

Listas de conteúdos disponíveis em [ScienceDirect](#)

Relatórios Físicos

www.sciencedirect.comFourier e a ciência de hoje / *Fourier e a ciência de hoje*

Fourier no coração da música de computador: de sons harmônicos a texturas

*Fourier no coração da música de computador: de sons harmônicos a texturas*Vincent Lostanlen ^{no, *}, Joakim Andén ^b, Mathieu Lagrange ^{vs}^{no} Laboratório de pesquisa de música e áudio, New York University, New York, NY, EUA^b Center for Computational Mathematics, Flatiron Institute, New York, NY, EUA^{vs} LS2N, CNRS, Escola Central de Nantes, Nantes, França

item	informação	resumo
<i>História do artigo</i>		
Disponível online xxxx		Além do âmbito da condução térmica, o tratado de Joseph Fourier sobre o <i>Teoria Analítica do Calor</i> (1822) alterou profundamente nossa compreensão das ondas acústicas. Ele postula que qualquer função de período unitário pode ser decomposta em uma soma de sinusóides, cujas respectivas contribuições representam alguma propriedade essencial do fenômeno periódico subjacente. Na acústica, tal decomposição revela os modos ressonantes de uma corda que vibra livremente. A introdução da série de Fourier, portanto, abriu novos caminhos de pesquisa sobre a modelagem do timbre musical - um tópico que se tornaria de importância crucial na década de 1960 com o advento dos sons gerados por computador. Este artigo se propõe a revisitar o legado científico de Joseph Fourier através das lentes da pesquisa em música computacional. Discutimos primeiro como a série de Fourier marcou uma mudança de paradigma em nossa compreensão da acústica, suplantando a teoria da consonância dos harmônicos no monocórdio pitagórico. Então, destacamos a utilidade do paradigma de Fourier por meio de três problemas práticos de análise - síntese: a imitação de instrumentos musicais, a transposição de frequência e a geração de texturas de áudio. Curiosamente, cada um desses problemas envolve uma perspectiva diferente sobre a dualidade tempo-frequência e estimula uma interação multidisciplinar entre a pesquisa e a criação que ainda está em andamento.
<i>Palavras-chave</i>		
Análise de Fourier		
Música de computador		
Processamento de sinal de áudio		
<i>Palavras-chave</i>		
Análise de Fourier		
Música de computador		
Processamento de sinal de áudio digital		

© 2019 Academia de Ciências. Publicado por Elsevier Masson SAS. Todos os direitos reservados.

resumo

Além de sua contribuição teórica no campo da condução térmica, a tese de Joseph Fourier sobre o *Teoria analítica do calor* (1822) revolucionou nossa concepção de ondas sonoras. Esta tese confirma que qualquer função de período unitário é decomposta em uma série de sinusóides, cada uma representando uma propriedade essencial do fenômeno periódico estudado. Na acústica, essa decomposição revela os modos de ressonância de uma corda vibrante. Assim, a introdução da série Fourier abriu novos horizontes na modelagem do timbre musical, tema que assumiu importância crucial a partir da década de 1960, com o início da música por

* —Autor correspondente.

Endereço de e-mail: vincent.lostanlen@nyu.edu (V. Lostanlen), janden@flatironinstitute.org (J. Andén), mathieu.lagrange@cnsr.fr (M. Lagrange).<https://doi.org/10.1016/j.crhy.2019.07.005>

1631-0705 / © 2019 Academia de Ciências. Publicado por Elsevier Masson SAS. Todos os direitos reservados.

Please cite this article in press as: V. Lostanlen et al., Fourier at the heart of computer music: From harmonic sounds to texture, C. R. Physique (2019), <https://doi.org/10.1016/j.crhy.2019.07.005>

computador. Este artigo se propõe a tematizar a obra de Joseph Fourier à luz de suas implicações na pesquisa musical. Em primeiro lugar, refazemos a mudança de paradigma que a série Fourier operou na acústica, suplantando um modo de pensamento baseado nas consonâncias do monocórdio pitagórico. Posteriormente, destacamos o interesse do paradigma de Fourier por meio de três problemas práticos de análise-síntese: a imitação de instrumentos musicais, a transposição de frequências e a geração de texturas sonoras. Cada um dos seus três problemas evoca uma perspectiva diferente sobre a dualidade tempo-frequência e dá origem a um diálogo multidisciplinar entre investigação e criação que ainda hoje é relevante.

© 2019 Academia de Ciências. Publicado por Elsevier Masson SAS. Todos os direitos reservados.

1. Introdução

“Os números com os quais um computador lida podem ser convertidos em sons que o ouvido pode ouvir?” Em um artigo de 1963 intitulado *O Computador Digital como Instrumento Musical*, O engenheiro da Bell Labs, Max Mathews, levanta esta questão visionária, à qual ele responde imediatamente de forma afirmativa [1] Ele argumenta que, de fato, o refinamento da conversão digital para analógico permite a composição da música sob a forma de uma sequência de tempo discreto de valores de amplitude. Depois de converter essa sequência em um sinal de tensão elétrica de tempo contínuo, o computador, por indução eletromagnética de algum alto-falante, emitiria uma onda semelhante às produzidas por instrumentos convencionais.

Embora o protocolo descrito acima possa parecer banal para o leitor moderno, é importante ressaltar a ruptura histórica que representou, na época, no que diz respeito às tecnologias anteriores. Por um lado, as fitas cassete tinham uma excelente fidelidade em termos de reprodução de material pré-gravado, mas ofereciam pouca flexibilidade para manipulação posterior: variações na velocidade da fita, por exemplo, normalmente afetariam o andamento e o tom proporcionalmente. Por outro lado, os osciladores analógicos tinham vários botões e controles desilzantes para um controle paramétrico preciso, mas não tinham a capacidade prática de aproximar uma gama diversificada de sons do mundo real. Mathews estava ciente desse enigma e pretendia demonstrar que os computadores poderiam, em um futuro próximo, obter um melhor equilíbrio entre expressividade e controle do que qualquer outra tecnologia disponível na época.

A perspectiva de usar um computador para renderizar sons veio com uma advertência importante. Em virtude do teorema de amostragem de Nyquist - Shannon, codificando uma onda contínua de largura de banda B sem perda de informação requer $2B$ amostras discretas por segundo [2] Dado que o sistema auditivo típico de humanos tem uma largura de banda de cerca de $B = 20$ kHz, esse número equivale a cerca de 40.000 samples por segundo, ou seja, milhões de samples para uma peça musical de poucos minutos. Por mais atraente que pareça manter o controle completo sobre a evolução temporal da peça - em escalas de tempo de microssegundos -, a tarefa de ajustar independentemente o valor da amplitude de cada amostra parecia, para o músico, ser de Sísifo.

“A conversão de números em som”, ressalta Mathews, “é inútil musicalmente, a menos que um programa adequado possa ser desenvolvido para calcular as amostras de um único conjunto de parâmetros.” A presença de uma unidade central de processamento (CPU) dentro do computador digital, além de componentes de armazenamento e acionamento, alivia o design de sons no domínio do tempo. Consequentemente, o desenvolvimento da música computacional requer um conjunto de habilidades na interseção de acústica, processamento de sinais e ciência da computação. A combinação dessas habilidades foi defendida por Mathews e colaboradores, notavelmente John Chowning e Jean-Claude Risset, e posteriormente promovida por muitos outros.

Este artigo propõe uma história resumida dos modelos matemáticos para a produção e percepção de sons musicais. Ao longo desta busca contínua por um musical *língua franca* entre músicos e cientistas, o legado de Joseph Fourier desempenha um papel fundamental em pelo menos três aspectos. Primeiro, representar uma função periódica por sua série de Fourier ordenadamente direangles sua frequência fundamental, um parâmetro contínuo e unidimensional, de seu envelope espectral, uma sequência potencialmente infinita de números de valores complexos. Em segundo lugar, o algoritmo de transformada rápida de Fourier (FFT) permite convoluções eficientes entre quaisquer dois sinais de tempo discreto, mesmo se um deles tiver uma resposta de impulso infinita. Em terceiro lugar, no estudo de texturas de áudio, o cálculo da transformada de Fourier da função de autocovariância revela a densidade espectral de potência do processo estacionário subjacente. Nenhum desses tópicos de pesquisa foi devidamente formulado, muito menos abordado, pela lei da ressonância aritmética entre os harmônicos de uma corda vibrante, ¹) No entanto, em cada um deles, o recurso à teoria de Fourier preenche bem a lacuna entre as representações numéricas e perceptivas do som musical.

Uma revisão abrangente das técnicas modernas de processamento de áudio baseado em Fourier está além do escopo deste artigo; para este fim, referimos o leitor a [3] Em vez disso, optamos por restringir nossa narrativa a cinco métodos básicos: o modelo senoidal aditivo, a transformada de Fourier de curto prazo (STFT), o cepstrum de potência, a síntese de modulação de frequência (FM) e o vocoder de fase. Sua integração em estações de trabalho de áudio digital moldou a história da música artística do pós-guerra e também fez seu caminho para a música pop. Neste artigo, nossa ambição principal é mostrar que a dualidade entre o domínio do tempo e o domínio de Fourier traz coerência conceitual para a profusão de sistemas musicais de computador que se seguiram à descoberta de Mathews. Além disso, mostramos que, embora o paradigma de Fourier *estritamente falando* é insuficiente para sintetizar fontes aperiódicas, este paradigma pode ser estendido a um nível maior de generalidade, recorrendo a um esquema de multiresolução. Nas últimas duas décadas, pesquisadores em processamento de sinais de áudio desenvolveram várias alternativas para o Fourier de curto prazo

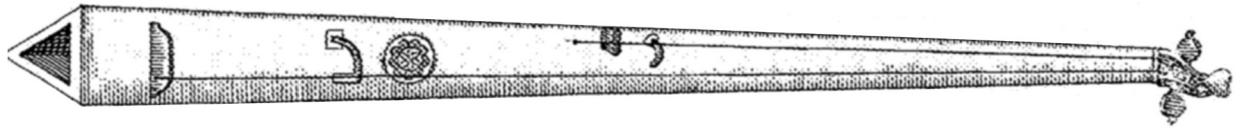


FIG. 1 Desenho de um monocórdio de Athanasius Kircher, publicado em seu tratado *Musurgia Universalis* (1650).

transformar para melhorar a análise - síntese de texturas de áudio naturais, como gravações de campo da vida selvagem. Este artigo apresenta uma dessas alternativas, a saber, o espalhamento conjunto tempo-frequência, que se baseia na extração de modulações espectrotemporais no escalograma wavelet. Ao fazer isso, nosso objetivo não é oferecer um relato definitivo do estado da arte em síntese de texturas de áudio, mas discutir a relevância do paradigma de Fourier nos empreendimentos atualmente ativos de pesquisa em música por computador.

Seção 2 refaz a história dos modelos reducionistas de ondas acústicas, do monocórdio pitagórico ao de Fourier.

Teoria Analítica do Calor, seguido por sua implementação digital na década de 1960. Seção 3 apresenta alguns recursos avançados de técnicas baseadas em Fourier para a análise e síntese de sons não estacionários. Por último, Seção 4 estende o paradigma de Fourier a uma estrutura de multiresolução profunda, introduzindo espalhamento de frequência de tempo e sua aplicação à síntese de textura de áudio.

2. Síntese de série de Fourier separável

As primeiras perguntas sobre a natureza do som vieram de uma perspectiva física: como objetos físicos, como cordas vibrantes, resultam em sons audíveis? Foi somente com a introdução da análise de Fourier que uma compreensão mais precisa dessa questão foi alcançada. Em particular, a noção de uma série de Fourier para descrever uma função periódica provou ser frutífera, primeiro, na análise de sons, mas eventualmente também em sua síntese. Esta última aplicação foi introduzida no trabalho de Max Mathews e seu software MUSIC, que usava uma série de Fourier modulada por um envelope temporal para criar alguns dos primeiros sons musicais gerados por computador.

2.1. Monocórdio pitagórico

Encontrar uma decomposição adequada da música em entidades atômicas é um problema que remonta, pelo menos, à Grécia Antiga e à institucionalização da prova matemática. O monocórdio pitagórico (*kanon*), uma cítara de uma corda compreendendo uma ponte móvel e uma régua graduada, encontra seu mais antigo traço escrito conhecido na obra de Euclides *Divisão do Cãnon*. Pode-se ajustar o comprimento da parte vibrante da corda movendo a ponte a uma distância específica da porca monocórdio (ver Fig. 1) Embora rudimentar e deliberadamente concebido como um experimento mental, a introdução do monocórdio estabeleceu dois princípios essenciais da acústica musical. A primeira é descritiva: todas as outras coisas sendo iguais, o intervalo musical entre duas cordas é determinado pela proporção de seus comprimentos vibratórios. O segundo, em contraste, é prescritivo: intervalos cujas proporções de comprimento irredutíveis têm um pequeno denominador inteiro evocam uma sensação de maior consonância.

Uma vez que esses dois princípios são considerados axiomas, a busca por um temperamento consonantal [4] equivale a o seguinte problema aritmético: o que são duas sequências inteiras finitas p_1, \dots, p_N e q_1, \dots, q_N de modo que todos os pares de termos cruzados p_i e q_j não tenham um pequeno produto comum? Para $N = 3$, a solução é dada pela quarta perfeita (4/3), pela quinta perfeita (3/2) e pela oitava (2/1). Esses três intervalos estão no cerne da teoria da afinação musical de Ptolomeu, onde eles são conhecidos como proporções epimóricas. Escalas diatônicas ($N = 7$), no entanto, necessariamente incorrem em dissonância severa em pelo menos um par de tons, conforme definido pelos axiomas de afinação pitagóricos acima. O tópico de mitigar essa dissonância e buscar uma escala "bem temperada" que procede do método de dividir a corda em subpartes de comprimento racional gerou uma controvérsia que durou dois milênios. Numerosos tratados, de Aristoxeno de Tarento e Boécio a René Descartes, trataram dessa controvérsia de um ponto de vista multidisciplinar, combinando matemática, música e filosofia [5]

Apesar de nunca ter visto a luz do palco, o monocórdio atuou como um paradigma compartilhado por estudiosos, compositores e fabricantes. Na verdade, promoveu, embora um tanto incipiente, duas noções fundamentais na física: ressonância modal e superposição de onda. A ressonância modal afirma que um corpo fixo pode, em seu regime estacionário de vibração, ser compreendido por um número contável de elementos elementares *modos próprios* [6] A superposição de ondas afirma que, na natureza, os corpos vibrantes não se sintonizam exclusivamente com um modo próprio ou outro, mas oscilam de acordo com alguma mistura dos mesmos [7] No jargão dos sistemas dinâmicos, o monocórdio é um excelente exemplo de ângulo reto conceitual entre a forma (condições de contorno) e o estado (condições iniciais).

2.2. A revolução de Fourier

A era do Iluminismo marca o ápice do paradigma do monocórdio, a ponto de revelar suas deficiências intrínsecas. No *Encyclopédie*, iniciada em 1749, Jean-Jacques Rousseau cunhou o termo *carimbado* para referir-se a qualia acústicas de opacas a brilhantes e de azedas a doces. Ele destacou que a percepção do timbre permite ao ouvinte reconhecer a identidade dos instrumentos, mesmo quando esses instrumentos tocam notas de intensidade e altura idênticas. Compreender o timbre requer investigar a interação física entre a forma e o estado e, em particular, como a técnica de execução de um

nota (por exemplo, puxada ou curvada) afeta as magnitudes relativas de todos os modos próprios superpostos ao longo do tempo. No entanto, o paradigma do monocórdio não oferece nenhum protocolo experimental para medir essas magnitudes relativas e, portanto, é inadequado para discutir as bases físicas da percepção do timbre musical. Da mesma forma, em seu tratado sobre *Harmonia reduzida a seus princípios naturais* (1722), o compositor Jean-Philippe Rameau concluiu que as consonâncias no monocórdio não bastavam para explicar a harmonia tonal no período de prática comum. Tanto Rameau quanto Rousseau conversaram com o matemático Jean Le Rond d'Alembert sobre a necessidade de reformular a estrutura acústica herdada da Grécia Antiga. Este d'Alembert levou a escrever um tratado sobre o *Elementos da música* (1752), que reconheceu a obsolescência do monocórdio e propôs algumas direções futuras para o estudo da música. Embora d'Alembert tenha descoberto, pouco antes (em 1749), a equação diferencial parcial governando o movimento das cordas vibrantes, a conexão entre o estado inicial dessa equação e a superposição dos modos próprios permaneceu obscura por mais sete décadas, até a publicação de Joseph Fourier's *Teoria Analítica do Calor* (1822).

Embora não esteja diretamente relacionado à música, o tratado de Fourier representou um ponto de inflexão na compreensão da equação de onda de d'Alembert [8] O tratado se propõe a isolar a questão da propagação do calor do estudo mais amplo do calor, e notadamente de seus aspectos químicos e dinâmicos [9] Ao fazer isso, estabeleceu um ramo novo do nome da física matemática, mais tarde conhecido como análise harmônica. Fourier certamente não foi o primeiro a estudar as séries trigonométricas da forma

$\rho \cos(2\pi p \xi t + \phi_p)$ para sequências arbitrárias de amplitudes ρ e fases ϕ_p : essas séries são encontradas, por exemplo, nos escritos de d'Alembert, bem como de Euler e Bernoulli. No entanto, a inovação de seu tratado reside na afirmação de que tal

as séries trigonométricas são universais, no sentido de que podem ser aplicadas não apenas para construir modos próprios do operador Laplaciano, mas também para representar a condição inicial de qualquer problema unidimensional de condução de calor.

A representação da série de Fourier é mais do que uma etapa computacional intermediária para reduzir as variáveis espaciais e temporais em uma equação diferencial parcial. De acordo com o epistemólogo Alain Herreman [10], inaugura a dualidade entre o conjunto de "linhas curvas arbitrárias" (em francês, *linhas curvas desenhadas arbitrariamente* [8]) e o conjunto de expressões analíticas baseadas em séries trigonométricas. Enquanto o primeiro conjunto é considerado natural como proveniente do mundo real, o segundo conjunto é escrito na linguagem do cálculo e é deliberadamente mantido à parte de qualquer instanciação física. Falando semioticamente, Fourier argumenta que este último está em uma relação de *conformidade incomensurável* (nas palavras de Herreman) com o primeiro. Uma vez traduzida para a acústica, esta afirmação significa que, embora as séries de Fourier não sejam uma consequência necessária da observação da vibração natural das cordas (incomensurabilidade), elas têm o mérito de representar potencialmente

qualquer forma de onda (conformidade). Por unicidade da série de Fourier, a identificação de amplitudes ρ e fases ϕ_p depende de integrais de valor complexo da forma

$$\rho \exp(i\phi_p) = \xi \int_0^{1/\xi} x(t) \exp(-2\pi i p \xi t) dt \quad \forall p \in \mathbb{N} \setminus 0 \quad \text{com} \quad x(t) = \sum_{p=1}^{\infty} \rho \cos(2\pi p \xi t + \phi_p) \quad (1)$$

A brilhante intuição de Fourier é que, na prática, a função x pode codificar a temperatura relativa inicial em uma haste de metal (nesse caso t é uma variável espacial unidimensional), mas também o deslocamento inicial de uma corda vibrante. "Se alguém aplicar esses princípios à questão do movimento das cordas vibrantes", escreve Fourier, "resolverá as dificuldades que sua análise de Daniel Bernoulli havia levantado no passado". Ele continua: "Esta questão difere muito daquela da distribuição de calor; mas as duas teorias têm pontos em comum; porque ambos são baseados na análise de derivadas parciais."

2.3. Selo de modelagem com o envelope espectral

Ao longo dos 142 anos entre a de Fourier *Teoria Analítica do Calor* e de Mathews *Computador Digital como Instrumento Musical*, vários estudiosos progressivamente reconheceram a importância do timbre na percepção musical e a conveniência de pensar sobre o som em termos de tempo e frequência. Nesse sentido, dois marcos históricos são a publicação de um *Tratado de Instrumentação* (1844) pelo compositor Hector Berlioz, e o *Sensações de Tom* (1863) pelo físico Hermann von Helmholtz. Na época em que os computadores digitais se tornaram comuns, o recurso à série de Fourier parecia evidente para Mathews. O principal apelo por trás da adoção da série Fourier como modelo de sinal para x é que eles destacam a frequência fundamental ξ como um parâmetro contínuo e unidimensional da percepção do som, cujo efeito se estende por todas as amostras de tempo discreto t . Além disso, a série de Fourier permite, até certo ponto, abordar a questão da modelagem do timbre parametrizando a amplitude co-

eficientes ρ_p independentemente da escolha da frequência fundamental ξ . Os coeficientes ρ_p correspondem aos valores do envelope espectral da fonte em frequências uniformemente espaçadas $p \xi$, em que o inteiro p corresponde ao índice de um de

os sinusóides, denotados como parciais harmônicos. Este envelope espectral, por sua vez, relaciona-se com a forma e o material do instrumento, bem como com o gesto do intérprete. Portanto, embora a identificação de uma fonte vá além da percepção

de seu envelope espectral, os valores de ρ_p certamente tem um papel central na percepção do timbre. Para o compositor, as escolhas

de ξ e ρ_p assemelham-se a escolhas em harmonia e orquestração, dois atributos bem estabelecidos de estilo artístico na música artística europeia. Tal redeangement é crucial para garantir que um programa de computador possa analisar, processar e renderizar qualquer

pontuação.

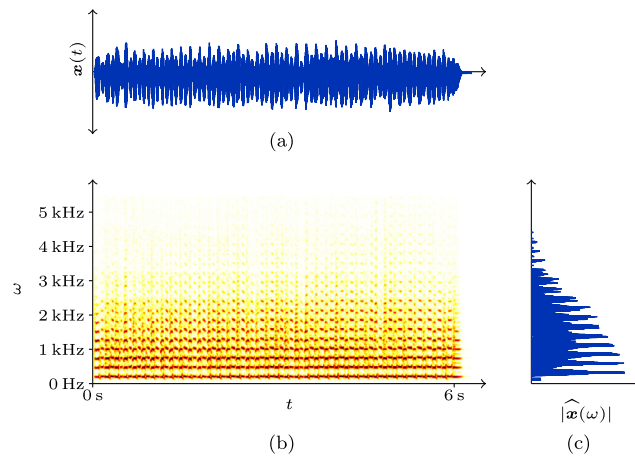


FIG. 2. (a) Uma gravação x de uma trombeta tocando um trinado com tom C4. (b) O espectrograma de x . (c) A magnitude da transformada de Fourier de x .

2.4. O ambiente de software MUSIC

O software que Max Mathews construiu em 1963, simplesmente apelidado de MUSIC, dependia muito da série Fourier para definir o timbre dos instrumentos virtuais. O layout dos primeiros programas de música de computador se assemelhava muito aos circuitos eletrônicos de osciladores e moduladores, com diferentes parâmetros de duração, intensidade e amplitude, como aqueles já em uso na época por compositores de música eletrônica analógica. No entanto, ao contrário de um circuito analógico tangível, a complexidade de um programa de música de computador não é limitada por restrições de operabilidade humana, mas apenas pela velocidade da unidade de processamento central. Considerando que soldar várias centenas de osciladores juntos seria uma tarefa complicada para o compositor, instanciar esses osciladores em um ambiente virtual alivia consideravelmente essa tarefa, pois delega a maior parte da parametrização ao computador. 11)) para ganhar em sofisticação e detalhes. Esses programas foram distribuídos

com uma coleção de valores predefinidos para a sequência de coeficientes de Fourier $no\ p$. Da mesma forma, funções de amplitude pré-definidas

$t \rightarrow \alpha(t)$ permitiu ao usuário modelar o perfil temporal de ataque, decadência, sustentação e liberação (ADSR) dessas notas, como

uma função linear por partes que compreende quatro segmentos. Por parte do compositor, os graus de liberdade criativa são então: frequência fundamental ξ , perfil temporal α , e tempo

de início T . Sob esta estrutura, para um instrumento virtual com

P harmônico $p \in \{1, \dots, P\}$ artiais tocando uma peça com NAO notas, a versão de computador da peça musical é dada por

$$x(t) = \sum_{n=1}^{NAO} \alpha_n(t - T_n) \sum_{p=1}^P no\ p \cos(2\pi p \xi_n t + \phi_p) \quad (2)$$

Para corresponder à qualidade do tom de instrumentos familiares, Max Mathews e sua colaboradora Joan Miller implementaram

Síntese baseada em Fourier em MÚSICA. Ao inferir valores predefinidos de $no\ p$ ao reproduzir a série de Fourier de dados externos não produziu novos sons per se, mas forneceu um ponto de partida conveniente para a exploração da tibia. Porque o número de parciais

P com amplitudes não desprezíveis $no\ p$ era tipicamente da ordem de 20, tal ajuste era humanamente tratável - ao contrário da descrição de amostra da forma de onda $x(t)$ no domínio do tempo, o que normalmente envolveria cem amostras

por período ou mais. Além disso, uma única predefinição de valores $no\ 1, \dots, no\ p$ seria suficiente para abranger uma ampla tessitura de valores de altura

ξ_n , para um determinado instrumento.

Para ilustrar o freq $\int_{-\infty}^{\infty}$ conteúdo uência de um som x , vamos considerar seu transformada de Fourier de curta duração (STFT):

$$\text{STFT}(x)(t, \omega) = \int_{-\infty}^{\infty} x(\text{você}) g(\text{você} - t) \exp(-2\pi i \omega \text{você}) d \text{você} \quad (3)$$

Onde $g(\text{você})$ é a função de janela associada ao STFT. Esta representação de x nos permite examinar seu conteúdo de frequência quando restrito à janela g centrado no tempo t . Tomando seu módulo, obtemos o *espectrograma* $|\text{STFT}(x)(t, \omega)|$.

FIG. 2 (a) mostra a forma de onda de seis segundos de um trompete tocando a nota C4 com um trinado. O espectrograma correspondente, com uma janela de Hann de 46 ms, é mostrado na Fig. 2 (b). Finalmente, a magnitude da transformada de Fourier de todo o sinal (ou seja, não localizada usando uma janela) é mostrada na Fig. 2 (c). Aqui, a estrutura harmônica regular induzida pelo pitch constante é prontamente vista tanto no espectrograma quanto na magnitude da transformada de Fourier. O espectrograma, além disso, revela a dinâmica do sinal em diferentes frequências - cada um traça um envelope distinto. Há também pequenas mudanças no tom, conforme revelado pela ligeira oscilação dos contornos parciais. Como veremos, esses dois fenômenos também contribuem muito para a percepção de um som.

2.5. Parametrizando o envelope espectral

Chá P valores $no\ 1 \dots, no\ P$ pode, por sua vez, ser codificado por ainda menos parâmetros, o que englobaria a forma geral do envelope espectral. Por exemplo, a razão de amplitude entre alguns números parciais ímpares $no\ 2\ p + 1$ e seu vizinho estranho $no\ 2\ p$ tem um significado físico interpretável, em termos de condições de contorno da equação de onda de d'Alembert. O orifício de uma flauta transversal é aberto em ambas as extremidades: como resultado, a razão de energia ímpar-para-par é da ordem de 1. O orifício de um clarinete, em por outro lado, é aberto na extremidade inferior e fechado na outra: a presença da palheta enfraquece os parciais pares, aumentando assim a razão de energia ímpar-para-par [12] Tal percepção sobre os fundamentos físicos da produção de som, derivada de uma decomposição de Fourier, permite ao compositor interpolar entre dois tipos bem conhecidos de timbre musical (flauta e clarinete) controlando um único parâmetro contínuo.

Outro exemplo de um recurso unidimensional para timbre musical que tem uma interpretabilidade física é a taxa de decaimento dos coeficientes de amplitude de Fourier, também conhecido como inclinação espectral. Por um lado, resolver a equação de d'Alembert com velocidade inicial zero e deslocamento inicial diferente de zero no ponto médio da corda produz uma série de Fourier cujos termos de amplitude $no\ p$ são proporcionais a $1 / p^2$ Por outro lado, resolver a mesma equação com deslocamento inicial zero e velocidade inicial diferente de zero leva, todas as outras coisas sendo iguais, a coeficientes de Fourier de magnitudes maiores: para os primeiros valores de p , magnitudes são quase proporcionais a $1 / p$. Essa discrepância na queda é refletida no domínio do tempo por uma discrepância na forma da forma de onda, do triangular ao quadrado. Matematicamente falando, esta dualidade entre regularidade no domínio do tempo e taxa de decadência no domínio de Fourier está no cerne da noção de continuidade de Hölder e espaços de Sobolev [13] Uma decadência mais lenta incorre em um aumento relativo na energia do sinal $x(t)$ em altas frequências, o que é percebido como uma sensação de brilho acústico [14] Na acústica musical, o caso de velocidade inicial zero e deslocamento inicial diferente de zero corresponde a um instrumento com cordas dedilhadas (por exemplo, um violão ou cravo), enquanto o caso de velocidade inicial diferente de zero e deslocamento inicial zero corresponde a um instrumento com cordas marteladas (por exemplo, um piano). Em um ambiente de música de computador, parametrizar as magnitudes de Fourier como proporcionais a $1 / p^s$ para arbitrário s produz uma interface perfeita para interpolação de brilho acústico entre a série típica de Fourier de uma corda dedilhada ($s \rightarrow 1$) e de uma corda martelada ($s \rightarrow 2$). Além da interpretabilidade física das duas características acima mencionadas, eles também respondem por aspectos independentes do envelope espectral. De fato, a razão de energia ímpar para par e a taxa de decaimento dos coeficientes de Fourier caracterizam o contraste local e global entre parciais, respectivamente.

2.6. Power cepstrum

Entre esses dois extremos, é possível definir uma família de descritores que calculam o contraste em todo o espectro de Fourier de acordo com várias escalas, variando de grosseiro (por exemplo, a taxa de decaimento) a fino (por exemplo, o estranho igualar relação de energia). Em um artigo de 1963 de Bruce Bogert, Michael Healy e John Tukey [15], os autores transformam $no\ p$ aplicando um alojamento pontual \sum ritmo seguido (pela fórmula de soma da série de Fourier:

$$vs(k) = \frac{\sum_{p=1}^{+\infty} \text{registro}(no\ p) \exp(2\pi i p k)}{P} \quad (4)$$

A soma de Fourier opera como se a variável de frequência discreta p eram uma dimensão temporal e produz uma variável contínua k , que é fisicamente homogêneo a um intervalo de tempo. Portanto, Tukey propôs referir-se a $vs(k)$ como um poder *cepstrum*, um anagrama na palavra "espectro". Da mesma forma, por meio de um anagrama em "frequência", a variável k é conhecido como um *quefrequency*, e é expresso em Hz -1 , ou seja, em segundos. Em alta quefrequency $k = P/2$, o exponencial complexo na Equação (4) resume-se a $(-1)^k$, e o coeficiente cepstral correspondente se aproxima do logaritmo médio da razão de energia par para ímpar. Por outro lado, em baixa frequência $k = 1/2$, o módulo de $vs(k)$ é proporcional ao expoente s na decadência dos coeficientes de Fourier.

3. Análise - síntese no domínio do tempo - frequência

O recurso a descritores de forma espectral e coeficientes cepstrais é central para a modelagem perceptual de similaridade de timbre e, até hoje, encontra aplicações tanto na análise quanto na síntese de sinais de áudio. No entanto, essas ferramentas são inadequadas para sons não estacionários; isto é, sinais cujo envelope espectral varia ao longo do tempo, seja por efeito da expressividade gestual ou devido a não linearidades do ressonador acústico em relação à amplitude de entrada. Dois exemplos de tais não-estacionariedades são o vibrato no caso do violino e o crescendo no caso do trompete. A imitação desses dois instrumentos, conduzida por Mathews e Risset, respectivamente, não foi concebida como um fim em si mesmo, mas sim como um banco de ensaio bem definido para o desenvolvimento de novos modelos de síntese, em última análise, levando a sons gerados por computador que nunca foram ouvidos antes. Esta seção descreve como estender os princípios fundamentais da análise de Fourier - síntese para melhorar sua gama de aplicabilidade para música de computador.

3.1. Preenchendo a lacuna entre expressividade e controle

Antes da democratização dos programas de música para computador, a ideia de Mathews de usar o computador digital como um instrumento musical foi recebida com ceticismo e descrença. Isso se deveu, em parte, à impressão generalizada de que os computadores

eram mais adequados para o raciocínio baseado em regras, não para a expressão criativa. Com a importante exceção de Ada Lovelace, que, em 1842, havia especulado sobre o uso potencial de *Motor analítico* para escrever música, poucos haviam previsto que o desdobramento temporal de uma peça musical poderia ser obtido como resultado de um cálculo. Em vez disso, até a década de 1960, a pesquisa científica na interseção entre música e tecnologia foi polarizada em torno de dois paradigmas: *música concreta* e *elektronische Musik*. O primeiro, defendido por Pierre Schaeffer no "Studio d'études" de Paris, baseava-se na gravação e manipulação de fitas magnéticas [16] O último, defendido por Karlheinz Stockhausen no Studio for Electronic Music em Colônia, foi baseado na associação modular de osciladores analógicos [17]

Neste contexto, Mathews enfrentou o desafio de demonstrar que os computadores digitais poderiam encontrar uma compensação satisfatória entre a versatilidade de *música concreta* com a controlabilidade de *elektronische Musik*. Portanto, ele empreendeu a tarefa demorada de melhorar a fidelidade dos presets no programa MUSIC, em particular para instrumentos que não foram bem modelados pela Equação (2), ou seja, como uma série de Fourier multiplicada por um envelope ADSR. Uma das principais descobertas de Schaeffer, de fato, foi exibir a natureza inerentemente espectrotemporal da percepção auditiva. Ao manipular o envelope temporal de sons pré-gravados, ele mostrou que a parte de ataque da nota (o A em ADSR) tinha uma importância decisiva na identificação de um instrumento musical. Por exemplo, enfraquecer o potenciômetro durante a parte de ataque do som de um sino é suficiente para apagar seus atributos percussivos e transformá-lo em um tom semelhante ao oboé [18] O mesmo é verdade para muitos outros aspectos dinâmicos da percepção do timbre, como vibrato e tremolo: embora sua influência nos coeficientes cepstrais seja minúscula, eles transmitem uma sensação de "vitalidade" acústica que, quando descartada, pode deixar uma sensação de ausência de afeto robótica.

3.2. Série de Fourier modulada em frequência

Para simular o vibrato típico do violino, Mathews implementou a modulação de frequência no modelo de síntese. De caminho de dois vocês definido por $\sum_{n=1}^P$ ametros, th (em $\alpha_{n\omega}(t - T_{\text{não}})$ (baço de odulação V) a) e a profundidade de modulação, o sinal de áudio é dado por

$$x(t) = \sum_{n=1}^P \alpha_{n\omega}(t - T_{\text{não}}) \cos(\xi_{\text{não}} t + V \sin(Vt)) \quad (5)$$

A equação acima produz notas musicais de frequência fundamental variável no tempo $\xi_{\text{não}} + \cos(Vt)$. Aliás, ele encontrou app = cãtions em música de computador muito além da modelagem de vibrato. Em 1967, John Chowning percebeu que definir o valor de $V \xi_{\text{não}}$, ou mesmo $V > \xi_{\text{não}}$, permitiu sintetizar um conjunto surpreendentemente rico de tons não vibrantes, mas com não mais do que três parâmetros: ξ_1 , V , e [19] Porque sua série de Fourier está implicitamente codificada na composição de duas senoidais, computar uma nota musical por meio da síntese de modulação de frequência (FM) de Chowning é consideravelmente mais rápido do que por meio da síntese aditiva de Mathews. A invenção da síntese FM inaugurou a distribuição de ferramentas de música digital baratas e em tempo real para produções de música pop.

3.3. Modulação espectrotemporal

Além do vibrato, outro efeito musical que está além do alcance do modelo de soma da série de Fourier na Equação (1) é crescendo em instrumentos de sopro. De fato, devido à propagação de ondas não lineares ao longo do furo, bem como variações no movimento labial e perda viscotérmica no sino [20], o brilho acústico de um instrumento de sopro é altamente dependente do nível de som. Voltando ao modelo aditivo da Equação (2), isso implica que as variações temporais da amplitude global $\alpha_{\text{não}}(t - T_{\text{não}})$ deve afetar cada parcial p diferentemente [21] Em 1964, compôs Jean-Claude Risset, durante uma visita ao Bell Labs, implementado $\sum_{p=1}^P$ idade Versão neralizada de síntese aditiva para MUSIC IV. Nesta versão, a forma de onda gerada por computador é

$$x(t) = \sum_{n=1}^P \alpha_{p,n}(t - T_{\text{não}}) \cos(p \xi_{\text{não}} t) \quad (6)$$

onde cada função $\alpha_{p,n}$ representa a curva de amplitude temporal de p para nota não . Risset empregou este modelo não estacionário para sintetizar alguns elementos melódicos em *Suite de Computador para Menino* (1968) e *Mutações* (1969), dois marcos peças no desenvolvimento de música para computador.

Apesar de permitir mais expressividade do que a Equação (2), Equação (6) su ferros de um aumento considerável em plexidade. Isso ocorre porque ele comercializa um modelo separável de envelope temporal $\alpha_{\text{não}}(t - T_{\text{não}})$ e envelope espectral $\alpha_{p,n}(t - T_{\text{não}})$ para um baseado modelo espectrotemporal, em que a amplitude funciona $\alpha_{p,n}(t - T_{\text{não}})$ não necessariamente fatorar como um produto separável de um envelope temporal e espectral. Com essa transição, o recurso a predefinições baseadas em dados, em vez de pura tentativa e erro, tornou-se cada vez mais importante. Para calcular essas predefinições, Risset aplicou a uma família de P filtros passa-banda $\psi_p \xi$ com centro frequências em $p \xi_{\text{não}}$ sobre notas pré-gravadas $y_{\text{não}}$ de frequência fundamental conhecida $\xi_{\text{não}}$.

3.4. Análise de espectrograma e ressíntese

Um wa | y para obter as funções de amplitude $\alpha_{p,n}$ de cada um desses filtros passa-banda é defini-los como $\alpha_{p,n}(t) = | y_{\text{não}} * \psi_p \xi(t) |$, onde o asterisco * denota a operação de convolução

$$(x * y)(t) = \int_{-\infty}^{\infty} x(\tau) y(t - \tau) d\tau \quad (7)$$

e as barras verticais ($|z| = \sqrt{z\bar{z}}$) denotam módulo complexo. Este procedimento foi o prenúncio de um paradigma de longa data na música computacional, conhecido como *análise - síntese*. Aqui, o domínio de Fourier não é nem um ponto de partida nem um ponto final, mas uma etapa intermediária, na qual os cálculos podem ser expressos de forma mais natural do que no domínio do tempo [22] Para produzir um

imitação de uma trombeta, Risset começou analisando um sinal do mundo real *y não* com um dispositivo eletrônico analógico chamado *espectrógrafo de som* [23], inventado no Bell Labs em 1946. Ele então reduziu a descrição de *y não* para alguns, variando lentamente a amplitude funções α_p, n . Por último, ele ressampleou o sinal original por um procedimento semelhante ao somatório da série de Fourier. Após seu retorno à França em 1965, Risset apresentou seus tons de trompete sintetizados para a Academia Francesa de Ciências [24]

Em 1966, a análise espectral de áudio repentinamente se tornou menos exigente computacionalmente, graças à invenção da transformada rápida de Fourier (FFT) por James Cooley e John Tukey. Na verdade, o FFT trouxe a complexidade das convoluções de tempo discreto entre dois sinais de comprimento *eu* a partir de $O(L^2)$ até $O(L \log L)$ registro *EU*). A transformada discreta de Fourier $F(\psi)$ de alguns

filtro passa-banda ψ de comprimento *eu* é definido como: \sum

$$F(\psi): \omega \in \{0, \dots, eu-1\} \rightarrow \mathbb{C} = \sum_{\tau=0}^{eu-1} \psi(\tau) \exp(-2\pi i \omega \tau / eu) \quad (8)$$

O teorema da convolução circular afirma que a aplicação da transformada discreta de Fourier converte um produto de convolução em uma multiplicação elemento a elemento: $F(x * y) = F(x) \times F(y)$. Portanto, é possível medir com eficiência todas as amplitudes

modula = tio | n Ff - Unti ons por m

$\alpha_p, n(t) = \sum_{\tau=0}^{eu-1} y(\tau) \psi_p(\xi) \exp(-2\pi i \omega \tau / eu)$ A transformação de fase é realizada no domínio do tempo:

frequência mental ξ , e até mesmo para sinais de entrada *y não*(*t*) sem nenhuma frequência fundamental discernível. O uso de filtros passa-banda ψ generaliza modelos de tempo - frequência de uma série de Fourier de curto prazo com um período dependente do sinal $\xi - 1$ para um

STFT com um tamanho de frame agnóstico de sinal *EU*, sob uma suposição de estacionariedade local, de sentido amplo. Essa generalização é importante para a música computacional, pois abre a possibilidade de realizar análises - síntese em sons aperiódicos, ruidosos ou mesmo fractais [25]

3.5. Fase de Vocoder e aplicações

O paradigma de análise - síntese é o equivalente digital do canal *vocoder*, inventado por Homer Dudley na Bell Labs em 1940 [26] Ambos os paradigmas, o analógico e o digital, contam com a transformada de Fourier para mapear os sinais em

o domínio da frequência do tempo. A formulação digital, no entanto, oferece controle mais fino sobre α_p, n e, notadamente, a possibilidade de realizar transposição de frequência e alongamento de tempo. Esta ideia foi proposta em 1966 pelo engenheiro da Bell Labs James Flanagan,

sob o nome do *fase vocoder* [27] Risset cunhou os termos de *microcirurgia sônica* e *transformações íntimas* para denotar tais operações de alongamento de tempo artificial e transposição de frequência [28] Nós nos referimos a [29] para uma introdução ao vocoder de fase e ao [30, 31] para uma visão geral do estado da arte no domínio.

Contanto que a quantidade de tempo de alongamento seja pequena ou suficiente, o resultado gerado por computador permanece perceptualmente realista e dificilmente pode ser discernido de uma gravação real em um novo tempo. Assim, o vocoder de fase encontrou aplicações na pontuação em tempo real seguindo [32], de modo a ajustar o andamento de um acompanhamento pré-gravado às flutuações de andamento expressivas de um artista humano. No entanto, para valores extremos de modificação da escala de tempo, a codificação de fase produz sons que não são mais realistas, mas ainda podem ser interessantes musicalmente. Por exemplo, o compositor Trevor Wishart e o cientista da computação Mark Dolson lançaram *Vox 5* em 1986, uma peça musical de computador em que uma voz falada se estende ao longo do tempo, perde progressivamente suas características reconhecíveis e estranhamente se transforma em sons não-falados, como um cavalo vizinho ou um trovão estrondoso [33]

Outra aplicação do vocoder de fase é o ajuste de pitch em tempo real. Esta tecnologia fez o seu caminho para a música pop sob a marca registrada de "Auto-Tune". Depois de rastrear a curva de pitch $\xi_{\text{não}}(t)$ de cada nota *não* em um stream de áudio ao vivo, Auto-Tune substitui $\xi_{\text{não}}(t)$ por seu vizinho mais próximo sobre alguma escala musical pré-definida. Essa substituição produz uma voz cantada perfeitamente afinada, muitas vezes a ponto de soar estranha. Até hoje, alguns músicos abraçaram isso estranhamente e usar o vocoder de fase para construir, nas palavras do artista Jace Clayton, um "dueto entre o eletrônico e o pessoal" [34]

4. Dispersão ondulada de texturas de áudio

De tudo o que foi dito acima, parece que a invenção da transformada rápida de Fourier permitiu que os pesquisadores da música computacional se afastassem do modelo rígido da série harmônica e explorassem o espaço de design da modulação de amplitude (AM), bem como da modulação de frequência (FM). Na verdade, a modelagem de transientes para instrumentos específicos por Risset e Mathews eventualmente deu lugar ao rastreamento parcial [35] e codificação de fase, dois algoritmos complementares para alongamento adaptativo dos contornos do pitch no domínio da frequência do tempo. No entanto, a aplicabilidade musical desses algoritmos é baseada na suposição de que o material sonoro em questão consiste em uma soma finita de sinusóides, e que esses sinusóides têm lentamente

amplitudes e frequências variadas em comparação com alguma escala de tempo conhecida. Embora essa suposição seja frequentemente válida para muitas amostras de música tonal, ela é insuficiente para texturas de áudio, como grandes grupos orquestrais, tambores e gravações de campo de vida selvagem.

Em alguns casos, continua sendo possível acomodar componentes ruidosos dentro de uma estrutura de modelagem senoidal por um procedimento conhecido como aumento de largura de banda [36]. No entanto, a perspectiva de modelar fontes aperiódicas exige uma generalização multiescala da análise de Fourier - síntese. Conhecida como decomposição wavelet, esta generalização multiescala é adequada para representar sinais altamente oscilatórios, mas também caracterizar fenômenos transitórios, conforme evidenciado por experimentos em psicoacústica [37]. Esta seção apresenta uma abordagem baseada em wavelet para a síntese de textura de áudio. Devemos enfatizar que, nesse aspecto, as wavelets não são de forma alguma hegemônicas; nós nos referimos a [38] para uma visão geral recente do estado da arte em modelagem de textura de som com base na representação STFT.

4.1. Transformada wavelet

Em 1971, logo após seu retorno do Bell Labs, Risset foi convidado pela Universidade de Marselha para fundar uma equipe de pesquisa voltada para a computação musical que, em 1978, passou a fazer parte do Laboratório de Mecânica e Acústica do CNRS. Em Marselha, Risset conheceu o físico quântico Alex Grossmann, dando início a uma colaboração entre o Centro de Física Teórica de Grossmann e seu próprio laboratório. O ano de 1984 marca uma aceleração dessa colaboração, em torno do emergente

tópico de *ondas*—A te (rm cunhado por Grossmann para denotar famílias de funções bem localizadas de forma constante e largura de banda variável. Um exemplo de su) ch função bem localizada é a wavelet de Morlet, definida como

$$\psi_{\lambda}(t) = \lambda \exp \left[-\frac{\lambda^2 t^2}{2Q^2} \times (\exp(i\lambda t) - \kappa) \right] \quad (9)$$

Onde λ é a frequência central, Q é o fator de qualidade, e K é calibrado para que $\psi_{\lambda}(t)$ tem média zero. A representação wavelet tem muitas propriedades teóricas interessantes que muitas vezes a tornam mais adequada para a análise de sinais do que o STFT, que é baseado em uma família de funções bem localizadas de largura de banda constante [39]. De uma perspectiva musical, a mais atraente dessas propriedades é que o fator de qualidade Q define uma diferença apenas perceptível na percepção do tom em torno de cada λ , como um intervalo em uma escala cromática. Por exemplo, configuração $Q = 1$ leva a uma largura de banda relativa de cerca de uma oitava, enquanto $Q = 12$ leva a uma largura de banda de cerca de um semitom em temperamento igual de doze tons. Em contraste, a escolha da janela no STFT define uma largura de banda absoluta em hertz, independentemente da frequência central.

Por um lado, o módulo do STFT produz um espectrograma cujo eixo vertical cresce linearmente com a frequência ω . Por outro lado, a transformada wavelet contínua produz uma representação tempo - frequência

$$\text{você } 1(x)(t, \text{registro } \lambda) = x * \psi_{\lambda}(t) \quad (10)$$

cujo eixo vertical cresce logaritmicamente com frequência λ . A representação **você 1** (x) é conhecido como o *escalograma* de x . Embora uma mudança na frequência fundamental ξ de uma nota musical resultaria em uma escala do eixo vertical do STFT, seu efeito no escalograma wavelet é uma tradução simples, devido a este mapeamento logarítmico. Portanto, o escalograma wavelet é particularmente adequado para a implementação de Risset's *transformações íntimas*, como alongamento de tempo ou transposição de frequência [40]

Em 1985, Risset e Grossmann iniciaram um grupo de interesse especial em torno do tópic o de wavelets que incluía os matemáticos Yves Meyer e Ingrid Daubechies, o geofísico Jean Morlet, o físico quântico Thierry Paul, os pesquisadores de processamento de sinais Daniel Ar fi b e Richard Kronland-Martinet e muitos outros . Dentro deste grupo, Kronland-Martinet, Grossmann e Morlet implementaram a transformação wavelet contínua no ambiente de software SYTER (em francês, *Sistema em tempo real*) para análise e re-síntese de áudio [41] Este software teve um impacto profundo na pesquisa musical de computador. No-

ação, o layout visual do escalograma wavelet **você 1 (x)** espelha a representação das notas musicais em uma partitura, com o tempo ao longo da horizontal e o tom ao longo do eixo vertical. Este layout é, portanto, particularmente intuitivo para uma pessoa com treinamento clássico músico.

4.2. Fase de recuperação

Sob suposições relativamente suaves sobre ψ , a transformada wavelet é um operador invertível, com uma forma estável e fechada inverso [39, Teorema 4.4]. O mesmo não pode ser dito do operador do escalograma wavelet **vocẽ 1**: porque a aplicação de módulo complexo pontual incorre em perda de informação de fase, recuperando o sinal x de seu escalograma **vocẽ 1** (x) está longe de ser trivial. No entanto, esta operação não linear de módulo complexo tem a vantagem de demodular oscilações periódicas localmente em ω e t . Assim, a função $\text{vocẽ 1}(x)$ pode ser considerada como uma coleção de valores reais. Como resultado, ele veicula uma função de valor complexo, que muda

$\psi / (t, \text{registro } \lambda).$

Em 2015, Irène Walspurger provou que, sob uma condição estrita, mas viável, em Ψ , o operador do escalograma é invertível até uma mudança de fase constante, que é inaudível [42]. Na verdade, apesar da perda de fase, há redundância suficiente entre as bandas de frequência adjacentes λ para codificar diferenças de fase. Embora a prova de Walspurger não forneça uma expressão de forma fechada

VOCÊ 1 É

Por falta de uma estrutura computacional adequada para manipular representações de escalogramas enquanto preserva a condição de reprodução do kernel, restam duas opções. A primeira opção, rotineiramente empregada no vocoder de fase e seus aprimoramentos adicionais [43 - 45], é confiar em uma heurística de coerência vertical para sintetizar uma função de fase substituta que é, se não livre de artefatos, pelo menos perceptualmente plausível. A segunda opção é transformar gradualmente x em um novo sinal cujo escalograma corresponde ao escalograma alvo. Em particular, pode-se construir uma trajetória de aproximações de gradiente descendente

Please cite this article in press as: V. Lostanlen et al., Fourier at the heart of computer music: From harmonic sounds to texture, C. R. Physique (2019), <https://doi.org/10.1016/j.crhy.2019.07.005>

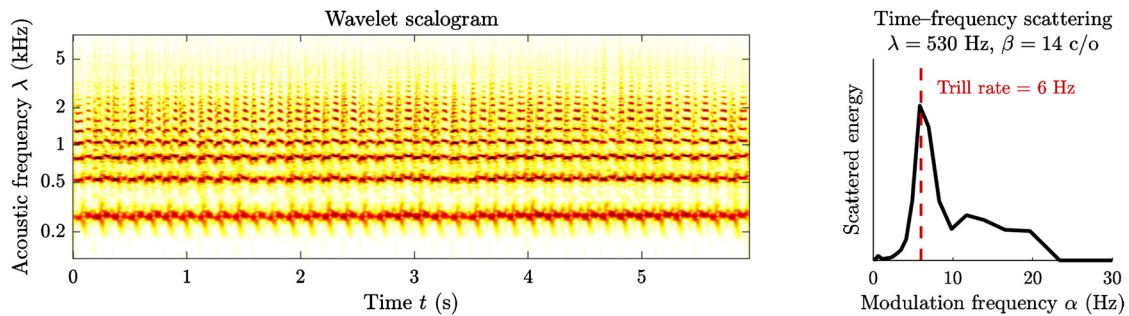


FIG. 3 Esquerda: escalograma wavelet de uma trombeta tocando um trinado com tom C4. Direita: perto do segundo harmônico ($\lambda = 520$ Hz, um pico de energia no tempo - coeficientes de espalhamento de frequência revela uma modulação espectral na taxa de trilo correspondente ($\alpha = 6$ Hz) e extensão do trilo ($\beta = 14$ canais / oitava).

Aplicando um algoritmo de descida gradiente no funcional distância euclidiana $\mathbf{E} \mathbf{x}: \mathbf{y} \rightarrow \|\mathbf{S}(\mathbf{y}) - \mathbf{S}(\mathbf{x})\|$, é possível atualizar \mathbf{y}_0 para coincidir com os coeficientes de espalhamento de frequência de tempo de \mathbf{x} . Este algoritmo foi originalmente proposto por Joan Bruna no caso mais simples de espalhamento no tempo, onde o operador **você 2** incluído na wavelet temporal ψ_α , mas sem wavelet frequencial

ψ_β [49]. Vários resultados teóricos de Waldspurger, notavelmente a invertibilidade de redes de espalhamento em profundidade infinita [42] e o decaimento exponencial dos coeficientes de espalhamento [50], sugere que o espalhamento de frequência de tempo tem a capacidade de com precisão

caracterizar dependências de longo alcance T do que o módulo de Fourier ou a média dos coeficientes do módulo wavelet. No caso de sons musicais, configuração T a 50 ms ou mais no cálculo dos coeficientes do espectrograma ou coeficientes do escalograma leva a artefatos perceptíveis na reconstrução. Em comparação, a configuração T até 500 ms produz uma reconstrução baseada em espalhamento

\mathbf{y} não que é perceptualmente semelhante a \mathbf{x} [51]

4.5. Compondo música com dispersão ondulada

O espalhamento de tempo - frequência foi originalmente desenvolvido como uma representação de sinal para classificação automática de sinais de áudio. Em 2016, porém, o compositor Florian Hecker propôs estender sua aplicação para além da mera análise de sons e redirecioná-la para uma aplicação criativa. De um fragmento de áudio \mathbf{x} de duração igual a 17 s, Hecker computou

uma forma de onda \mathbf{y}_0 de acordo com a Equação (11) combinando as amplitudes entre \mathbf{x}/e e \mathbf{y}_0 no domínio de Fourier, objetivo randomizar as fases correspondentes. Ele então realizou uma descida gradiente para sintetizar $\mathbf{y}_1, \dots, \mathbf{y}_{50}$ iterativamente, assim com aproximando-se de \mathbf{x} no sentido da distância Euclidiana associada funcional $\mathbf{E} \mathbf{x}$, com $Q = 12$ anos $T = 188$ ms. A peça resultante, chamada *FAVN*, foi estreada na Alte Oper em Frankfurt, apresentada novamente no festival Geometry of Now

em Moscou, e se tornou uma exposição de dois meses no Kunsthalle em Viena, com um catálogo retrospectivo dedicado [52]. Na hora do concerto, o intérprete de *FAVN* tem que reproduzir a primeira iteração do primeiro fragmento, e então avançar progressivamente na reprodução digital da peça, tanto em termos de tempo composicional (fragmentos) quanto tempo computacional (iterações). Desde então, Hecker compôs três peças originais com dispersão de frequência de tempo: *Modulador (Transformação de Dispersão)* (2012), *Palimpsesto Experimental* (2016), e *Inspeção* (2016).

Além da estética da música experimental, a dispersão da frequência do tempo encontrou recentemente um público mais amplo, aparecendo em um disco eletrônico chamado *A forma de RemiXXXes que virão*, lançado pela gravadora independente Warp [53]. Em seu remix, Hecker isolou alguns loops de uma barra de Lorenzo Senni *XAllegroX* e reconstruindo-os a partir de seus

tempo - coeficientes de espalhamento de frequência. Embora a estimativa inicial baseada em Fourier \mathbf{y}_0 soa nebuloso e estático, a reconstrução recupera parte do conteúdo rítmico original em iterações subsequentes, produzindo assim uma sensação de aumento sônico. No

No contexto da dance music, tal sensação transmite a antecipação de uma mudança repentina, ou "queda". Para produzir um aumento, uma técnica de engenharia de áudio amplamente difundida consiste em isolar uma amostra percussiva e reproduzi-la repetidamente enquanto reduz progressivamente a duração dessa amostra. Uma segunda técnica consiste em aplicar um filtro passa-altas cuja frequência de corte aumenta progressivamente com o tempo. Essas duas técnicas operam no domínio do tempo e no domínio de Fourier, respectivamente. Em contraste, o espalhamento de frequência de tempo atinge uma sensação de aumento ao afetar a organização espectral-temporal do som, de grosso a fino, independentemente das manipulações puras no domínio do tempo ou da frequência. O lançamento do

XAllegroX (scattering.m remix) subverte a articulação clássica de tensão e liberação na música de dança, deixando a percepção da métrica musical e a percepção do registro musical inalteradas e, em vez disso, modulando a complexidade espectral-temporal da própria textura.

5. Conclusão

A democratização do armazenamento de áudio digital na década de 1980, seguida pela adoção massiva da comunicação pela Internet na década de 1990, intensificou a necessidade de modelos computacionais eficientes de percepção auditiva. Da codificação compressiva ao aprimoramento da fala, muitas das tecnologias que permitem a rede de áudio sem fio contam com a transformação rápida de Fourier ou FFT - saudada pelo jornal *IEEE Computing in Science & Engineering* como um dos dez algoritmos mais importantes do século XX. Assim, a onipresença dos algoritmos de processamento digital de sinais nas sociedades industrializadas é a prova de que o legado científico

de Fourier permanece, no 250º aniversário de seu nascimento, altamente relevante para enfrentar os desafios matemáticos da era da informação.

Surpreendentemente, a publicação do *Teoria Analítica do Calor* (1822) por Joseph Fourier precede a invenção da dig-computadores italianos por mais de cem anos. Apesar desse intervalo de tempo, a teoria das séries trigonométricas aqui descrita é fundamental na história do processamento digital de sinais. O caso da acústica musical exemplifica a longevidade das representações de Fourier em muitos domínios da ciência aplicada. Na verdade, a introdução da série de Fourier constitui um divisor de águas em nossa compreensão do timbre - isto é, como a forma e a técnica de tocar de um instrumento influencia nossa percepção qualitativa do som que ele produz, além do tom e da intensidade. Praticamente falando, o principal apelo por trás das decomposições de Fourier está na imitação de instrumentos acústicos com um pequeno conjunto de parâmetros contínuos fisicamente interpretáveis.

Em seu tratado, Fourier apontou que seu método proposto para resolver a equação do calor em uma haste de metal também poderia ser aplicado, *mutatis mutandis*, para resolver a equação de d'Alembert em uma corda vibrante. Embora a aplicação inicial desta estrutura fosse meramente analítica, o advento da música computacional na década de 1960 introduziu o uso prático da série de Fourier em ambas as fases da modelagem de som: análise (ou seja, visualização e classificação de sons pré-gravados) e síntese (ou seja, transformando pré-sons gravados e geração de novos sons). Além disso, os esforços coordenados de pesquisadores e compositores demonstraram que as séries de Fourier também podem gerar timbres inéditos e, assim, servir como uma interface ergonômica para a criação musical contemporânea.

Na produção musical, uma ampla gama de efeitos de áudio digital (DAFX) depende de alguma forma de análise espectral-temporal e ressíntese após manipulação no domínio do tempo - frequência. Um exemplo bem conhecido desse procedimento é o vocoder de fase, no coração do algoritmo "Auto-Tune" para ajuste de tom de voz. A disponibilidade desses DAFX não só inspirou compositores com novas ideias, mas também criou novas necessidades tecnológicas. Em particular, embora o atual estado da arte da análise baseada em Fourier - a síntese agora reproduza fielmente o instrumentário comum, é inadequado para replicar sons que não são considerados fala nem música, por exemplo, sons domésticos ou vocalizações de vida selvagem. De forma geral, a questão da modelagem de texturas de áudio na ausência de qualquer conhecimento prévio de periodicidade temporal permanece amplamente em aberto. Longe de ser puramente acadêmica, essa questão representa um gargalo técnico no caminho para interações computador-humano ainda mais intuitivas.

Na última seção deste artigo, apresentamos o espalhamento tempo-frequência, uma representação não linear de sinais de áudio que, ao lado de outros modelos recentes baseados em modulações espectral-temporais, oferece novas possibilidades para a síntese de textura. Seguindo a tradição metodológica das primeiras pesquisas em música computacional, em que a arte e a ciência desempenham papéis complementares, o desenvolvimento da dispersão de frequência no tempo foi associado à criação de novas peças, tanto na estética dita "vanguardista" quanto na estética "pop". O espalhamento de frequência de tempo é fortemente inspirado pela análise de Fourier porque se baseia em convoluções com wavelets de Morlet, que são bem localizadas no domínio do tempo e no domínio de Fourier. No entanto, o espalhamento de frequência de tempo também toma emprestado de outros paradigmas de concepção mais recente, e notavelmente redes convolucionais profundas.

Agradecimentos

Este trabalho é financiado pelo ERC InvariantClass grant 320959. O Flatiron Institute é uma divisão da Simons Foundation.

Referências

- [1] MV Mathews, The digital computer as a musical instrument, *Science* 142 (3592) (1963) 553–557.
- [2] C. Shannon, Comunicação na presença de ruído, *Proc. Inst. Radio Eng.* 37 (1) (1949) 10–21.
- [3] U. Zölzer, *DAFX: Digital Audio Effects*, John Wiley & Sons, 2011.
- [4] F. Jędrzejewski, *Mathematics of acoustic systems*, L'Harmattan, Paris, 2002.
- [5] M. Andreatta, F. Nicolas, C. Alunni, À luz da matemática e à sombra da filosofia, em: *Proceedings of the Mamuphi seminar, Mathematics– Music - Philosophy*, Ircam / Delatour, França, 2012.
- [6] M. Kac, pode-se ouvir a forma de um tambor? *Am. Math. Minhas.* 73 (4, Parte 2) (1966) 1–23.
- [7] M. Mersenne, *Universal Harmony*, contendo a teoria e prática da música, Bibliothèque nationale de France, 1636, No. FRBNF30932210.
- [8] J. Fourier, *Analytical Theory of Heat*, National Library of France, 1822, No. FRBNF30454516.
- [9] RM Friedman, A criação de uma nova ciência: a teoria analítica do calor de Joseph Fourier, *Hist. Viga. Phys. Sci.* 8 (1977) 73–99.
- [10] A. Herberman, The Inauguration of Trigonometric Series in Fourier's Analytical Theory of Heat and in the Vibrating String Controversy, *Rev. History Math., Soc. Matemática. Fr.* 19 (2) (2013) 151–243.
- [11] MV Mathews, et al., *The Technology of Computer Music*. Voar. 969, MIT Press, Cambridge, MA, USA, 1969.
- [12] NH Fletcher, TD Rossing, *The Physics of Musical Instruments*, Springer-Verlag, 1991.
- [13] H. Triebel, *Theory of Function Spaces*, Birkhäuser Verlag, 1992.
- [14] S. McAdams, et al., Escala perceptiva de timbres musicais sintetizados: dimensões comuns, cidades específicas e classes de assuntos latentes, *Psychol. Res.* 58 (3) (1995) 177–192.
- [15] B. Bogert, M. Healy, J. Tukey, The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariância, cross-cepstrum e quebra da safra, em: *Proceedings of a Symposium on Time Series Analysis*, John Wiley & Sons, Inc., 1963, pp. 209–243.
- [16] P. Schaeffer, *Treatise on Musical Objects: An Essay Across Disciplines*, University of California Press, Berkeley, EUA, 2017.
- [17] J. Harvey, *The Music of Stockhausen: An Introduction*, University of California Press, Berkeley, CA, EUA, 1975.
- [18] P. Schaeffer, *Em Busca de uma Música Concreta*, Vol. 15, University of California Press, Berkeley, CA, EUA, 2012.

- [19] JM Chowning, A síntese de espectros de áudio complexos por meio de modulação de frequência, *J. Audio Eng. Soc.* 21 (7) (1973) 526–534.
- [20] H. Berjamine, et al., Modelagem numérica no domínio do tempo de instrumentos de latão, incluindo propagação de onda não linear, perdas viscoelásticas e vibração labial, *Acta Acust. United Ac.* 103 (1) (2017) 117–131.
- [21] M. Castellengo, Ouvindo música e acústica, Eyrolles, Paris, 2015.
- [22] X. Rodet, P. Depalle, Spectral envelopes and inverse FFT synth, in: *Proc. 11th Int. AES Conference: Audio Test & Measurement*, Portland, OR, USA, 29-31 de maio de 1992.
- [23] W. Koenig, H. Dunn, L. Lacy, O espectrógrafo de som, *J. Acoust. Soc. Am.* 18 (1) (1946) 19–49.
- [24] J.-C. Risset, Síntese de sons usando computadores eletrônicos aplicados ao estudo de sons de trombeta, *CR Acad. Sci. Paris, Ser. B* 263 (1966) 111-114.
- [25] MR Schroeder, Paradoxo auditivo baseado na forma de onda fractal, *J. Acoust. Soc. Am.* 79 (1) (1986) 186–189.
- [26] H. Dudley, O vocoder - recriação elétrica da fala, *J. Soc. Motion Pict. Eng.* 34 (3) (1940) 272–278.
- [27] JL Flanagan, et al., Phase vocoder, *J. Acoust. Soc. Am.* 38 (5) (1965) 939–940.
- [28] J.-C. Risset, Do sonho ao som. Entrevistas com Matthieu Guillot, L'Harmattan, Paris, 2008.
- [29] J. Laroche, M. Dolson, Improved vocoder time - scale modi? Cation of audio, *IEEE Trans. Processo de áudio de fala.* 7 (3) (1999) 323–332.
- [30] M. Betser, et al., Estimation of frequency for AM / FM models using the phase vocoder framework, *IEEE Trans. Processo de sinal.* 56 (2) (2008) 505–517.
- [31] M. Liuni, et al., Adaptação automática do tempo - resolução de frequência para análise e re-síntese de som, *IEEE Trans. Audio Speech Lang. Processar.* 21 (5) (2013) 959–970.
- [32] A. Cont, ANTESCOFO: sincronização antecipatória e controle de parâmetros interativos em música de computador, em: *Proc. 2008 International Computer Music Conference, ICMC 2008*, Belfast, Irlanda, 24–29 de agosto de 2008, Michigan Publishing, 2008, pp. 33–40.
- [33] T. Wishart, A composição de *Vox-5*, *Comput. Music J.* 12 (4) (1988) 21–27.
- [34] R. Strachan, *Sonic Technologies: Popular Music, Digital Culture and the Creative Process*, Bloomsbury Publishing, EUA, 2017.
- [35] X. Serra, JO Smith, Síntese de modelagem espectral: um sistema de análise / síntese de som baseado em uma decomposição determinística e estocástica, *Comput. Music J.* 14 (4) (1990) 12–24.
- [36] K. Filtz, L. Haken, P. Christensen, Um novo algoritmo para associação de largura de banda em modelagem de som aditivo de largura de banda aprimorada, em: *Proc. Int. Comput. Music Conf. (ICMC 2000)*, Berlim, 27 de agosto - 1 de setembro de 2000, pp. 384-387.
- [37] JH McDermott, EP Simoncelli, percepção da textura do som via estatísticas da periferia auditiva: evidências da síntese de som, *Neuron* 71 (5) (2011) 926–940.
- [38] W.-H. Liao, A. Roebel, AW-Y. Su, Sobre a modelagem de texturas sonoras com base na representação STFT, in: *Proc. 16th Int. Conf. on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2–4 de setembro de 2013.
- [39] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 2009.
- [40] ES Ottosen, et al., A phase vocoder based on nonstationary Gabor frames, *IEEE Trans. Audio Speech Lang. Processar.* 25 (11) (2017) 2199–2208.
- [41] R. Kronland-Martinet, J. Morlet, A. Grossmann, Análise de padrões de som através de transformadas wavelet, *Int. J. Pattern Recognit. Artif. Intell.* 1 (2) (1987) 273–302.
- [42] I. Waldspurger, *Wavelet Transform Modulus: Phase Retrieval and Scattering*, tese de doutorado, École normale supérieure, Paris, 2015.
- [43] A. Röbel, A new approach to transient processing in the phase vocoder, in: *Proc. 6th Int. Conf. on Digital Audio Effects (DAFx-03)*, Londres, Reino Unido, 8–11 de setembro de 2003.
- [44] A. Röbel, Um vocoder de fase invariante de forma para transformação de fala, em: *10th Int. Conf. on Digital Audio Effects (DAFx-10)*, Graz, Áustria, 6–10 de setembro de 2010.
- [45] J. Laroche, M. Dolson, Novas técnicas de vocoder de fase para mudança de tom em tempo real, coro, harmonização e outras modi fi cações de áudio exóticas, *J. Audio Eng. Soc.* 47 (11) (1999) 928–936.
- [46] S. Mallat, Compreendendo redes convolucionais profundas, *Philos. Trans. R. Soc. A* 374 (2065) (2016) 20150203.
- [47] D. Depireux, et al., Spectro-temporal response field characterization with dynamic ripples in furet primary auditory cortex, *J. Neurophysiol.* 85 (3) (2001) 1220–1234.
- [48] J. Andén, V. Lostanlen, S. Mallat, Joint time - frequency scattering, *IEEE Trans. Processo de sinal.* 67 (14) (2019) 3704–3718.
- [49] J. Bruna, S. Mallat, síntese de textura de áudio com momentos de espalhamento, pré-impressão arXiv, arXiv: 1311.0407, 2013.
- [50] I. Waldspurger, Exponential decay of scattering coefficients, in: *Proc. 2017 Int. Conf. on Sampling Theory and Applications (SampTA)*, Tallinn, Estônia, 3–7 de julho de 2017, pp. 143–146.
- [51] J. Andén, V. Lostanlen, S. Mallat, Joint time-frequency scattering for audio classification, in: *IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 17–20 de setembro de 2015, Boston, EUA, 2015, 6 pp.
- [52] V. Lostanlen, On time - frequency scattering and computer music, em: *VJM Nicolaus Schafhausen (Ed.), Florian Hecker: Halluzination, Perspektive, Synthese*, Sternberg Press, Berlin, 2019.
- [53] V. Lostanlen, F. Hecker, The shape of RemiXXes to come: síntese de textura de áudio com espalhamento de frequência de tempo, em: *Proc. 22nd Int. Conf. on Digital Audio Effects (DAFx-19)*, Birmingham, Reino Unido, 2 a 6 de setembro de 2019.