesse sentido, a implementação do aplicativo *hetro* – desenvolvido desde 1986 por Tom Sullivan⁵, como parte do pacote Csound – e que permitia a decomposição de sinais de áudio em *pitch tracks* que poderiam ser utilizadas em processos de ressíntese aditiva. Já em ambientes voltados ao processamento/síntese em tempo real, a introdução de extensões voltadas à análise de sinais de áudio desenvolvidos por Miller Puckette para os ambientes Pure Data e Max/MSP – como os objetos *fiddle~* e *sigmund~* – abriram ainda novas possibilidades, disponibilizando à criação musical/sonora processos de *machine listening* de detecção de f0 e *pitch tracking*.

Mais recentemente, pesquisas voltadas à implementação de processos de *Music Information Retrieval* e *machine listening* resultaram em outros *softwares* voltados ao modelamento e à implementação de processos de escuta. Sem a pretensão de empreender uma listagem exaustiva, podemos elencar: em *Pure Data*, além dos objetos anteriormente mencionados de Puckette, extensões como *timbreID*, de William Brent, e as abstrações PDescriptors, de Adriano Monteiro (MONTEIRO, 2011; 2012); em *Max/MSP*, as extensões de Tristan Jehan; e, em *SuperCollider*, a biblioteca SCMIR, de Nick Collins⁶.

O uso de tais extensões é, evidentemente, muito mais aberto a experimentações criativas e a usos variados do que softwares cujo paradigma pré-estabelecido de interação delimita as possíveis utilizações desses recursos – como no já mencionado AudioSculpt ou mesmo em softwares como Sonic Visualizer⁷. Disponibilizados no formato de extensões e em código aberto, esses recursos computacionais podem ser reconfigurados e interconectados de maneira a subsidiar a estruturação de processos os mais variados de escuta mediada ou automatizada através de recursos computacionais: o que vai desde a sua utilização para a inferência de parâmetros específicos no contexto de escrita musical assistida por computador até utilizações mais voltadas à automatização completa de processos de escuta pela máquina.

É o caso, por exemplo, de *Autocousmatic*, de Collins. Valendo-se da já mencionada SCMIR, desenvolvida pelo autor, é descrita, no resumo, como

...um sistema algorítmico que cria música de arte eletroacústica usando processos de machine listening no ciclo de design. (...) Um sistema de processamento interativo e automático é conectado a componentes de machine listening, o que inclui a avaliação do mérito de arquivos intermediários para continuarem até um estágio final de mixagem. A geração da forma estrutural das peças de saída utiliza modelos derivados de um pequeno corpus de músicas eletroacústicas

exemplares, e uma técnica de *dynamic time warping* que serve na medição da similaridade extraída dos processos de *music information retrieval* e é empregada para decidir entre candidatas a mixagens finais. (COLLINS, 2012, p.1)

5. A situação da escuta e do solfejo com e pela máquina

A criação algorítmica de Collins, bastante particular quanto ao uso de recursos de *machine listening*, levanta diversas questões. O processo implementado não se resume a uma “composição eletroacústica algorítmica”: antes de simplesmente automatizar o que Schaeffer denominava por *thème* (o fazer musical associado à composição e à criação de sons) Collins emula o que Schaeffer denominava por *version* (a escuta e a descrição metódica de sons).

Certamente provocativa, já que uma das “avaliações dos resultados” do trabalho consistiu em mandar peças compostas pelo processo para concursos e chamadas de obras, sua proposta aproxima a questão da escuta e o projeto de pesquisa de Schaeffer (portanto, o próprio *solfejo generalizado*) de problematizações que se relacionavam, nas décadas de 1950 a 1970, às primeiras empreitadas da composição com o computador. Analogamente àquele contínuo imaginado por Laske entre uma *composição manual* e uma *composição pela máquina*, pode-se pensar em uma linha de continuidade entre a *escuta humana* (ou, mais amplamente, *biológica*) e a *escuta pela máquina*. Nesse contínuo, poderíamos talvez incluir técnicas já estabelecidas de escuta/*version*: todo o *Solfège de l’Objet Sonore*, aliás, poderia ser situado, junto com utilizações da *machine listening* como recurso de escuta e solfejo mediados por recursos tecnológicos. Se o som registrado e passível de uma auscultação exploratória com recursos eletroacústicos somou-se aos sons dos instrumentos tradicionais em processos os mais diversos de *solfejo* – entendido, evidentemente, enquanto prática muito mais ampla que aquela da leitura de pentagramas sem uso de instrumentos quaisquer –, as técnicas de *machine listening* associadas a recursos de visualização e discretização de informações introduzem possibilidades que não deveriam ser menosprezadas, seja em contextos de criação, seja em contextos pedagógicos.

Em um caso ou em outro, as transformação das técnicas, induzem, também, uma renovação das práticas que deles fazem uso e a uma reconsideração conceitual mais ampla. Nesse sentido, se a noção de *solfejo* foi de tal modo ampliada por Schaeffer, é inevitável a pergunta: pode o *solfejo*, ainda, ser definido como a arte de se “exercitar em melhor escutar”? (SCHAEFFER, 1966, p. 62). Uma hipótese a se considerar é que, tendo em vista as técnicas e recursos disponíveis na atualidade e as práticas de criação, investigação e escuta que delas fazem uso, torna-se cada vez mais relevante um *solfejo dos objetos técnicos*. Em outros

termos, um estudo ao mesmo tempo criativo, técnico e reflexivo voltado à compreensão, transformação e desenvolvimento de recursos técnicos envolvidos em processos de escuta, criação e pesquisa.

Referências:

- BUXTON, William. A composer's introduction to computer music. **J. of New Music Res. Journal of New Music Research**, v. 6, n. 2, p. 57–71. 1977.
- CHION, Michel. **Guide des objets sonores Pierre Schaeffer et la recherche musicale**. Paris: Buchet-Chastel : Institut national de la communication audiovisuelle, 1983.
- COLLINS, Nick. Machine Listening in SuperCollider. In: WILSON, Scott; COLLINS, Nick; COTTLE, David (Eds.). . **The SuperCollider book**. Cambridge, Mass: MIT Press, 2011.
- COLLINS, Nick. Automatic composition of electroacoustic art music utilizing machine listening. **Computer Music Journal**, v. 36, n. 3, p. 8–23. 2012. . Acesso em: 14 jun. 2014 .
- DEERY, Joe. The “Real” History of Real-Time Spectrum Analyzers. **Sound & vibration**, v. 41, n. 1, p. 54–59. 2007. . Acesso em: 3 mai. 2015 .
- GRISEY, Gérard. **Écrits ou L'invention de la musique spectrale**. [Paris]: MF, , 2008.
- HILLER, Lejaren; ISAACSON, Leonard M. **Experimental music; composition with an electronic computer**. New York: McGraw-Hill, , 1959.
- JEHAN, Tristan. **Creating music by listening**. [s.l.] Massachusetts Institute of Technology, 2005. Disponível em: <http://web.media.mit.edu/~tristan/Papers/PhD_Tristan.pdf>. Acesso em: 20 jun. 2014 .
- KLAPURI, Anssi; DAVY, Manuel. **Signal processing methods for music transcription**. New York: Springer, , 2006. Disponível em: <<http://site.ebrary.com/id/10161140>>. Acesso em: 16 jun. 2014 .
- KOENIG, W.; DUNN, H. K.; LACY, L. Y. The Sound Spectrograph. **The Journal of the Acoustical Society of America**, v. 18, n. 1, p. 19–49. 1946. . Acesso em: 3 mai. 2015 .
- LASKE, Otto. Composition Theory in Koenig's Project One and Project Two. **Computer Music Journal**, v. 5, n. 4, p. 54–65. 1981. . Acesso em: 4 mai. 2015 .
- MONTEIRO, Adriano. **Criação e Performance Musical no Contexto dos Instrumentos Musicais Digitais**. Campinas: Universidade Estadual de Campinas, 2012.
- PADOVANI, José Henrique. “Acerca da transdução: princípios técnicos, aspectos teóricos e desdobramentos”. In: Anais do XXIV Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música. São Paulo: 2014.
- ROWE, Robert James. **Machine listening and composing—making sense of music with cooperating real-time agents**. [s.l.] Massachusetts Institute of Technology, 1991. Disponível em: <<http://dspace.mit.edu/handle/1721.1/13835>>. Acesso em: 16 jun. 2014 .
- SCHAEFFER, Pierre. **Traité des objets musicaux**. Paris: Éditions du Seuil, , 1966.

Notas

¹ Ver, por exemplo, os aplicativos Shazam e Soundhood <<http://www.shazam.com/>> e <<http://www.soundhound.com/>>. Acesso: 2 de Maio de 2015

² Ver <<http://www.orelia.fr/en/security/audiosense-the-audio-surveillance.html>> Acesso: 2 de Maio de 2015

³ Em Csound processos de síntese eram originariamente codificados em um arquivo .orc enquanto os parâmetros de execução desses processos ao longo do tempo eram determinados em um arquivo .sco.

⁴ Quanto a isso, é significativo que a concepção original do microfone de Johann Philipp Reis seja baseada no aparelho auditivo. A esse respeito, ver PADOVANI, 2014.

⁵ Ver comentários do código do aplicativo em

<http://csound.sourceforge.net/documentation/4.23f12/hetro_8c-source.html>. Acesso: 2 de Maio de 2015

⁶ Ver, respectivamente: <<http://williambrent.conflations.com/pages/research.html>>;

<<https://sites.google.com/site/pdescriptors/>>; <<http://web.media.mit.edu/~tristan/>>. Acesso: 2 de Maio de 2015.

⁷ Ver: <<http://www.sonicvisualiser.org/>> e <<http://anasynth.ircam.fr/home/english/software/audiosculpt>>.

Acesso: 2 de Maio de 2015.