Identificação de nota musical e classificação de timbre

Gisele Goulart Tavares da Silva[1], **Guilherme Almeida Félix da Silva**[1]

[1]Departamento de Mecânica Aplicada e Computacional
Universidade Federal de Juiz de Fora

giselegoulart@ice.ufjf.br, guilherme.felix@engenharia.ufjf.br

Resumo. *O objetivo do presente relatório é apresentar os resultados obtidos pelo algoritmo implementado no primeiro trabalho da disciplina Trabalho Multidisciplinar. O método realiza a leitura de uma nota musical no formato .wav e através de modelagem matemática da onda como uma série de Fourier retorna a nota referente na escala musical. Também é realizada uma análise de timbre, por meio da análise do perfil da onda no domínio da frequência.*

1. Introdução

1.1. O que é uma onda sonora?

Alteração de pressão que se propaga no espaço tridimensional. No caso da música, fenômeno periódico produzido pela vibração de cordas, colunas de ar, paletas, membranas.

1.2. Como representá-la?

O movimento em questão é o deslocamento do meio (corda, ar) em relação a uma posição de equilíbrio sujeito a uma força restauradora, proporcional ao deslocamento. Esse movimento é descrito por uma equação diferencial de segunda ordem cuja solução geral é uma função seniodal [Benson 2008].

1.3. O que é uma nota? Como representá-la?

Uma nota musical é nada mais que um som com frequência definida. Quando uma corda vibra, entretanto, o som produzido contém várias ondas senoidais de diversas amplitudes e frequências (vários harmônicos) de forma que o sinal enviado para o cérebro é mais complexo. Por essa razão a decomposição de uma onda sonora é feita em soma de funções seno, chamada de série de Fourier. Para cada frequência de vibração, o cérebro atribui um som diferente (uma nota diferente) [Charles 1992]. A Figura 1 apresenta as diferentes vibrações de uma corda.

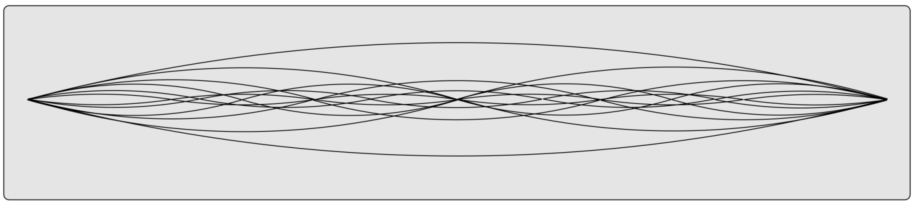

Figura 1. Diferentes modos de vibração de uma corda.

1.4. O timbre de uma fonte geradora

Duas notas de mesma frequência emitidas por instrumentos diferente serão percebidas de forma diferente pelo ouvinte pelo fato de as amplitudes das ondas componentes do som resultante terem valores distintos para cada fonte. Seja um piano e um violão emitindo a mesma nota e exatamente à mesma frequência (440 Hz, por exemplo). Ao ser decomposto, o som emitido pelo piano conterá várias ondas seno com amplitudes diferentes. Essas amplitudes, porém, não serão as mesmas nas mesmas funções seno obtidas da decomposição do som emitido pelo violão. Essa diferença é que chama-se timbre.

O objetivo do presente trabalho é a implementação de um procedimento computacional para leitura e determinação de nota musical a partir da leitura de arquivos .wav com o uso de modelagem matemática. A partir do formato gráfico dos dados no domínio da frequência, uma análise do timbre também será realizada, de modo que o instrumento utilizado para a reprodução da nota possa ser de detectado.

2. Séries de Fourier

Série de Fourier é uma maneira de representar uma função como soma de ondas senoidais. No seu estudo sobre análise harmônica, Fourier estabeleceu que é possível decompor qualquer função periódica em uma soma de infinita de funções seno e cosseno, ou exponenciais complexas. Dessa forma, uma função periódica x(t) pode ser escrita na forma [Bell 2017]:

$$x(t) = a_0 + \sum_{k=1}^{\infty} (a_k cos\left(\frac{2\pi kt}{T}\right) + b_k sin\left(\frac{2\pi kt}{T}\right)) \tag{1}$$

onde os coeficientes a_0, a_k e b_k são conhecidos como coeficientes de Fourier e podem ser determinados por:

$$\begin{aligned} a_0 &= \frac{1}{T}\int_0^T x(t)dt \\ a_k &= \frac{2}{T}\int_0^T x(t)cos\left(\frac{2\pi kt}{T}\right)dt \\ b_k &= \frac{2}{T}\int_0^T x(t)sin\left(\frac{2\pi kt}{T}\right)dt \\ k &= 1, 2, 3, \ldots \end{aligned}$$

Neste caso, o valor de cada coeficiente corresponde a amplitude de cada onda componente da onda original, definida por $x(t)$, onde cada k corresponde a um harmônico, sendo $k = 1$ chamado de harmônico fundamental.

Isso significa dizer que, uma onda periódica com frequência f_1 pode ser decomposta em soma de senos e cossenos de frequências diferentes, porém múltiplos inteiros de f_1, ponderados pelos coeficientes a_k.

2.1. Harmônicos musicais

Instrumentos musicais produzem som como resultado da vibração de um meio físico, como uma paleta, uma corda ou uma coluna de ar. Essa vibração causa variação periódica da pressão no ar, que se propaga e sensibiliza o aparelho auditivo humano, sendo interpretado pelo cérebro como som. Entendendo o som como uma sucessão de compressões e rarefações do ar no entorno da fonte, o período T de uma onda é o tempo necessário para que o pulso se repita, ou seja, o intervalo de tempo necessário para que duas compressões (ou rarefações) sucessivas passem pelo mesmo ponto do espaço. Nesse intervalo de tempo, tem-se que exatamente um comprimento de onda viajou por esse espaço. O número de comprimentos de onda que viaja pelo espaço em exatamente um período T é chamado de frequência f da onda. Essa frequência $f_1 = 1/T$, que define a altura da nota produzida, é chamada de frequência fundamental ou primeiro harmônico. Quando um instrumento produz uma nota, não apenas o harmônico fundamental é produzido, mas também outras frequências, f_2, f_3, f_4, tais que são múltiplos inteiros do harmônico fundamental. Cada um desses múltiplos, porém, é produzido com uma amplitude diferente, de forma que alguns deles possuem grande amplitude e outros valores muito pequenos.

Quando se decompõe uma onda sonora em série de Fourier, encontram-se as frequências que somadas resultam naquela determinada onda, naquele determinado som. Encontrar os coeficientes de Fourier para uma onda sonora periódica significa encontrar as amplitudes de cada harmônico que somado resulta na onda original.

Por meio dos coeficientes de Fourier de uma função, pode-se determinar o harmônico de maior amplitude e, dessa forma, determinar a frequência do som emitido. Conhecendo a frequência do som emitido pode-se determinar a nota emitida.

Musicalmente falando, duas notas cujas frequências estão na relação 1:2 formam um intervalo conhecido como "oitava". Assim, uma nota a 440 Hz e uma nota a 880 Hz estão em um intervalo de uma oitava um em relação ao outro. Note que se forem considerados $f_1 = 440$ Hz e $f_2 = 880$ Hz, então tem-se $f_2 = 2f_1$. O sistema musical usado atualmente é tal que divide o intervalo de uma oitava em 12 partes iguais resultando, portanto, em 12 notas com frequências bem definidas. O conjunto de frequências a seguir indica, para o intervalo de uma oitava, as 12 notas [Nov 2017]:

Nota	Frequência (Hz)
Lá (A)	440 Hz
Lá sustenido (A#)	466.16 Hz
Si (B)	493.88 Hz
Dó (C)	523.25 Hz
Dó sustenido (C#)	554.37 Hz
Ré (D)	587.33
Ré sustenido (D#)	622.25 Hz
Mi (E)	659.26 Hz
Fá (F)	698.46 Hz
Fá sustenido (F#)	739.99 Hz
Sol (G)	783.99 Hz
Sol sustenido (G#)	830.61 Hz
Lá (A)	880 Hz

No referido intervalo a razão entre quaisquer duas frequências adjacentes é igual a $\sqrt{12}$, de maneira que dada uma frequência de referência f_0 qualquer outra frequência f_i pode ser determinada por meio de:

$$f_i = f_0.2^{\frac{i}{12}} \tag{2}$$

Usando essa relação é possível determinar se a frequência de maior amplitude na série de Fourier corresponde à frequência da nota esperada.

3. Materiais e Métodos

Foram usadas amostras de áudio (PCM a 44100 Hz) referentes a 4 notas diferentes para a identificação da nota e 3 notas Lá (440 Hz) executadas por 3 instrumentos diferentes, para a identificação de timbre. Cada amostra foi submetida a uma transformada rápida de Fourier e o espectro resultante foi analisado. Para a identificação da nota, a frequência de maior intensidade foi usada. O resultado comparado com o da expressão $f_i = f_0 2^{\frac{i}{12}}$, onde f_0 é uma frequência de referência. Para a identificação do timbre a união das maiores amplitudes foi utilizada.

Foram utilizadas amostras musicais de piano para os testes para determinação de nota musical e de clarinete, flauta e guitarra para a análise de timbre. Os amostras utilizadas para teste de determinação de nota foram coletadas com o uso da biblioteca de sons de softwares de instrumentos virtuais, cujas amostras são gravadas em alta qualidade a partir de instrumentos reais.

O procedimento computacional foi desenvolvido com o uso da linguagem de programação Python, por conta de diminuir a complexidade das funções implementadas através de bibliotecas e pacotes da linguagem. Algumas das bibliotecas utilizadas foram: Numpy, Scipy, Matplotlib e Math.

A teoria de Fourier, descrita anteriormente, decompõe ondas periódicas (infinitas) em somas de senos e cossenos. É desejável analisar funções não periódicas de forma similar, o que leva à teoria da transformada de Fourier. O som captado nas amostras utilizadas não são periódicas de fato, uma vez que ondas periódica não apresentam início ou fim. Por isso, deseja-se uma análise da forma de onda por uma “janela”de tempo definida. De fato, portanto, o que é feito é uma transformada de Fourier e não uma série [Benson 2008].

Além disso, para a abordagem computacional, a função não é contínua no tempo, mas discreta, o que leva a uma outra abordagem para a transformada de Fourier, conhecida como transformada discreta de Fourier. A função usada no procedimento deste trabalho é uma implementação otimizada desta operação, conhecida por transformada rápida de Fourier e é amplamente utilizada em diversos algoritmos e processos de tratamento de ondas[1][Higuti 2017] e presente em bibliotecas de diversas linguagens, em particular Python. Dessa forma, procedimento computacional foi executado de acordo com as seguintes etapas:

1. Determinação da frequência fundamental f_0 a ser utilizada como referência na Equação 2 e atribuição do array de 12 notas musicais, partindo do C até o B;

[1]O padrão de codificação de áudio mpeg layer 1, por exemplo, utiliza essa operação em uma de suas etapas

2. Leitura e processamento do arquivo de entrada .wav;
3. Normalização dos dados de entrada. Este passo é necessário para o correto cálculo da Transformada de Fourier;
4. Cálculo da Transformada de Fourier, responsável por trazer os dados do domínio do tempo para o domínio da frequência. Este passo foi realizado através da função fft (Transformada Rápida de Fourier) do pacote Scipy do Python;
5. Procura do máximo global, frequência onde ocorre a maior intensidade de som;
6. Cálculo da variável i da Equação 2, responsável por indicar a posição no vetor de notas musicais para a nota emitida;
7. Montagem dos resultados graficamente, através da plotagem de gráficos da intensidade em função da frequência.

4. Resultados

O algoritmo foi executado para 4 arquivos musicais distintos com o objetivo de determinar a nota que foi emitida. Os resultados obtidos pelo procedimento computacional podem ser comparados com os valores que constam na Tabela 1.

As Figuras 2, 3, 4 e 5 apresentam os resultados encontrados para as 4 notas musicais testadas. Para a primeira nota, o resultado encontrado foi F# (Fá sustenido), através do pico de valor 92.05 e na segunda E (Mi) com o pico 1322.01. Observando os valores constantes na Tabela 1, é possível constatar que os resultados encontrados pelo algoritmo possuem correspondências aproximadas e as notas retornadas através da análise do gráfico no domínio da frequência refletem o som emitido. Na terceira nota o máximo global ocorreu em 466.92 e na quarta em 417.08, correspondendo a A# e G#, respectivamente.

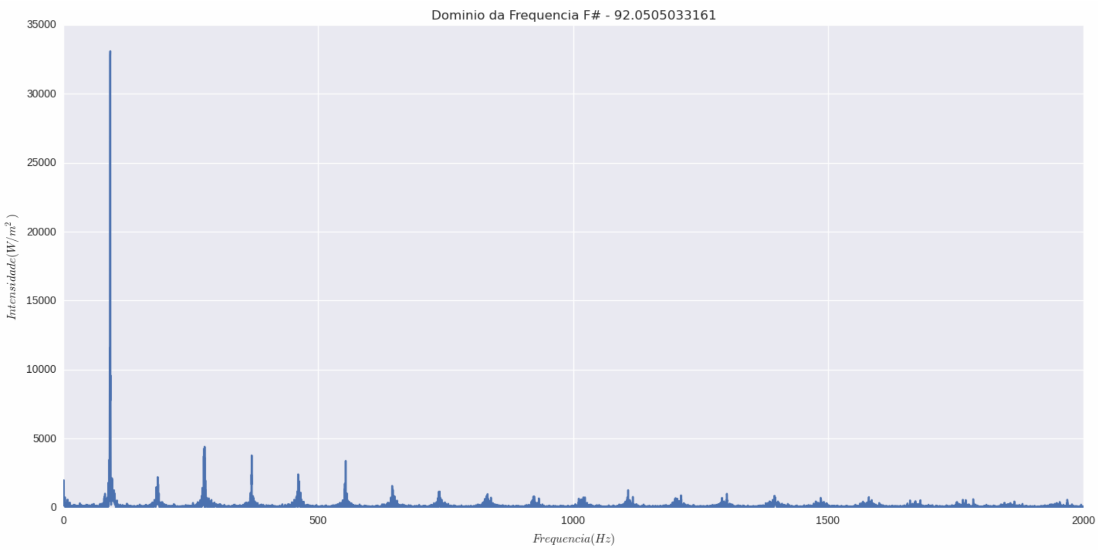

Figura 2. Resultado encontrado para a primeira nota testada. O máximo global ocorre no 92.05, correspondendo à nota F#.

As Figuras 6, 7 e 8 apresentam as comparações realizadas para os perfis de timbre de 3 notas musicais, sendo de piano, clarinete e trompa. Podemos notar certa similaridade entre os gráficos obtidos e os de referência [Bell 2017], mas por conta de ser baixa não conseguimos verificar os instrumentos utilizados através do timbre de forma direta. A

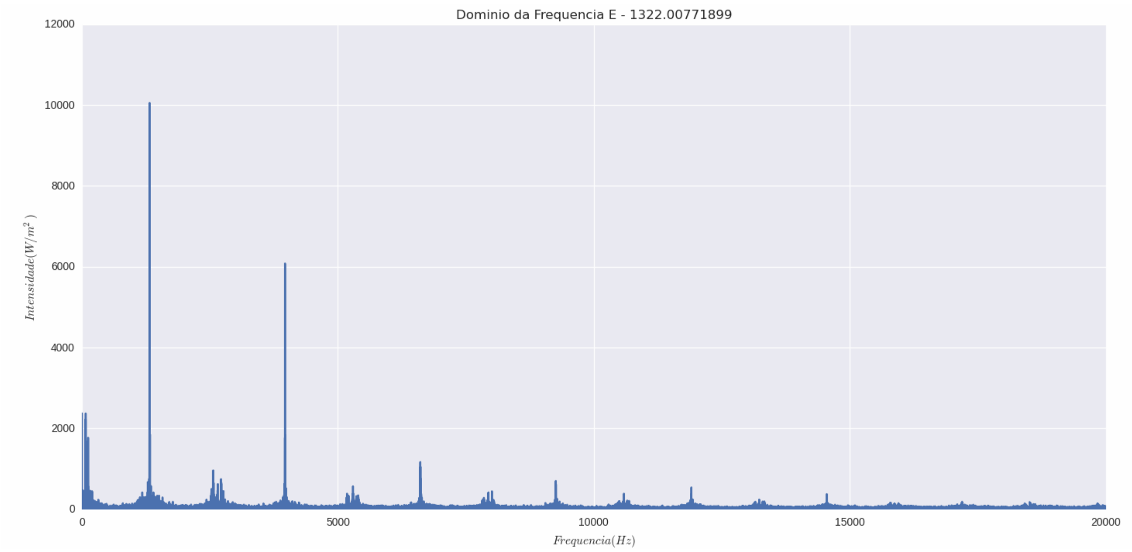

Figura 3. Resultado encontrado para a segunda nota testada. O máximo global ocorre no 1322.01, correspondendo à nota E.

implementação de um tratamento nos dados é necessária para a correta comparação das curvas.

5. Conclusões

Os resultados indicam concordância da determinação da nota emitida com o método utilizado. Todos os exemplos apresentaram diferença muito pequena para o valor esperado de frequência.
A determinação do timbre, em contrapartida, não foi bem sucedida. Inspecionando-se visualmente o espectro produzido pelas amostras testadas e comparando com a referência, não é possível concluir que há grande similaridade. Destaca-se o fato de a referência ter reescrito o eixo horizontal em termos de cada harmônico e normalizado a escala vertical, na qual o valor do primeiro harmônico é ajustado para 1. Os gráficos aqui apresentados possuem no eixo horizontal a frequência e no eixo vertical a amplitude absoluta obtida a partir da transformada das amostras.
Dado o exposto conclui-se que a inspeção do espectro de frequências para a determinação do timbre de determinado instrumento requer maior nível de aprofundamento, possivelmente sendo necessária a utilização de métodos como busca local ou mesmo o uso de outro tipo de transformada.

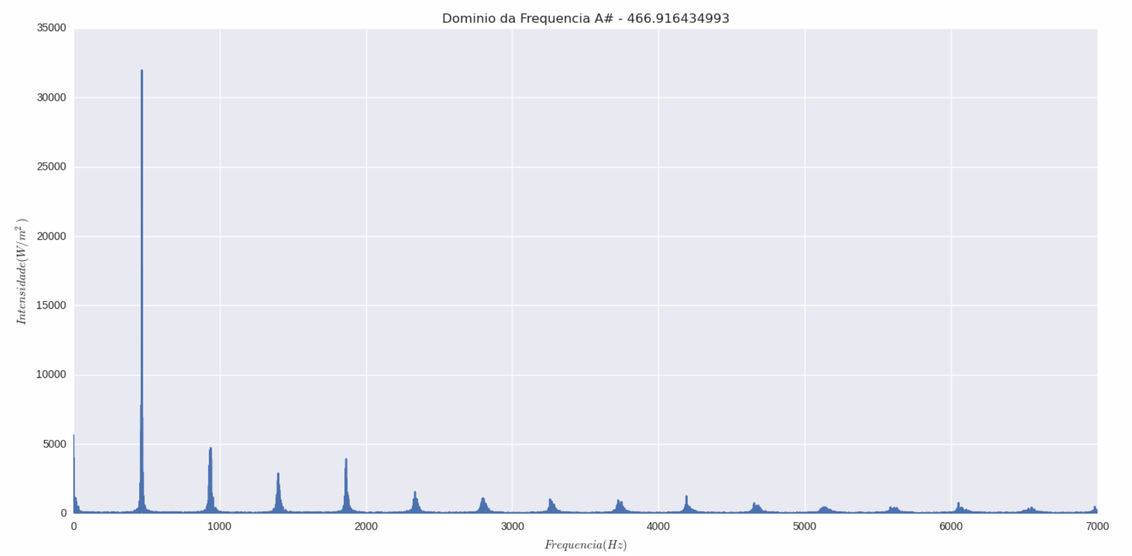

Figura 4. Resultado encontrado para a terceira nota testada. O máximo global ocorre no 466.92, correspondendo à nota A#.

Referências

Bell, M. (2004 (accessed September 10, 2017)). *Fourier Analysis in Music*.

Benson, D. J. (2008). Music: a mathematical offering. *The Mathematical Intelligencer*, 30(1):76–77.

Charles, T. (1992). *Exploring music: The science and technology of tones and tunes*. CRC Press.

Higuti, R. T. (2004 (accessed September 10, 2017)). *Processamento de Sinais - Transformada Rápida de Fourier - FFT*.

Nov, Y. (accessed September 3, 2017). *Explaining the Equal Temperament*.

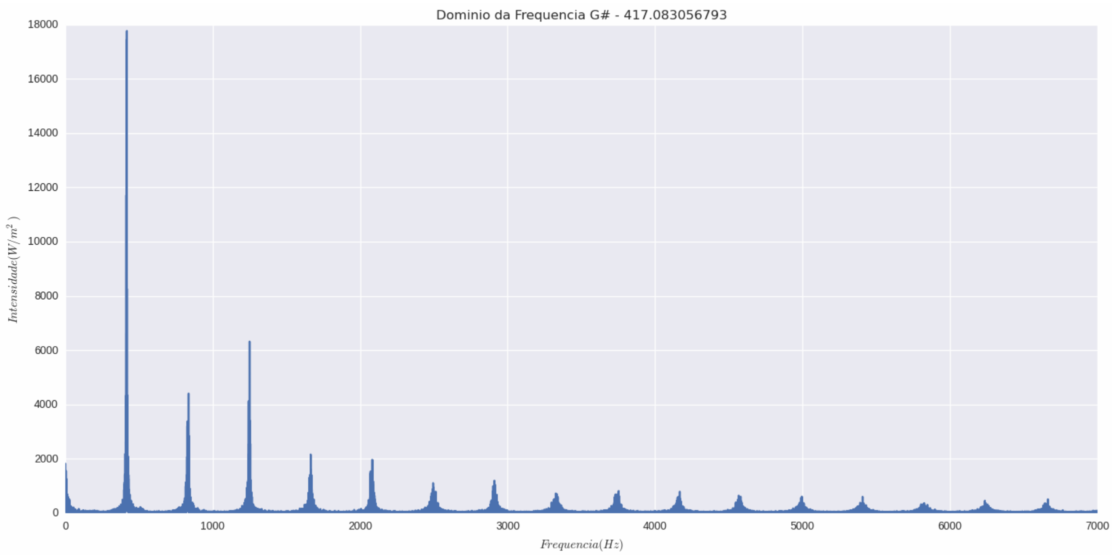

Figura 5. Resultado encontrado para a quarta nota testada. O máximo global ocorre no 417.08, correspondendo à nota G#.

Tabela 1. Tabela de referência utilizada para comparação das notas encontradas pelo algoritmo de acordo com os picos de frequência.

Nota	Frequência (Hz)										
	-	1° Oitava	2° Oitava	3° Oitava	4° Oitava	5° Oitava	6° Oitava	7° Oitava	8° Oitava	9° Oitava	-
C	16.35	32.70	65.41	130.81	261.63	523.25	1046.50	2093.00	4186.01	8372.02	16744.03
C#	17.32	34.65	69.30	138.59	277.18	554.37	1108.73	2217.46	4434.92	8869.84	17739.69
D	18.35	36.71	73.42	146.83	293.66	587.33	1174.66	2349.32	4698.64	9397.27	18794.54
D#	19.45	38.89	77.78	155.56	311.13	622.25	1244.51	2489.02	4978.03	9956.06	19.912.125
E	20.60	41.20	82.41	164.81	329.63	659.26	1318.51	2637.02	5274.04	10548.08	21096.17
F	21.83	43.65	87.31	174.61	349.23	698.46	1396.91	2793.83	5587.65	11175.30	22350.61
F#	23.12	46.25	92.50	185.00	369.99	739.99	1479.98	2959.96	5919.91	11839.82	23679.64
G	24.50	49.00	98.00	196.00	392.00	783.99	1567.98	3135.96	6271.93	12543.86	25087.71
G#	25.96	51.91	103.83	207.65	415.30	830.61	1661.22	3322.44	6.644.875	13289.75	26579.50
A	27.50	55.00	110.00	220.00	440.00	880.00	1760.00	3520.00	7040.00	14080.00	28160.00
A#	29.14	58.27	116.54	233.08	466.16	932.33	1864.65	3729.31	7458.62	14917.24	29834.49
B	30.87	61.74	123.47	246.94	493.88	987.77	1975.53	3951.07	7902.13	15804.26	31608.53

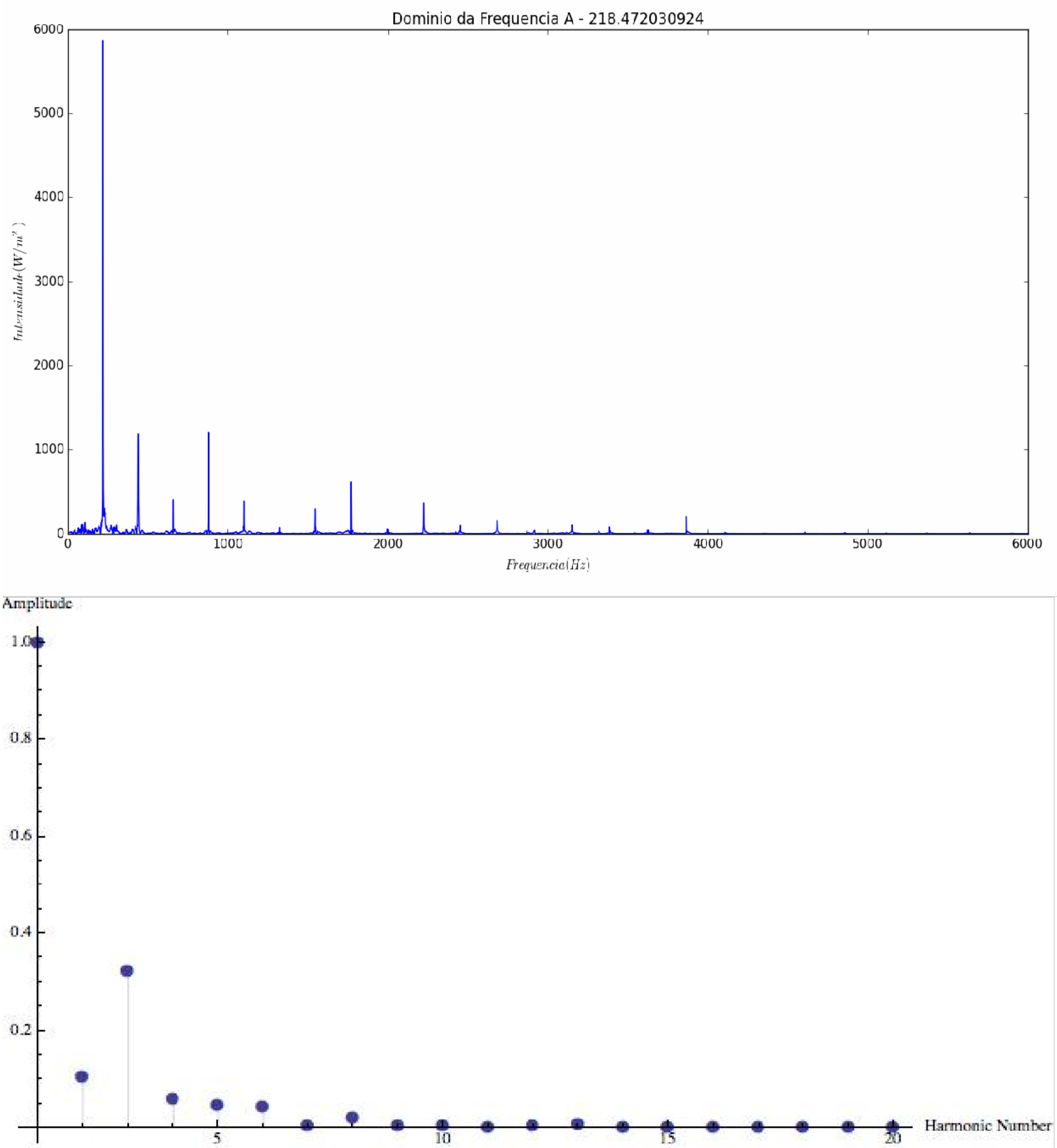

Figura 6. Comparação do perfil de timbre encontrado para o piano em relação à referência.

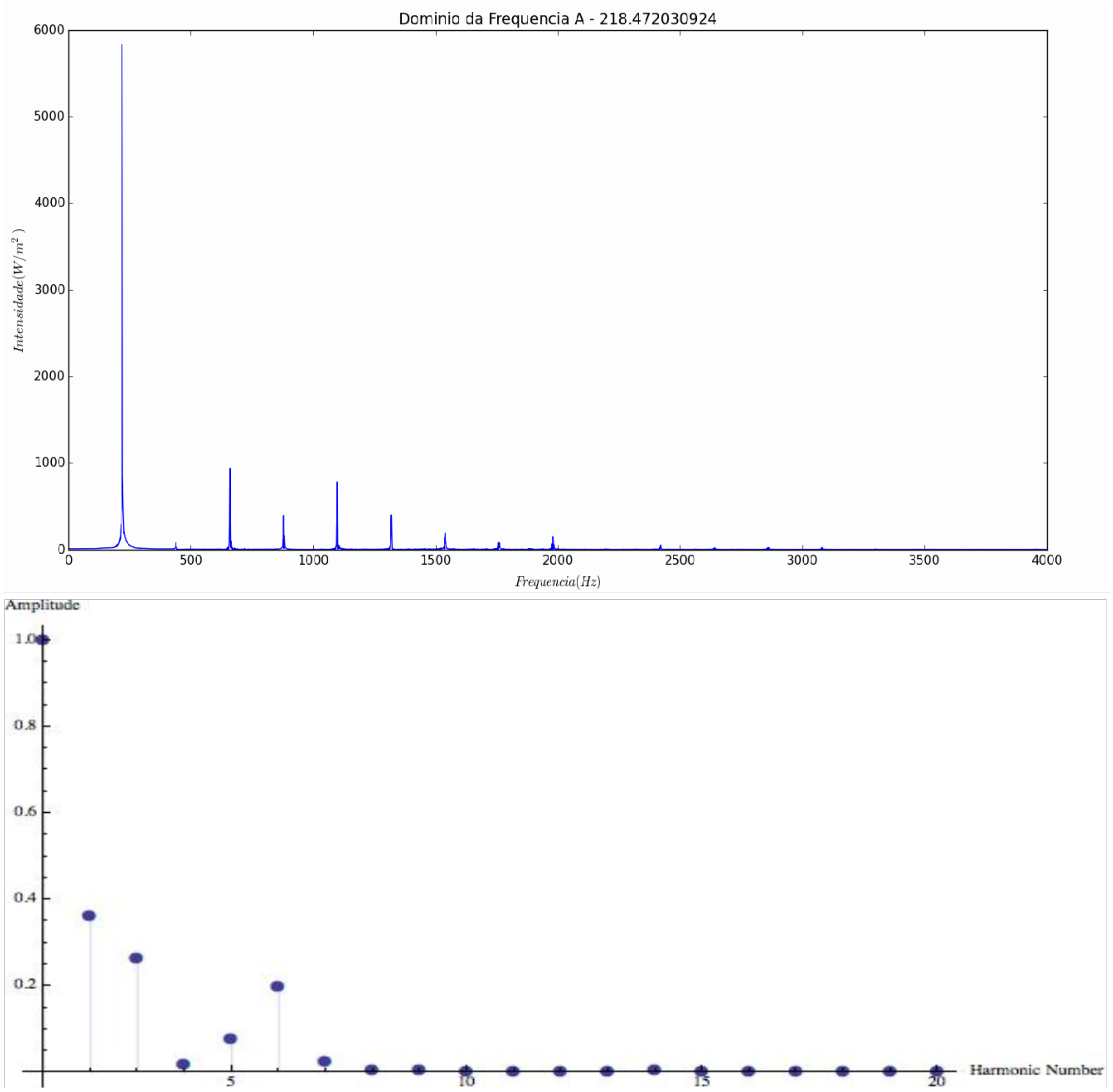

Figura 7. Comparação do perfil de timbre encontrado para o clarinete em relação à referência.

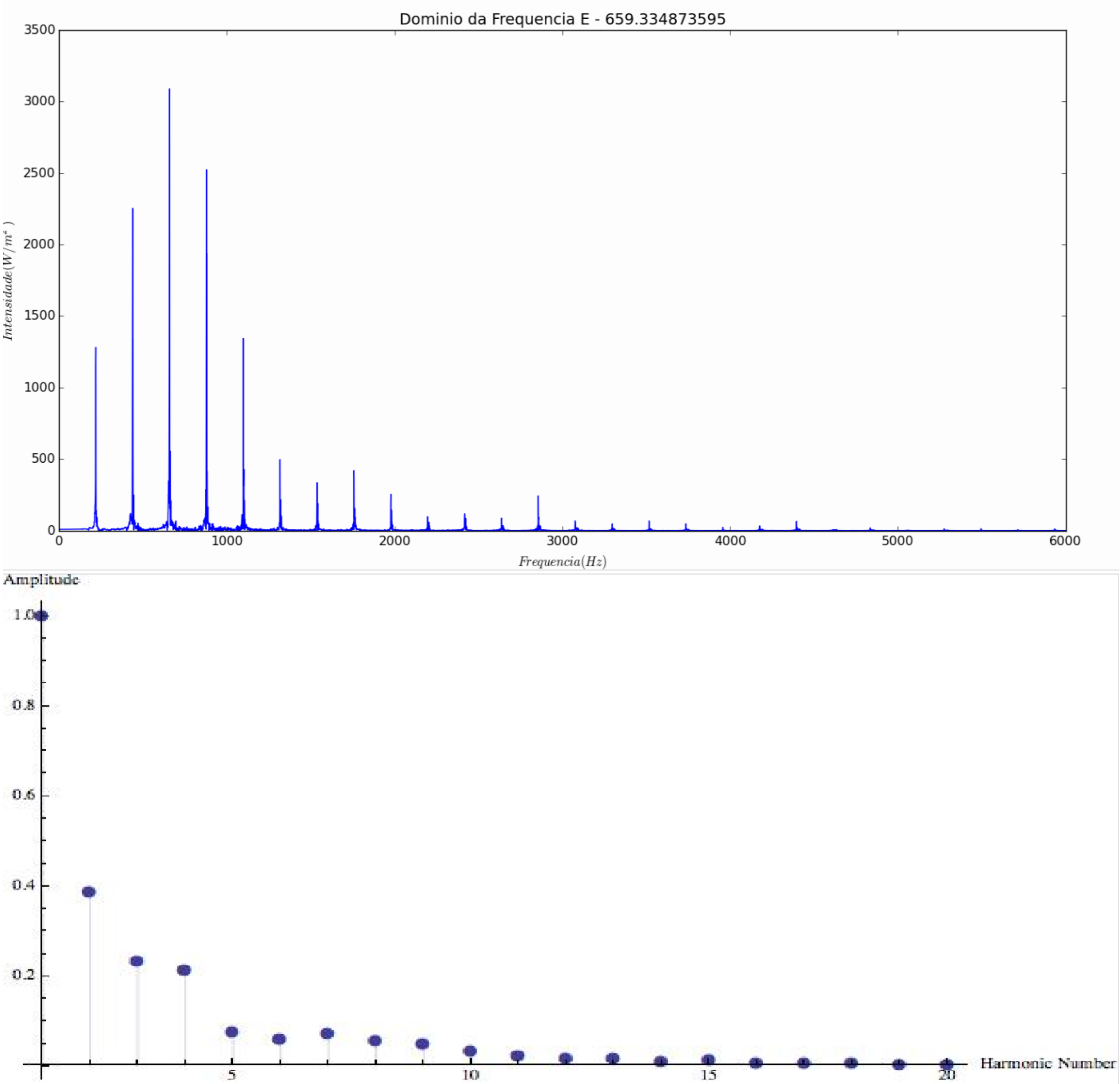

Figura 8. Comparação do perfil de timbre encontrado para a trompa em relação à referência.