SIMULAÇÃO DE INSTRUMENTOS MUSICAIS ACÚSTICOS EM TEMPO REAL UTILIZANDO REDES NEURAIS PARALELAS NO DOMÍNIO DA FREQUÊNCIA

Carlos Henrique Tarjano Santos (Universidade Federal Fluminense)
tesserato@hotmail.com
Valdecy Pereira (Universidade Federal Fluminense)
valdecypereira@yahoo.com.br

Para exibir características sonoras verossímeis, instrumentos musicais digitais que buscam emular instrumentos acústicos ainda utilizam-se majoritariamente da reprodução de sons pré gravados. Enquanto a capacidade computacional atual da maiioria dos comput

Palavras-chave: Redes Neurais, Desenvolvimento de Produtos, Instrumentos Musicais Virtuais, Síntese Sonora em Tempo Real

1. Introdução

Enquanto o interesse por instrumentos musicais digitais cresceu significativamente na última década (Staudt, 2016), instrumentos de ponta, utilizados em estúdios para gravações profissionais ainda baseiam-se prioritariamente em coleções de amostras sonoras pré-gravadas (Smith, 2008), demandando uma grande quantidade de armazenamento digital e poder computacional do hardware (geralmente compradores pessoais) onde estão implementados.

Para plataformas onde o armazenamento ou poder de processamento são fatores limitantes, a exemplo de teclados digitais e baterias eletrônicas, é inevitável o uso de bibliotecas de qualidade e tamanho reduzidos, que, a aliados a algoritmos menos sofisticados, comprometem a qualidade sonora do instrumento: qualidade profissional do pode ser alcançados conectando esses instrumentos a compradores, onde passam a funcionar como meus controladores que permitem o acesso a bibliotecas e algoritmos mais sofisticados.

Os desenvolvimentos recentes no campo de redes neurais sugerem seu potencial em mitigar algumas dessas limitações: na área de visão computacional, por exemplo, pode-se observar uma variedadede trabalhos que regularmente expandem as fronteiras da área.

A maior parte do trabalho sobre redes neurais aplicadas ao áudio aborda o problema a partir de um nível mais alto de abstração, que exclui a representação discreta de ondas sonoras: são usualmente sobre a manipulação de representações musicais como partituras.

A alta dimensionalidade usual da representação de ondas sonoras é uma das principais razões dessa abordagem, já que, por exemplo, no caso de áudio na qualidade padrão CD, a síntese de 10 segundos de áudio envolve a criação de quase 1 milhão de amostras.

O trabalho desenvolvido pelas equipes por trás do Google Brain e DeepMind (Engel et al., 2017) é uma notável exceção: trata-se de uma arquitetura neural baseada na rede Wavenet (Van Den Oord et al., 2016) que é usada para gerar sons diretamente após o treinamento a partir de amostras de áudio de vários instrumentos musicais. Os resultados mostram que uma arquitetura convolucional multicamadas é capaz de aprender representações no domínio do tempo para vários tipos de instrumentos. Uma extensão experimental deste trabalho, denominada Magenta (Eck, 2018), explora, do ponto de vista probabilístico, as representações latentes para sequências sonoras no domínio do tempo(Roberts, Engel, Oore e Eck., 2018; Roberts, Engel, Raffel, Hawthorne e Eck., 2018).

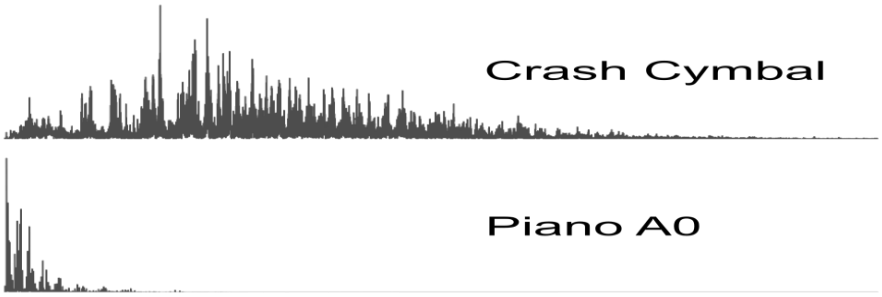
Nota-se, no entanto, um descolamento das áreas mais tradicionais de processamento de sinais: Podemos, por exemplo, tirar vantagem do caráter periódico das amostras e representá-las no domínio da frequência. A transformada Discreta de Fourier fornece uma representação perfeita e reversível de uma onda no domínio da frequência. Considerando o fato de que estamos, no presente trabalho, interessados somente em representações temporais no domínio real, as representações do domínio da frequência consistirão em um vetor de números complexos, metade do tamanho do número original de amostras.

Embora essa transformação de domínio não introduza por si só uma representação de onda mais compacta, do ponto de vista de armazenamento, já que números complexos são representados por pares de números reais na maioria das linguagens de programação, ela introduz vantagens conceituais consideráveis. Levando em conta, por exemplo, que o ouvido humano não é capaz de perceber frequências fora da faixa aproximada de 20 Hz a 20 kHz, uma das vantagens de trabalhar no domínio da frequência é que podemos truncar o resultado da transformada neste esse intervalo (tomando cuidado para traduzi-lo em termos das frequências locais da transformação).

Outra vantagem vem na forma de sua independência da duração do sinal, que permite o uso de uma arquitetura neural densa na previsão de ondas com durações arbitrárias. Outras vantagens desta abordagem serão ilustradas nas próximas seções, levando em conta as características físicas do instrumento a ser emulado e as propriedades da transformação, e fornecerão base teórica para o método aqui apresentado.

Até onde sabemos, as representações do som no domínio da frequência, uma técnica popular no campo do processamento digital de sinais, raramente eram usadas no contexto da síntese sonora a partir de redes neurais. (Embora seja comum em tarefas de classificação nesta área). A figura 1 compara as duas representações no domínio da frequência, para o caso de sons harmônicos e não harmônicos.

Figura 1 - Comparação do espectro audível, no domínio da frequência, entre o som produzido por um prato (crash) de uma bateria e a primeira tecla, A0, de um piano. O som produzido pelo prato não é harmônico, enquanto o som da tecla do piano é.



Fonte: Elaboração própria

Está claro que a mudança de domínio, que pode ser alcançada através do algoritmo eficiente Fast Fourier Transform (FFT), simplifica muito a representação dos sons harmônicos. O entendimento da física específica do instrumento pode simplificar ainda mais essas representações, aliviando a carga de previsão das redes neurais.

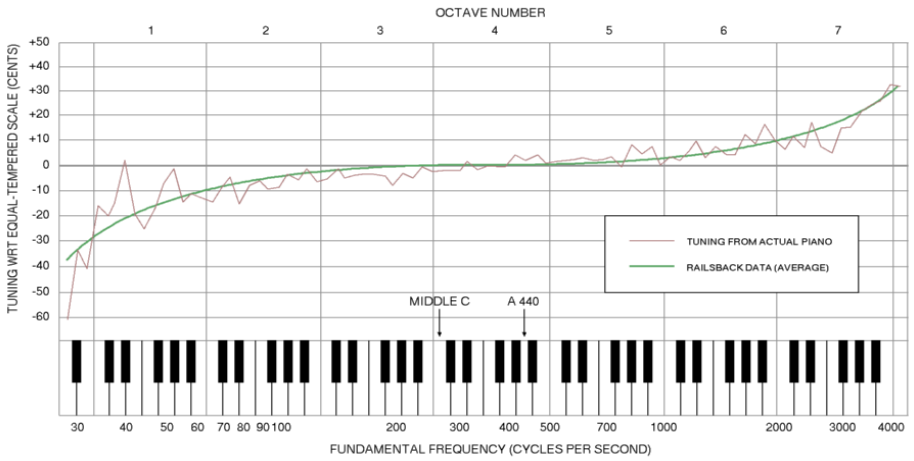
Considerando o caso específico de um piano de cauda padrão, podemos chegar a um modelo básico, a ser estendido posteriormente, que nos dá uma aproximação inicial razoável.

Observando que em instrumentos de igual temperamento, como é o caso do piano, a proporção de frequências (teóricas) em notas adjacentes é fixada em $2^{1/12}$, podemos chegar a uma fórmula da forma $f_0[k] = 440\ 2^{\frac{k-49}{12}}$ relacionando as 88 teclas do piano às suas frequências fundamentais, onde k representa o número da tecla do piano, de 1 a 88, e 440 Hz é a frequência padrão da tecla 49, com o tom A4.

Na prática, porém, as frequências fundamentais desviam do seu valor teórico, devido a uma técnica de afinação denominada alongamento da oitava que achata as oitavas mais baixas e alonga as mais altas, no que diz respeito às suas frequências teóricas fundamentais, na tentativa de atenuar o embate entre parciais de diferentes teclas (Koenig e Fandrich, 2015).

Esse comportamento consistente da parte de afinadores de piano, decorrente da metodologia utilizada na afinação do piano, foi investigado pela primeira vez por Railsback em um artigo publicado em 1938 no *The Journal of the Acoustical Society of America*. Os dados podem ser vistos na figura 2, com a linha verde suave representando a média dos desvios para vários pianos.

Figura 2 - A curva de Railsback, exibindo os desvios das frequências fundamentais teóricas médias em cada uma das teclas de um piano



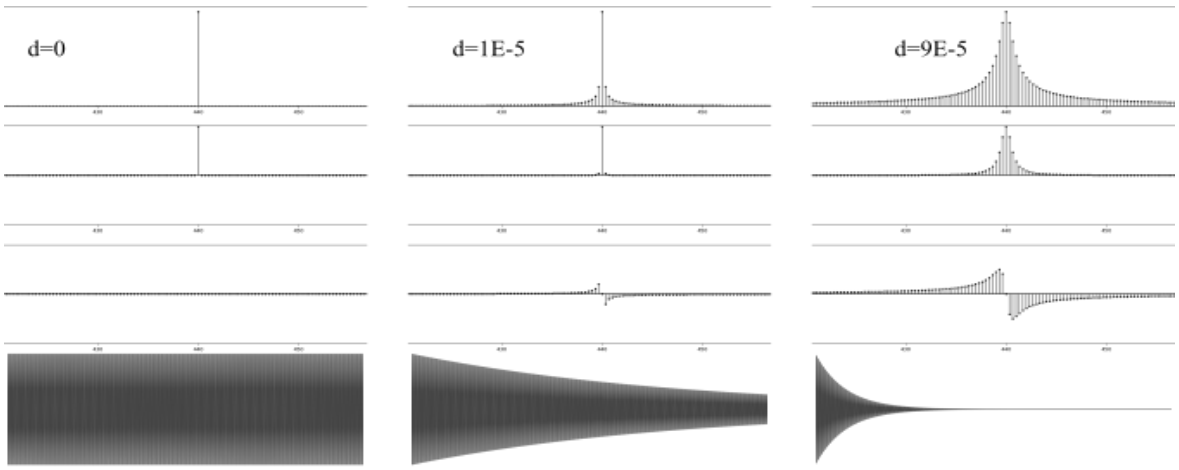
Fonte: Tung, 2006

Isso pode ser incorporado ao modelo antes do tratamento da rede neural. Para explicar as frequências parciais, uma estratégia simples consiste em assumir o caso de sequência ideal, em que parciais são múltiplos inteiros da frequência fundamental da tecla; podemos assim escrever $f[p,k] = 440.2^{\frac{k-49}{12}}(p+1)$ para denotar a freqüência da p-ésima parcial da k-ésima tecla.

Enquanto Fletcher (1964), por exemplo, propõe uma equação que relaciona a frequência fundamental de uma nota de piano com sua p-ésima parcial, incorporando a inarmonicidade presente nas cordas de piano, podemos ver em Koenig e Fandrich (2014) que os coeficientes de inarmonicidade por tecla não são, em geral, bem comportados. Esse também é o caso das amplitudes parciais; as redes neurais estão, portanto, em melhor posição para extrair características ocultas e aprender as associações subjacentes necessárias para reproduzir e generalizar essas informações.

É importante notar que a base de picos no domínio da frequência é proporcional ao decaimento da frequência correspondente no domínio do tempo: a figura 3 ilustra essa relação.

Figura 3 –Relação entre a variação das representações no domínio do tempo e da frequência de uma senoide pura de 440 Hz com decaimento exponencial



Fonte: Elaboração própria

Pode-se observar que o decaimento introduz novas frequências em torno da frequência nominal, além de mudanças de fase na representação no domínio da frequência; observa-se empiricamente que o efeito primordial dessas frequências e fases é reproduzir o decaimento (ou, mais genericamente, o envelope) da onda.

A partir dessa observação, duas importantes intuições podem ser traçadas: A primeira é que, com um razoável grau de aproximação, podemos descrever um som harmônico gerado por uma excitação impulsiva em função da localização de algumas de suas frequências (parciais), suas respectivas intensidades e seus decaimentos.

Como envelopes parciais devem ser contabilizados diretamente, informações originais de fase podem ser descartadas sem efeitos perceptuais significativos, como é sugerido empiricamente a partir da reconstrução de ondas com suas fases originais zeradas ou randomizadas.

No repositório Github dedicado a este trabalho (Tarjano, 2018), em uma pasta chamada "RandomPhaseReconstructions/", podem ser encontradas reconstruções das amostras usadas para treinar a rede, onde as informações de fase foram randomizadas para as primeiras 100 frequências parciais.

As amostras originais podem ser encontradas no mesmo repositório, em "00_samples / piano /".Comparando o som, pode-se ver que a reconstrução é bastante plausível. A maior parte da diferença perceptual entre eles se origina no número de parciais considerados, que não inclui todas as frequências presentes na fase transitória da onda nas teclas mais baixas.

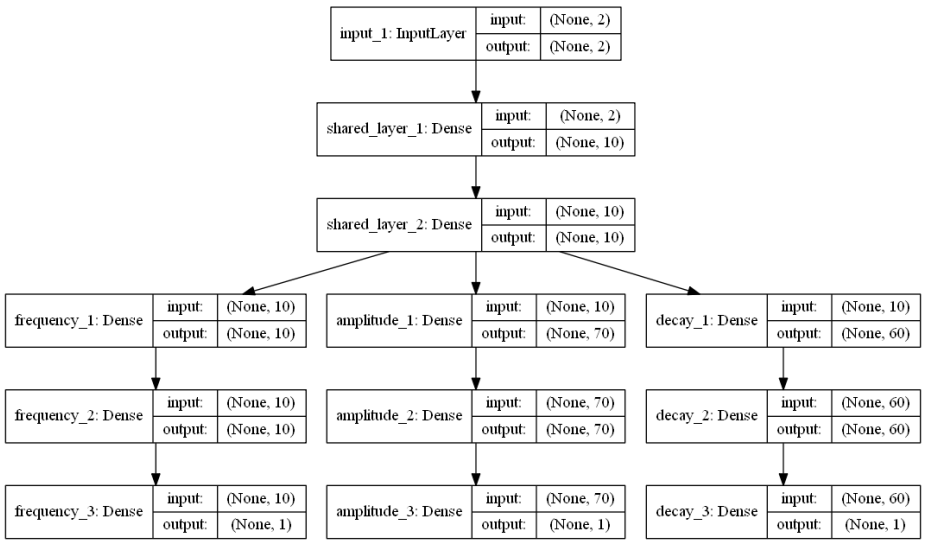
Notando que os instrumentos harmônicos têm uma distribuição de frequência bem comportada, consistindo basicamente de picos em sua representação no domínio da

frequência, e assumindo decaimentos exponenciais da forma e^{-dt}, com um valor de d por parcial e por tecla, pode-se elegantemente explicar o envelope de cada uma das parciais com conhecimento da amplitude da parcial e a taxa de decaimento d.

2. Metodologia

Com base na discussão apresentada na introdução, foi criada uma rede neural, usando a biblioteca Keras em cima do backend Tensorflow. A arquitetura pode ser vista na figura 4. Recebendo como entrada uma matriz da forma [k, p] onde k é a tecla piano normalizada, na faixa de 1 a 88, e p é a frequência normalizada de cada um das 100 primeiras parciais, a rede foi treinada para produzir, paralelamente, a inarmonicidade residual, o decaimento e a amplitude de cada par tecla-parciais.

Figura 4–Arquitetura da rede utilizada



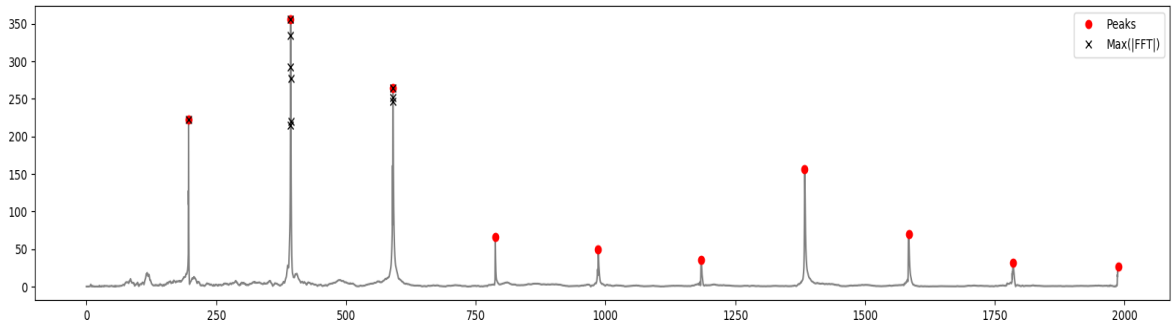
Fonte: Elaboração própria

As amostras de som usadas para treinar o modelo foram obtidas no site da Universidade de Iowa Electronic Music Studios (Fritts, 2011). A biblioteca original consiste de um total de 260 amostras gravadas de um Steinway & Sons modelo B Grand Piano com um microfone Neumann KM 84, e são codificadas em arquivos .aiff estéreo de 16 bits, 44,1 kHz.

A partir desta biblioteca de amostras, utilizamos as 88 articulações *fortissimo* (uma por tecla). Os arquivos .aiff originais foram convertidos em arquivos mono .wav com a mesma taxa de bits e taxa de amostragem original. Os silêncios no início dos arquivos foram removidos e a intensidade do áudio foi normalizada. As amostras preparadas estão disponíveis no repositório do Github preparado para este trabalho.

Tendo em conta que, conforme discutido, as frequências parciais são aproximadamente múltiplos inteiros da frequência fundamental de cada tecla, pode-se procurar os máximos no domínio da frequência considerando os intervalos apropriados. A Figura 5 compara esse algoritmo com a enumeração ingênua dos valores de intensidade mais altos, para uma onda correspondente ao som emitido pela tecla 35 de um piano, a partir do qual as primeiras 30 frequências parciais são investigadas.

Figura 5–Frequências parciais detectadas (pontos vermelhos) em comparação com a enumeração dos maiores valores ("x"
pretos)



Fonte: Elaboração própria

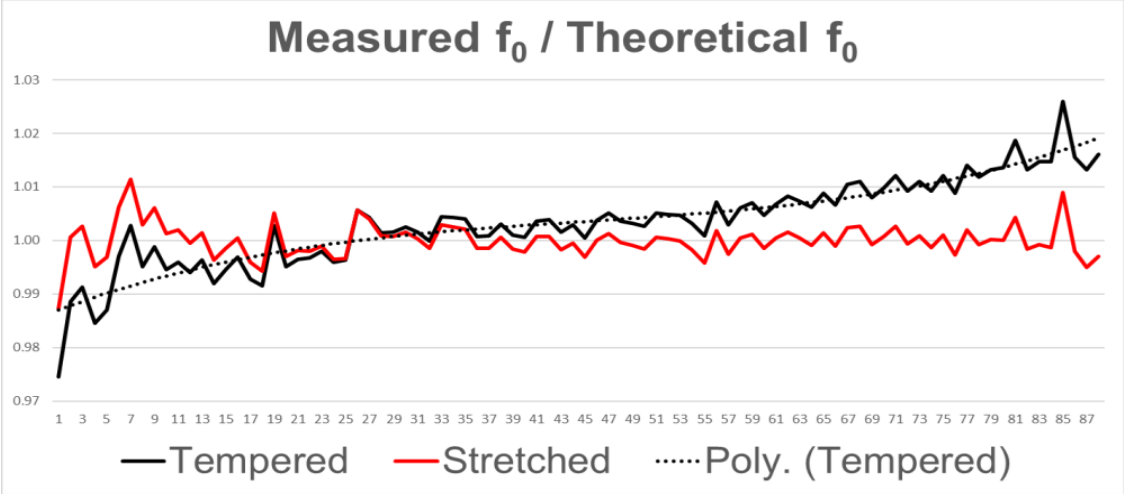
A Figura 6 oferece uma visão da afinação original do piano a partir da qual as amostras foram gravadas; um polinômio de grau 3 oferece um bom compromisso entre simplicidade e precisão na aproximação do ajuste esticado e foi ajustado usando a biblioteca numérica *Numpy*.

Podemos melhorar as frequências teóricas fundamentais $f_0(k) = 440.2^{\frac{k-49}{12}}$ com um termo que representa o alongamento da oitava; como se viu, essa equação original desconsidera as inarmonicidades presentes nos instrumentos, responsáveis por importantes características de seus timbres.

No entanto, apresenta uma aproximação inicial bastante razoável que serve tanto para reforçar as características harmônicas básicas no modelo final quanto para aliviar o esforço de predição da rede, na medida em que podemos adicionar um termo de inarmonicidade na equação acima a ser aprendida pela rede.

Podemos então escrever $f_0[k]$ como $440.2^{\frac{k-49}{12}}(c_1k^3 + c_2k^2 + c_3k)$. O esforço é justificado pelo fato de que melhorias em f_0 são levadas a todas as frequências teóricas parciais, já que são múltiplos de f_0, simplificando muito a arquitetura neural necessária no modelo. O algoritmo utilizado para extrair as frequências fundamentais, comparar com as teóricas e proceder ao processo de adaptação pode ser encontrado no arquivo "00_tuning_stretch.py", e a figura 6 mostra o polinômio, juntamente com uma comparação entre os desvios das afinações temperada e esticada as frequências fundamentais observadas por tecla.

Figura 6–Comparação entre a afinação temperada, teórica, e a afinação alongada utilizada na prática



Fonte: Elaboração própria

Assim, uma frequência teórica arbitrária, em função da tecla do piano e da parcial considerada, será representada, no modelo, da seguinte forma:

$$f_0[k,p] = 440.2^{\frac{k-49}{12}}(c_1k^3 + c_2k^2 + c_3k)(p+1)i[k,p], \quad k \in \{1,2,\dots,88\}, p \in \{0,1,2,\dots\},$$

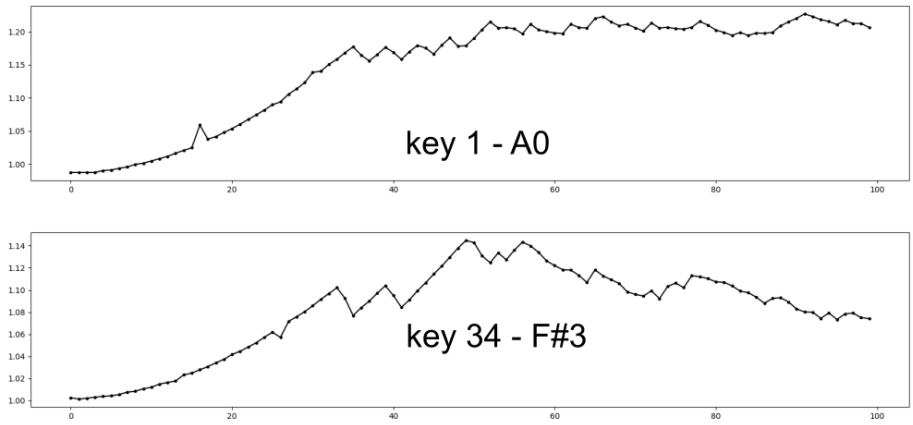
onde $i[k,p]$ é a inarmonicidade dada pela rede.

Considerando-se tanto as frequências fundamentais quanto as parciais no espectro audível, o modelo teria que aprender um intervalo de frequências de 27 Hz, a partir da frequência de fundamental de A0, para pouco mais de 20 kHz, correspondendo, por exemplo, ao quinto parcial de C8, caso fossemos estimar diretamente as frequências parciais de um piano.

Trabalhar com inarmonicidades, por outro lado, reduz o intervalo para o limite entre 0,98 e 1,02. Além disso, o comportamento das desarmonias é razoavelmente previsível, com um caráter ligeiramente exponencial, como ilustrado na figura 7.

Apesar de ser formulado e usado neste trabalho no contexto de um piano, este framework é conveniente para uma ampla gama de instrumentos, uma vez que as 88 teclas de um piano variam de A0 a C8, cobrindo o espectro de frequências da maioria dos instrumentos de interesse; Particularmente, para treinar a implementação aqui apresentada a partir de qualquer instrumento, basta etiquetar as amostras dos sons relevantes com o número da tecla equivalente de um piano.

Figura 7-Inarmonicidades



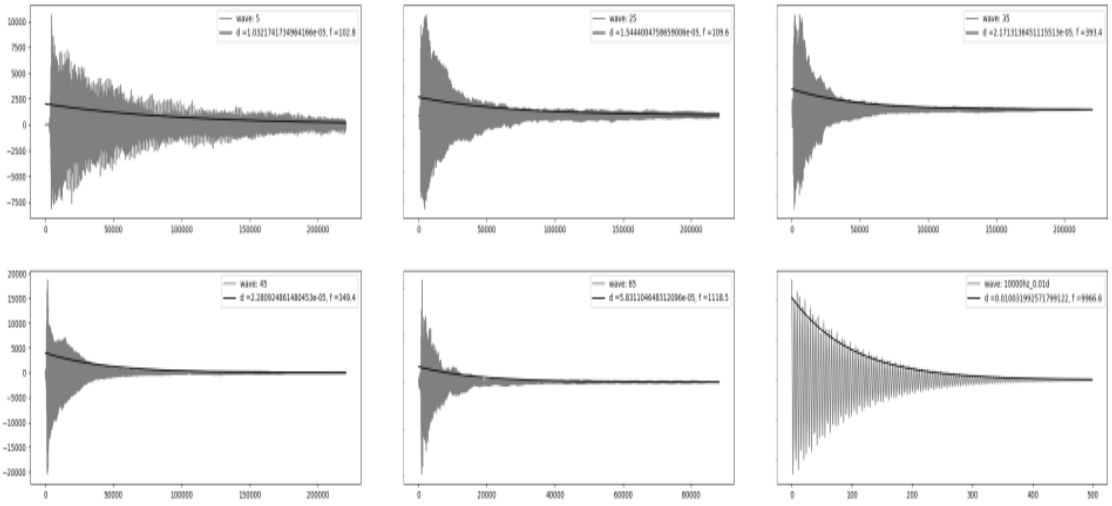
medidas

Fonte: Elaboração própria

Este mesmo raciocínio é aplicado às amplitudes, que é previsto como uma fração dos valores máximos encontrados em cada uma das teclas (ou notas, mais geralmente, no caso de um instrumento arbitrário sendo treinado), e residindo no intervalo fechado entre 0 e 1.

A curva de decaimento pode ser estimada considerando a diferença de intensidade de uma frequência arbitrária na primeira e segunda metades da onda, usando a fórmula $d = \frac{2\ln\left(\frac{a_1}{a_2}\right)}{l}$. A Figura 8 ilustra a aplicação desse procedimento em várias amostras de som. Note que somente no último caso, a onda sendo uma senoide pura, o decaimento extraído corresponde ao envelope da amostra.

Fig. 8. Estimativas de decaimento da frequência predominante



Fonte: Elaboração própria

Como foi observado, com uma estimativa dos decaimentos das frequências parciais, a informação de fase tem impacto insignificante sobre o som da onda reconstruída; Escolhemos, portando, na presente implementação, aleatorizar as fases: Esta abordagem tem a vantagem de transmitir um caráter mais orgânico e variado à saída sintetizada do modelo final, uma vez que nenhuma onda gerada será exatamente igual a qualquer outra onda. Outra opção, menos rica do ponto de vista da percepção, seria atribuir às fases um valor arbitrário (zero, por exemplo).

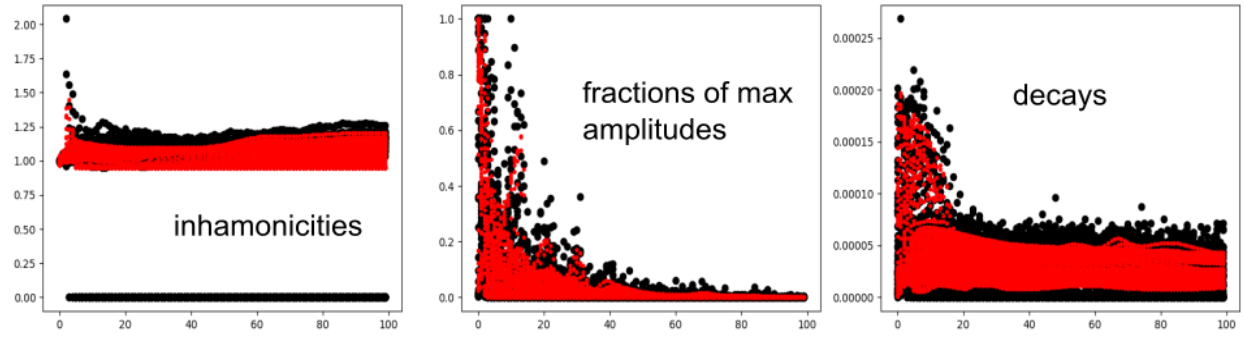
Como função de ativação em cada uma das camadas da rede, foi utilizada uma versão modificada da tangente hiperbólica, na forma $a(x) = \frac{\tanh(6x-3)}{2} + \frac{1}{2}$, para melhor cobrir o intervalo [0,1] em que as entradas e saídas residem.

A rede neural usada no atual modelo tem um total de 10.563 parâmetros treináveis, com um tamanho em disco de aproximadamente 200 KB. As duas camadas iniciais, com 10 neurônios cada, são comuns a todos os 3 resultados previstos e fornecem uma representação inicial compartilhada, a ser estendida mais tarde por cada uma das ramificações responsáveis pelas quantidades individuais.

Essa abordagem fornece menos redundância no modelo, permitindo que o número de neurônios nas camadas independentes seja personalizado para levar em conta a otimização individual de cada uma das saídas; Vale a pena notar que as aproximações de frequência

fundamental feitas a priori permitiram um pequeno número de neurônios no ramo responsáveis pelas inarmonicidades. A Figura 9 compara os alvos originais e o comportamento aprendido pela rede.

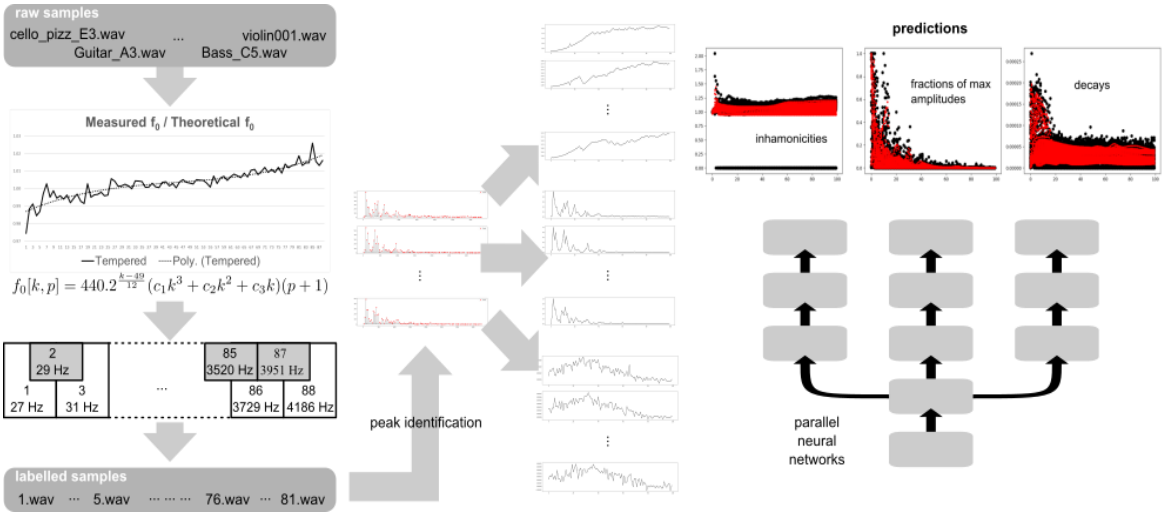
Figura 9 – Predições



Fonte: Elaboração própria

Assim, podemos definir uma nova metodologia para modelagem de instrumentos harmônicos, apelidada de "Neuro-Spectral Synthesis", conforme resumido na figura 10.

Figura 10–Visão esquemática do modelo final



Fonte: Elaboração própria

3. Resultados

Para avaliar a qualidade do modelo proposto, comparamos seus resultados e comportamento com as implementações das duas abordagens mais utilizadas na síntese sonora, a saber: os

Digital Waveguides (Smith, 2006) e o método das diferenças finitas (Bilbao, 2009). Para uma comparação entre esses métodos, o leitor é referido a Karjalainen e Erkut (2004).

Em nome da consistência, as implementações dos dois métodos foram preparadas, na linguagem de programação Python, e estão disponíveis no repositório do Github preparado para este trabalho (Tarjano, 2018), sob o nome de "Digital_Waveguides.py" e "Finite_Differences.py". Alguns exemplos de saídas dessas implementações podem ser ouvidos nas pastas "Demo Digital Waveguides" e "Demo Finite Differences", respectivamente.

O modelo neural apresenta resultados mais realistas do que as implementações tradicionais, e é pelo menos 10 vezes mais eficiente, em comparação com as implementações acima mencionadas.

Para comparação, levando-se em conta a geração de 5 ondas de 1 segundo a 44100 FPS, o método proposto foi, em média, 17 vezes mais rápido que o método baseado em Digital Waveguides e 26 vezes mais rápido que as diferenças finitas. A Tabela 1 resume os resultados.

Tabela 1. Comparação da latência para as 3 implementações

	DiferençasFinitas	Digital Waveguides	SínteseNeuroespectral
1	16.95213819	13.81114888	0.5896253586
2	19.62042928	13.02065825	0.8354649544
3	24.25646234	11.33273816	0.7555158138
4	17.89053726	12.3500874	0.7475204468
5	18.00246739	12.12223339	0.7735056877
Average	**19.34440689**	**12.52737322**	**0.7403264523**
Standard Deviation	**2.908692366**	**0.937300851**	**0.09102950916**

Fonte: Elaboração própria

Os sons gerados pelo modelo proposto podem ser ouvidos, em um contexto musical, no Soundcloud do autor soundcloud.com/carlos-tarjano/sets/synthesis-espectral-espectral

Algumas faixas fazem uso de uma bateria acústica gerada a partir de uma abordagem semelhante à apresentada neste trabalho. Vale a pena notar um instrumento híbrido usado em outras faixas, como a interpretação de Caravan de Duke Ellington; um amálgama alucinado por uma rede semelhante à apresentada aqui, baseada em treinamento usando amostras de um baixo acústico, para as notas mais baixas, um violoncelo para o meio do registro e um violino para as notas mais altas.

O tamanho em disco do modelo final, como pode ser visto no repositório do Github, é muito menor do que o apresentado sintetizadores comerciais de piano padrão da indústria, dos quais o Pianoteq 6 é o mais leve, com aproximadamente 30 Mb.

A principal limitação da metodologia proposta é que todos os parâmetros a serem manipulados no modelo final devem primeiro ter sido incorporados ao processo de treinamento.

Os dois paradigmas de modelagem física usados para comparação, tanto o método de diferenças finitas quanto o método de Digital Waveguides, permitem a manipulação em tempo real de seus parâmetros: nas implementações apresentadas, o ponto de excitação pode ser alterado de onda para onda. Além disso, a qualquer momento o ponto de captação pode ser alterado, mesmo durante a simulação, refletindo no timbre do som gerado.

Ademais, esses modelos se prestam à incorporação razoavelmente trivial de uma fonte, contínua ou periódica, de excitação, e podem ser usados para a simulação de instrumentos de sonoridade contínua, como violinos movidos a arco.

4. Conclusão

O presente trabalho, introduzindo uma nova técnica de modelagem de instrumentos acústicos, demonstra uma possibilidade do uso de redes neurais para a síntese de áudio, estabelecendo o potencial das aplicações em tempo real baseadas nesta técnica. O modelo gera resultados mais confiáveis do que o algoritmo de modelagem acústica em tempo real mais utilizado e eficiente encontrado na literatura, a um menor custo computacional.

Outra vantagem em relação aos modelos baseados em simulação física convencional é que ele pode aprender características de som relevantes que surgem de partes do sistema difíceis de modelar, como a influência de ressonadores de geometria complexa.

O presente trabalho mostra que arquiteturas densas, com uma representação adequada, são capazes de aprender recursos que permitem a reprodução e generalização de amostras sonoras de forma direta; a partir da introdução de uma representação compacta e fisicamente informada de ondas sonoras harmônicas, o trabalho mostra o potencial sinergético entre desenvolvimentos de pesquisa sobre a acústica de instrumentos musicais e o uso de redes neurais para basear modelos para a emulação desses instrumentos ou famílias de instrumentos
.

Além disso, através do uso de funções de ativação especialmente projetadas para acomodar os parâmetros de representação pertinentes e métodos de inicialização apropriados de pesos e vieses, estas arquiteturas podem ser simplificadas, e o número requerido de parâmetros treináveis diminuiu consideravelmente, tornando-os mais efetivos para a síntese de som em tempo real.

As possibilidades de desenvolvimentos futuros nesta área de interseção entre redes neurais e modelagem acústica são numerosas, dada a escassez de investigações similares: seria interessante, por exemplo, usar as saídas de um modelo baseado no método de diferenças finitas, que pode ser formulado de modo a simular características mais sofisticadas de um instrumento, como rigidez de corda, ressonância e vários termos de perda de um dado sistema acústico, ao custo de uma alta demanda de recursos computacionais, para treinar um modelo baseado em Digital Waveguides com uma rede neural no ponto em que as perdas e outros cálculos são concatenados.

Devido ao alto grau de recursão do algoritmo de Digital Waveguides, o treinamento direto baseado nos resultados esperados do modelo é bastante complexo de ser implementado; os vetores de saída de uma simulação baseada em diferenças finitas, entretanto, são totalmente compatíveis com essa abordagem, e a inserção de uma rede neural poderia levar a um modelo que retenha pelo menos parte da precisão da simulação pelo método das diferenças finitas, com eficiência computacional próxima ou até superior à apresentada pelo algoritmo digital de guias de onda.

Em Gully et al. (2017), por exemplo, encontramos um exemplo de trabalho nesta linha, que explora o uso de redes neurais para a identificação de parâmetros relevantes a uma simulação por guias de ondas digitais do trato vocal humano.

Relaxar a simplificação adotada durante o trabalho em relação a decaimentos exponenciais é outro desenvolvimento futuro com potencial interessante: para algumas categorias de som; a

voz humana, por exemplo, o envelope da onda apresenta maior impacto que as frequências nas características perceptivas do som, como inteligibilidade.

Estimar os envelopes com a técnica aqui apresentada, considerando um maior número de pontos e usar uma rede neural para aprender as características daquele envelope para um conjunto de amostras de um instrumento arbitrário, ou até mesmo da voz humana, é outra direção interessante a ser investigada.

Outro desenvolvimento seria implementar o método em uma linguagem de programação mais eficiente, como C ou C ++, com a adição de uma interface visual e compatibilidade com controladores MIDI. Essas modificações podem basear uma linha de produtos comercialmente viável, para serem comercializados em formato autônomo ou VST.

5. Referências

BILBAO, S., Numerical Sound Synthesis, John Wiley & Sons, 2009.

Magenta, [online],disponível em:https://magenta.tensorflow.org/ , acessado em: 02/05/2019 .

ENGEL, J. et al., Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders, CoRR, v. abs/1704.01279, 2017.

FLETCHER, H., Normal Vibration Frequencies of a Stiff Piano String, The Journal of the Acoustical Society of America, v. 36, n. 1, p. 203-209, 1964.

University of Iowa Electronic Music Studios, [online], disponível em: http://theremin.music.uiowa.edu/MIS.html, acessado em: 02/05/2019.

GULLY, A. J. et al., Articulatory Text-to-Speech Synthesis Using the Digital Wave-guide Mesh Driven by a Deep Neural Network, in Interspeech 2017, 2017.

KARJALAINEN, M. e ERKUT, C., Digital Waveguides versus Finite Difference Structures: Equivalence and Mixed Modeling, Journal on Advances in Signal Processing (EURASIP), v. 2004, n. 7, 2004.

KOENIG, D. M. e FERICH, D. D., Spectral Analysis of Musical Sounds with Emphasis on the Piano, PAPERBACKSHOP UK IMPORT, 2015.

VAN DEN OORD, A. et al., WaveNet: A Generative Model for Raw Audio, CoRR, v. abs/1609.03499, 2016.

ROBERTS, A. et al., Learning Latent Representations of Music to Generate Interactive Musical Palettes, in Proceedings of the 2018 ACM Workshop on Intelligent Music Interfaces for Listening and Creation, MILC@IUI 2018, 2018.

ROBERTS, A. et al., A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music, in Proceedings of the 35th International Conference on Machine Learning, v. 80, p. 4364-4373, 2018.

SMITH, J. O., A Basic Introduction to Digital Waveguide Synthesis (for the Technically Inclined), Center for Computer Research in Music and Acoustics (CCRMA), 2006.

SMITH, J. O., Digital Waveguide Architectures for Virtual Musical Instruments, in Handbook of Signal Processing in Acoustics, p. 399-417, 2008.

STAUDT, P., Development of a Digital Musical Instrument with Embedded Sound Synthesis, 2016.

TARJANO, C., Neurospectral Audio Synthesis Repository, [online], disponível em: https://github.com/tesserato/neurospectral-audio-synthesis ,acessado em: 02/05/2019 .

The Railsback Curve, [online],disponível em: https://commons.wikimedia.org/wiki/File:Railsback2.png , acessado em: 02/05/2019