

Análise dos métodos de reconstrução de imagens aplicado a sinais de áudio

Carlos Eduardo Leal de Castro

Orientador: Dr. Leonardo Koller Sacht
Universidade Federal de Santa Catarina



2017

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Introdução/Motivação

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Introdução/Motivação

Em sua Tese de Doutorado, SACHT (2014) desenvolveu um método de Quasi-Interpoladores Otimizados para reconstrução de imagens, com resultado superior aos apresentados anteriormente nesta área de pesquisa.

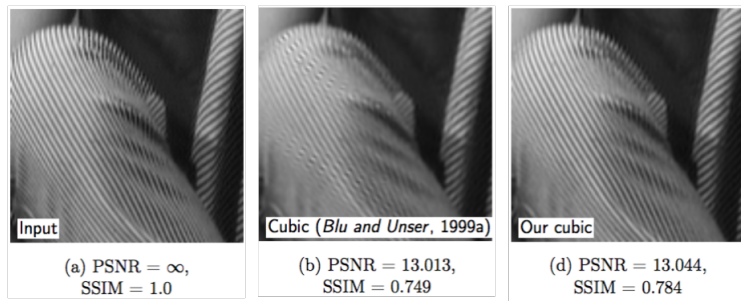


Figura: Comparação entre diferentes interpoladores. Imagem retirada da Tese de SACHT (2014).

Introdução/Motivação

De acordo com GIANNAKOPOULOS; PIKRAKIS (2014) a manipulação de Áudio digital é uma área que vem crescendo muito ao longo dos anos. Principalmente, no mundo estão trabalhando nas seguintes áreas de pesquisa relacionadas a áudio:

- Reconhecimento de discurso;
- Identificação e visualização do orador;
- **Recuperação de informações musicais:**
- Detecção de eventos em áudio;

Introdução/Motivação

- Criação de novos instrumentos musicais;
- Playback musical via uso de Inteligência artificial;



Figura: Imagem retirada do Youtube a partir do link <https://www.youtube.com/watch?v=967MuViGRm8>

Pergunta de Pesquisa

Pergunta de Pesquisa: O Método de Quasi-Interpoladores Otimizados para reconstrução de imagens é aplicável a sinais de áudio? Quais métodos de reconstrução de imagens dá o melhor resultado ao aplicarmos em sinais de áudio? Como podemos ajustá-los para reconstruir de maneira eficiente um sinal de áudio digital?

Organização

O trabalho está organizado em 4 capítulos:

- Conhecimentos preliminares: Introdução às Séries de Fourier;
- Reconstrução: Conceitos principais e admissibilidade;
- Construção de Scripts e Resultados;
- APÊNDICE A: Scripts do Matlab.

Pipeline de Reconstrução

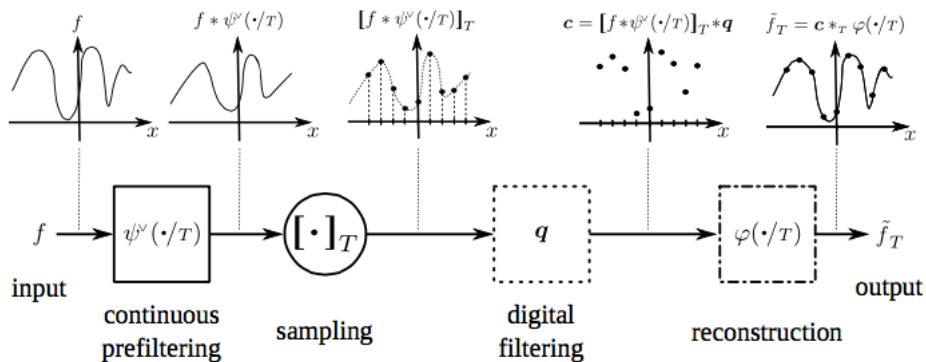


Figura: Imagem retirada da Tese de Doutorado de SACHT (2014).

Espaço de Funções

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Espaço de Funções

Definição

Seja $f : \mathbb{R} \rightarrow \mathbb{C}$ uma função. Dizemos que $f \in \mathcal{L}_2$ se

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty \quad (1)$$

Definição

Seja $q : \mathbb{Z} \rightarrow \mathbb{C}$ uma sequência. Dizemos que $q \in \ell_2$ se

$$\sum_{k \in \mathbb{Z}} |q_k|^2 < \infty \quad (2)$$

Espaço de Funções

Definição

O Espaço de Schwartz ou espaço das funções rapidamente decrescentes, que denotamos por \mathbb{S} , é o subespaço vetorial formado pelas funções $\zeta \in C^\infty(\mathbb{R}^n)$ tais que

$$\lim_{|x| \rightarrow \infty} \|x\|^k \frac{\partial^\alpha}{\partial x^\alpha} \zeta(x) = 0,$$

quaisquer que sejam $k \in \mathbb{N}$ e $\alpha \in \mathbb{N}^n$.

Reconstrução: Conceitos principais e admissibilidade

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

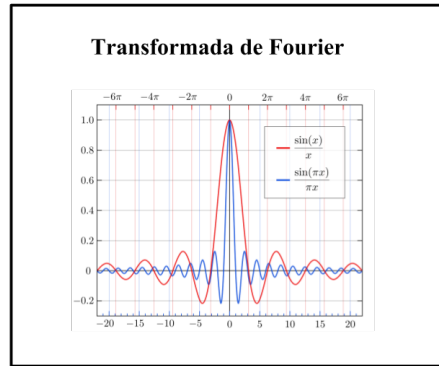
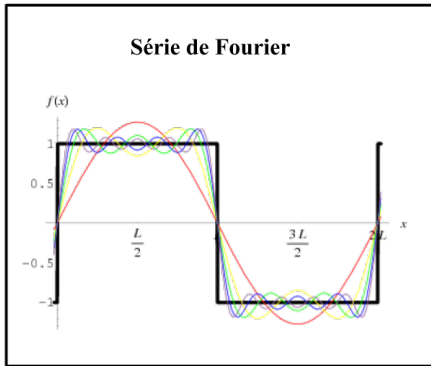
Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Série de Fourier - Transformada de Fourier



Sequências e Séries de Funções \rightarrow Teorema sobre a Convergência Uniforme da Série de Fourier \rightarrow Transformada de Fourier e sua Inversa.

Convolução (OSGOOD, 2009)

Definição

Dadas funções f e g e sequências c e q e $T \in \mathbb{R}$ a distância entre dois pontos, suas convoluções contínua, discreta e mista são dadas por:

$$\textbf{Contínua} : (f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x - t)dt \quad (3)$$

$$\textbf{Discreta} : (c * q)_n = \sum_{k \in \mathbb{Z}} c_k q_{n-k} \quad (4)$$

$$\textbf{Mista} : (q *_T f)(x) = \sum_{k \in \mathbb{Z}} q_k f(x - kT) \quad (5)$$

Filtro Digital

Filtros FIR

Um filtro digital \mathbf{q} é dito FIR quando contém um número finito de elementos diferentes de zero.

$$\text{fir_dalai1} : \mathbf{q} = [0, \dots, 0, \frac{7}{720}, \frac{-11}{90}, \frac{49}{40}, \frac{-11}{90}, \frac{7}{720}, 0, \dots, 0]$$

Filtros IFIR

Um filtro digital é dito IFIR se é a inversa de um filtro FIR. Se \mathbf{q} é um filtro FIR, o filtro IFIR \mathbf{q}^{-1} é tal que

$$\mathbf{q} * \mathbf{q}^{-1} = \delta = [\dots, 0, 0, 1, 0, 0, \dots].$$

Amostragem em grade T

Definição

Seja $f : \mathbb{R} \longrightarrow \mathbb{C}$ uma função contínua. Consideramos sua **Amostragem em grade T** sua discretização com espaçamento T , ou seja,

$$[f]_T = [\dots, f(-T), f(0), f(T), \dots].$$

Subespaço Vetorial de Geradores

O subespaço Vetorial das funções geradas por cópias de um gerador φ , com amostragem em grade T é dado por

$$V_{\varphi,T} = \{\tilde{f} = \sum_{k \in \mathbb{Z}} c_k \cdot \varphi(x - kT) \mid \forall c \in \ell_2\}.$$

Exemplo

O subespaço B-spline é a família de geradores mais fundamental, que é dada por

$$\beta^0 = \begin{cases} 1, & \text{se } |x| < 0.5 \\ 0.5, & \text{se } |x| = 0.5 \\ 0, & \text{se } |x| > 0.5 \end{cases} \quad \text{e } \beta^n = \beta^{n-1} * \beta^0 \quad (6)$$

Reconstrução: Admissibilidade

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Admissibilidade

Teorema

Seja $T > 0$. Se $f \in W_2^r$, para algum $r > \frac{1}{2}$ e seja ψ um filtro digital tal que $\exists k \in \mathbb{R}$ em que $\|\hat{\psi}\|_\infty \leq k < \infty$, então $[f * \hat{\psi}^\vee(\cdot/T)]_T \in \ell_2$

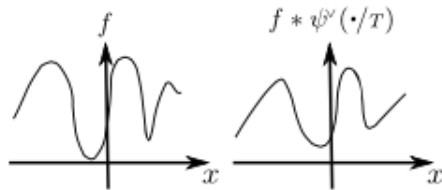


Figura: Recorte do Pipeline de reconstrução

Admissibilidade (cont.)

Teorema

Seja $c \in \ell_2$ e seja q um filtro digital FIR, IFIR ou FIR-IFIR. Então $c * q \in \ell_2$.

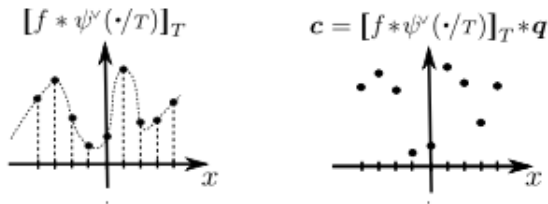


Figura: Recorte do Pipeline de reconstrução

Teorema de Shannon

Teorema

Seja f uma função para qual existe $W > 0$ tal que $\hat{f}(w) = 0, \forall w \in \mathbb{R} \setminus [-W, W]$. Então, a aproximação dada em nosso esquema de reconstrução produz $\tilde{f}_T = f$, se $T \leq \frac{1}{2W}$, $\psi = \delta$, $q = \delta$ e $\varphi(x) = \text{sinc}(x) = \frac{\sin \pi x}{\pi x}$

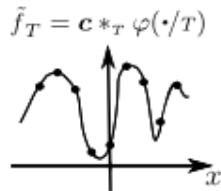


Figura: Recorte do Pipeline de reconstrução

Admissibilidade (cont.)

Este resultado não é usado na prática, pois:

1. Sons e imagens, geralmente, não são “banda limitada”;
2. $\varphi(x) = \frac{\sin \pi x}{\pi x}$ não tem suporte compacto.

Isso tornaria a convolução $c *_T \varphi$ impossível de ser implementada.

Admissibilidade (cont.)

Por conta disso, precisamos encontrar qual função do espaço $V_{\varphi, T}$ está mais próxima de uma dada função f . Na métrica do espaço \mathcal{L}_2 , a função que queremos é a projeção ortogonal da função f no espaço $V_{\varphi, T}$, ou seja, a função $c *_T \varphi \in V_{\varphi, T}$ que é solução para o problema

$$\arg \min_{c \in \ell_2} \| f - (c *_T \varphi) \|_{\mathcal{L}_2} := P_{\varphi, T}(f) \quad (7)$$

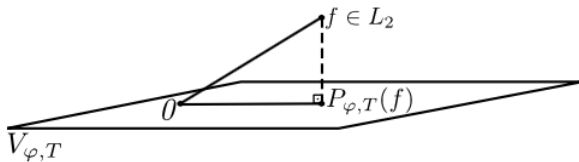


Figura: Representação da projeção ortogonal de uma função $f \in \mathcal{L}_2$ em $V_{\varphi, T}$. Retirada da tese de SACHT (2014).

Filtros e Geradores

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Filtros e Geradores

Filtros Finite Impulse Response

<i>Dalai 1</i>	<i>Dalai 2</i>	<i>Dalai 3</i>
----------------	----------------	----------------

Filtros Inverse Finite Impulse Response

β -spline 1	β -spline 2	β -spline 3
-------------------	-------------------	-------------------

Condat 1	Condat 2	Condat 3
----------	----------	----------

Sacht-Nehab 1	Sacht-Nehab 2	Sacht-Nehab 3
---------------	---------------	---------------

Omoms 2	Omoms 3	
---------	---------	--

Geradores

β -spline 1	β -spline 2	β -spline 3
-------------------	-------------------	-------------------

Sacht-Nehab 1	Sacht-Nehab 2	Sacht-Nehab 3
---------------	---------------	---------------

Omoms 2	Omoms 3	
---------	---------	--

Geradores

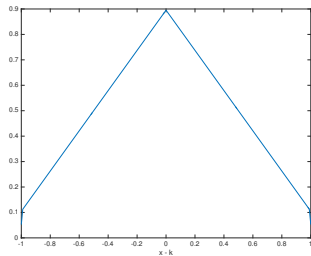


Figura: Gerador *sacht_nehab1*

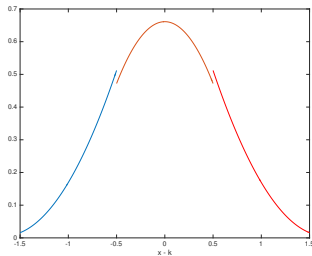


Figura: Gerador *sacht_nehab2*

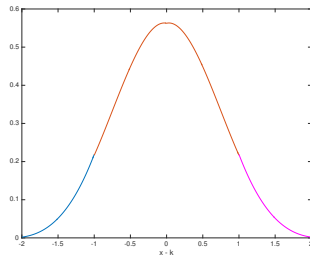


Figura: Gerador *sacht_nehab3*

Filtros e Geradores

Das quais, na reconstrução de imagens, são combinados da seguinte forma:

Filtros e Geradores	
<i>ifir$_{\beta}$-spline 1 e β-spline 1</i>	<i>ifir$_{\beta}$-spline 2 e β-spline 2</i>
<i>ifir$_{\beta}$-spline 3 e β-spline 3</i>	<i>ifir_condat 1 e β-spline 1</i>
<i>ifir_condat 2 e β-spline 2</i>	<i>ifir_condat 3 e β-spline 3</i>
<i>ifir_sacht_nehab 1 e sacht_nehab 1</i>	<i>ifir_sacht_nehab 2 e sacht_nehab 2</i>
<i>ifir_sacht_nehab 3 e sacht_nehab 3</i>	<i>fir_dalai 1 e β-spline 1</i>
<i>fir_dalai 2 e β-spline 2</i>	<i>fir_dalai 3 e β-spline 3</i>
<i>ifir_omoms 2 e omoms 2</i>	<i>ifir_omoms 3 e omoms 3</i>

Exemplos

Exemplo

Filtro IFIR - Sacht Nehab Linear:

$$q = [0, \dots, 0, q_{-2}, q_{-1}, q_0, q_1, q_2, 0, \dots, 0],$$

em que $q_{-2} = q_2 = -0.00272602$, $q_{-1} = q_1 = 0.11566267$ e $q_0 = 0.77412669$.

Exemplo

Generate Function - Sacht Nehab 1: $r = x - k, x \in \text{Dom}(\tilde{f}), k \in \mathbb{Z}$:

```

if       $|r| < 1$   return  $\varphi(r) = 0.89538176 - 0.79076352 \cdot |r|$ 
else if  $|r| = 1$   return  $\varphi(r) = 0.05230912$ 
else          return  $\varphi(r) = 0.$ 

```

Aplicando os Filtros: FIR

Para aplicarmos a convolução nos filtros FIR, devemos analisar ponto a ponto do domínio, ou seja, para algum $T > 0$,

$$c_k = \sum_{i \in \mathbb{Z}} q_{i-k} \cdot f(T \cdot i).$$

Aplicando os Filtros: FIR

$$c_1 = \sum_{i \in \mathbb{Z}} q_{1-i} \cdot f(i) \quad \left\{ \begin{array}{l} 1-i \leq 2 \Leftrightarrow i \geq -1 \\ 1-i \geq -2 \Leftrightarrow i \leq 3 \end{array} \right.$$

$$c_1 = \sum_{i=-1}^3 q_{1-i} \cdot f(i) = q_2 \cdot f(-1) + q_1 \cdot f(0) + q_0 \cdot f(1) + q_{-1} \cdot f(2) + q_{-2} \cdot f(3)$$

⋮

$$c_N = \sum_{i \in \mathbb{Z}} q_{N-i} \cdot f(i) \quad \left\{ \begin{array}{l} N-i \leq 2 \Leftrightarrow i \geq N-2 \\ N-i \geq -2 \Leftrightarrow i \leq N+2 \end{array} \right.$$

$$c_N = \sum_{i=N-2}^{N+2} q_{N-i} \cdot f(i) = q_2 \cdot f(N-2) + q_1 \cdot f(N-1) + q_0 \cdot f(N) + q_{-1} \cdot f(N+1) + q_{-2} \cdot f(N+2).$$

Aplicando os Filtros: FIR

Portanto, temos a convolução $c = q * f$ representada no produto matriz-vetor a seguir:

$$c = q * f = \begin{bmatrix} q_0 & 2q_1 & 2q_2 & 0 & 0 & 0 & \dots & 0 \\ q_{-1} & q_0 & q_1 & 2q_2 & 0 & 0 & \dots & 0 \\ q_{-2} & q_{-1} & q_0 & q_1 & q_2 & 0 & \dots & 0 \\ 0 & q_{-2} & q_{-1} & q_0 & q_1 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & q_{-2} & q_{-1} & q_0 & q_1 & q_2 & 0 \\ 0 & \dots & 0 & q_{-2} & q_{-1} & q_0 & q_1 & q_2 \\ 0 & \dots & 0 & 0 & 2q_{-2} & q_{-1} & q_0 & q_1 \\ 0 & \dots & 0 & 0 & 0 & 2q_{-2} & 2q_{-1} & q_0 \end{bmatrix} \cdot \begin{bmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \\ \vdots \\ f(N-3) \\ f(N-2) \\ f(N-1) \\ f(N) \end{bmatrix}.$$

Aplicando os Geradores

Iremos construir o sinal $\tilde{f} : \mathbb{Z} \rightarrow \mathbb{C}$ aplicando a convolução mista entre o sinal filtrado c e a função geradora $\varphi : \mathbb{R} \rightarrow \mathbb{C}$, polinomial por partes, e imprime condições para $|x - k|$, na convolução

$$\tilde{f}(x) = \sum_{k \in \mathbb{Z}} c(k) \varphi(x - k).$$

Aplicando os Geradores

$$\tilde{f}(1,1) = \sum_{k \in \mathbb{Z}} c(k) \cdot \varphi(1,1-k) \quad \left\{ \begin{array}{l} 1,1-k \leq 2 \Leftrightarrow k \geq -0,9 \\ 1,1-k \geq -2 \Leftrightarrow k \leq 3,1 \end{array} \right.$$

$$\tilde{f}(1,1) = c(0) \cdot \varphi(1,1) + c(1) \cdot \varphi(0,1) + c(2) \cdot \varphi(-0,9) + c(3) \cdot \varphi(-1,9)$$

$$\vdots$$

$$\tilde{f}(N) = \sum_{k \in \mathbb{Z}} c(k) \cdot \varphi(N-k) \quad \left\{ \begin{array}{l} N-k \leq 2 \Leftrightarrow k \geq N-2 \\ N-k \geq -2 \Leftrightarrow k \leq N+2 \end{array} \right.$$

$$\tilde{f}(N) = c(N-2) \cdot \varphi(2) + c(N-1) \cdot \varphi(1) + c(N) \cdot \varphi(0) + c(N+1) \cdot \varphi(-1) + c(N+2) \cdot \varphi(-2).$$

Aplicando os Geradores

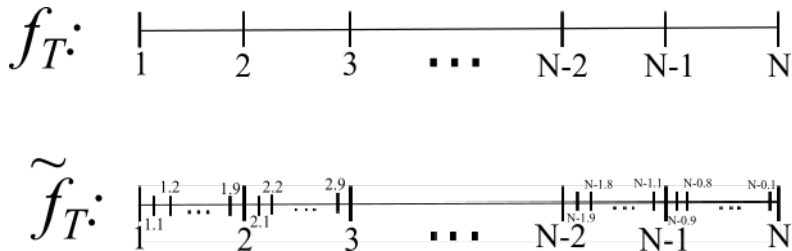


Figura: A figura nos indica que a função geradora irá, a partir dos pontos da f , construir no sinal \tilde{f} os pontos que fariam parte da amostragem original.

Resultados

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Resultados: Som reconstruído a 44100 Hz (usando sacht_nehab2)

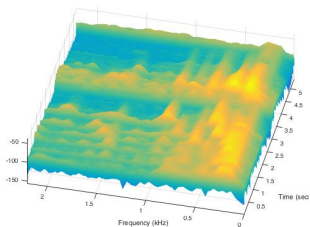


Figura: Som “mal amostrado” a 4410 Hz.

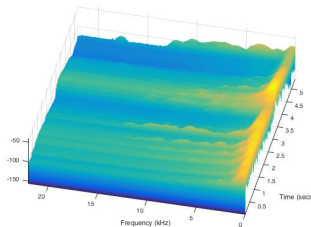


Figura: Som original a 44100 Hz.

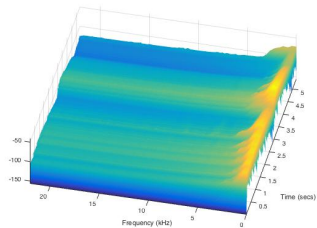


Figura: Som reconstruído com 44100 Hz (sacht_nehab2).

Banco de Sons

Para compararmos os métodos em diferentes tipos de som, construímos o seguinte banco de sons:

Sons baixados

- Áudio de fala masculina e feminina, obtidos no site *Free Sound* (<https://freesound.org/>);

Som Acústico

- Violão gravado tocando frequências graves e agudas;

Sons Sintéticos

- Um som de sino;

- Uma sitar;
- Uma nota solo de violoncelo;
- Uma clarineta;
- Um som de orquestra completa;
- Um som de bateria em ritmo.

PSNR (*Peak signal-to-noise ratio*)

Se f é o sinal original, \tilde{f} sua reconstrução e N o tamanho dos sinais, calculamos o *Mean Squared Error*

$$MSE = \frac{1}{N} \sum_{i=1}^{N-1} \left[f(i) - \tilde{f}(i) \right]^2,$$

e em seguida, calculamos o PSNR

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right).$$

Comparando os Erros

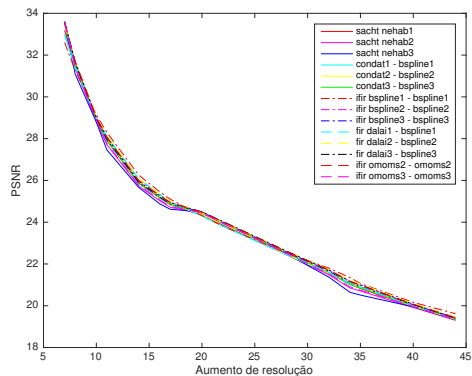
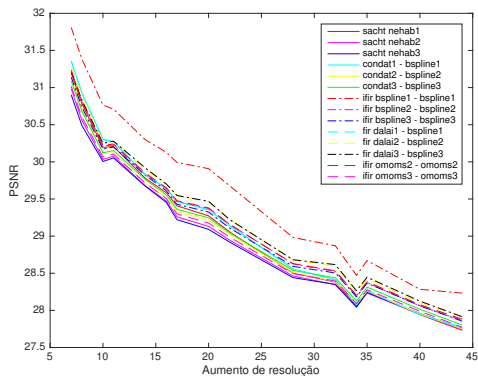


Figura: Gráfico PSNR de reconstrução do som de uma bateria e som de sinos.

Comparando os Erros

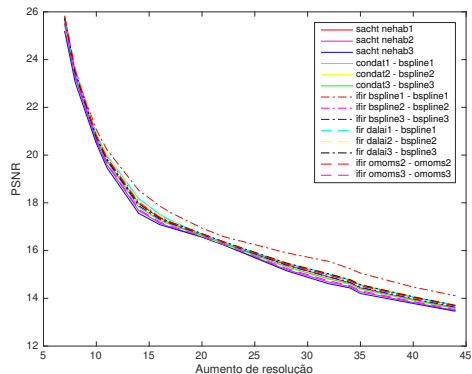
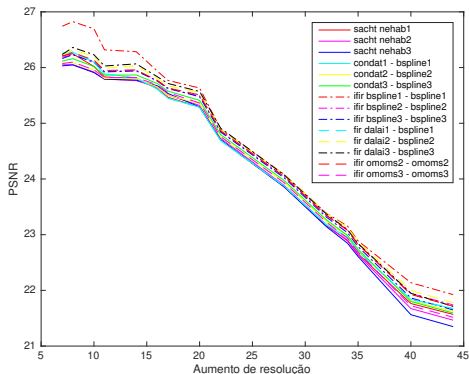


Figura: Gráfico PSNR de reconstrução do som de fala feminina e orquestra completa sintetizada.

Comparando os Erros

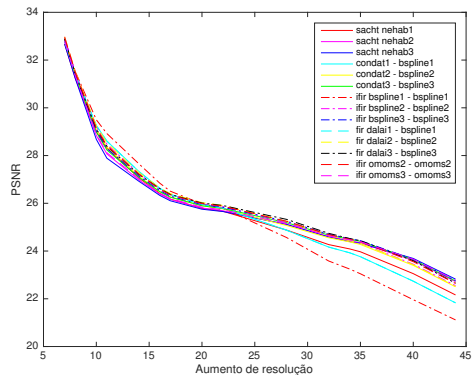
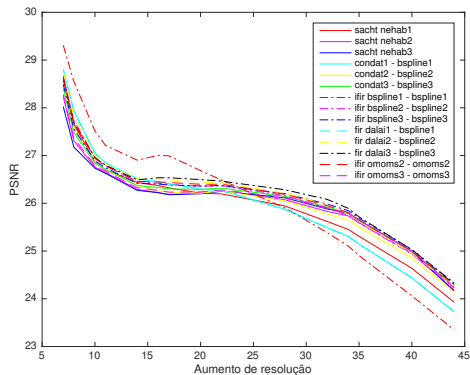


Figura: Gráfico PSNR de reconstrução do som de uma fala masculina e guitarra.

Comparando os Erros

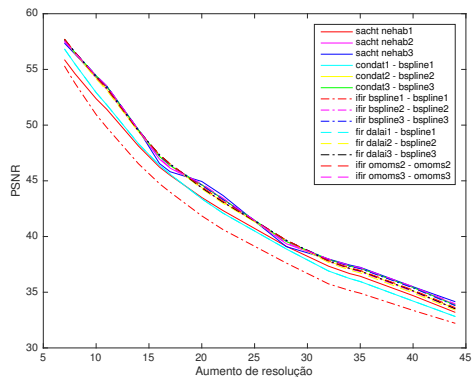


Figura: Gráfico PSNR de reconstrução do som de um violão.

Conclusão

- É possível usar os métodos de reconstrução de imagens para recuperar informações musicais;
- IFIR β spline1 – β spline1 apresentou-se como o melhor método de reconstrução de sons em diversos casos;
- FIR dalai3 – β spline3 apresenta-se, consistentemente, entre os melhores métodos para essas amostras de som.

Referências

Introdução/Motivação

Espaço de Funções

Reconstrução: Conceitos principais

Reconstrução: Admissibilidade

Filtros e Geradores

Resultados

Referências

Referências

BHALSHANKAR, S.; GULVE, A. K. Audio steganography: Lsb technique using a pyramid structure and range of bytes. arXiv preprint arXiv:1509.02630, 2015.

BLU, T.; UNSER, M. Approximation error for quasi-interpolators and (multi-) wavelet expansions. Applied and Computational Harmonic Analysis, Elsevier, v. 6, n. 2, p. 219–251, 1999.

BRUCKNER, A. M.; BRUCKNER, J. B.; THOMSON, B. S. Real analysis. [S.l.]: Classical Real Analysis, 1997.

FIGUEIREDO, D. G. d. Análise I. Rio de Janeiro: LTC, 1996.

FIGUEIREDO, D. G. d. Análise de Fourier e equações diferenciais parciais. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2014.

Referências (cont.)

GASQUET, C.; WITOMSKI, P. Analyse de Fourier et application: Filtrage, calcul numérique, ondelettes. [S.l.]: Paris: Masson, 1990.

GIANNAKOPOULOS, T.; PIKRAKIS, A. Introduction to Audio Analysis: A MATLAB Approach. [S.l.]: Academic Press, 2014.

MULLER, M. et al. Signal processing for music analysis. IEEE Journal of Selected Topics in Signal Processing, IEEE, v. 5, n. 6, p. 1088–1110, 2011.

OLIVEIRA, J. C. Cálculo IV. mar. 2017 a jun. 2017, 162 f. Notas de Aula, 2017.

Referências (cont.)

OSGOOD, B. The fourier transform and its applications. Lecture Notes for EE, v. 261, p. 20, 2009.

SACHT, L. K. Optimized Quasi-interpolators for Image Reconstruction and Consistent Volumetric Discretizations Inside Self-Intersecting Surfaces. Tese (Doutorado) — IMPA-Rio, 2014.

SHANNON, C. E. Communication theory of secrecy systems. Bell Labs Technical Journal, Wiley Online Library, v. 28, n. 4, p. 656–715, 1949.

Obrigado!

E-mail: lealdecastro@gmail.com

Acesse: **<http://mtm.ufsc.br/~leo/som>**