

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/272351208>

Componentes Mel Cepstrais

Technical Report · February 2015

DOI: 10.13140/2.1.1143.6327

CITATIONS

0

READS

254

Some of the authors of this publication are also working on these related projects:



Classificação em Classes de Sons Semelhantes [View project](#)



Doutorado [View project](#)

Componentes Mel Cepstrais

Adelino Pinheiro Silva*

2015, v-0.0

Resumo

O Cepstrum de um sinal foi uma das primeiras ferramentas para a realização da desconvolução de dois ou mais sinais que envolve também a predição linear, filtragem inversa. As variações básicas são o cepstrum de potência, o cepstrum complexo e o cepstrum real (CHILDERS; SKINNER; KEMERAIT, 1977).

Os Componentes Mel Cepstrais (*Mel Frequency Component Cepstrum* - MFCC) é o cálculo cepstral em intervalos do espectro separados por uma escala logaritmo melódica, denominada escala Mel. Este texto busca resumir (se possível) as bases para o cálculo dos Componentes Mel Cepstrais, e suas variações, aplicadas a sinais de tempo discretos.

Palavras-chaves: Componentes Mel Cepstrais, Cepstrum, MFCC.

Cepstrum de um sinal

O Cepstrum de um sinal é definido basicamente como uma transformação sobre o espectro do sinal, o que leva a duas operações em cadeia (CHILDERS; SKINNER; KEMERAIT, 1977). Devido a esta segunda transformação, os parâmetros obtidos também foram sugeridos por Bogert, Healy e Tukey (1963), introduzindo aos parâmetros análogos de acordo com o intercâmbio das sílabas, como mostra tabela 1 abaixo:

Tabela 1 – Tabela de nomenclatura dos parâmetros cepstrais.

Parâmetro Espectral		Parâmetros Cepstrais
Português	Inglês	
espectro	spectrum	cepstrum
frequência	frequency	quefreny
fase	phase	saphe
amplitude	amplitude	gamnitude
filtragem	filtering	liftering
harmônico	harmonic	rahmonic
período	period	repiod

Dadas estas definições vamos as definições de cepstrum de potência, cepstrum de fase e cepstrum complexo.

*adelinocpp@gmail.com

Cepstrum de Potência

O cepstrum de potência foi introduzido como uma técnica de obter o tempo de eco em um sinal composto (CHILDERS; SKINNER; KEMERAIT, 1977). Sendo assim, dado um sinal discreto no tempo $y[n]$ podemos definir seu cepstrum de potência em função da *quefrequency* q da forma:

$$Y[q] = \mathcal{C}\{y[n]\} = (\mathcal{F}^{-1}\{\log_{10}(|\mathcal{F}\{y[n]\}|^2)\})^2$$

Tem-se a transformada de Fourier inversa do logaritmo do módulo quadrático da transformada de Fourier¹ do sinal $y[n]$. No caso particular de dois sinais $x_1[n]$ e $x_2[n]$ ocuparem extensões de *quefrequency* diferentes e $y[n] = x_1[n] + x_2[n]$, temos o cepstrum de potência de $y[n]$ como:

$$\mathcal{C}\{y[n]\} = \mathcal{C}\{x_1[n]\} + \mathcal{C}\{x_2[n]\}$$

Cepstrum Complexo

O cepstrum complexo é definido como a transformada inversa de Fourier do logaritmo da transformada de Fourier de um sinal, da forma:

$$Y[q] = \mathcal{C}\{y[n]\} = \mathcal{F}^{-1}\{\log_{10}(\mathcal{F}\{y[n]\})\}$$

O cepstrum complexo é uma ferramenta que permite o desacoplamento de modelos fonte-filtro, pois, dado que $y[n]$ é a correlação de um sinal $x[n]$ com um filtro $h[n]$, tem-se a seguinte relação cepstral:

$$y[n] = x[n] * h[n]$$

$$Y[k] = X[k] \cdot H[k]$$

Onde $Y[k] = \mathcal{F}\{y[n]\}$, em seguida, tomando o logaritmo tem-se:

$$\log_{10}(Y[k]) = \log_{10}(X[k]) + \log_{10}(H[k])$$

Por fim, tomando a transformada inversa de Fourier:

$$\mathcal{C}\{y[n]\} = \mathcal{F}^{-1}\{\log_{10}(X[k])\} + \mathcal{F}^{-1}\{\log_{10}(H[k])\}$$

$$Y[q] = X[q] + H[q]$$

Cepstrum Real

O cepstrum real, definido por Oppenheim e Schaffer (2010), é uma adaptação do cepstrum complexo definido como a transformada inversa de Fourier do logaritmo do módulo da transformada de Fourier de um sinal, da forma:

$$Y[q] = \mathcal{C}\{y[n]\} = \mathcal{F}^{-1}\{\log_{10}(|\mathcal{F}\{y[n]\}|)\}$$

Assim como o cepstrum complexo, o cepstrum real permite o desacoplamento das amplitudes dos modelos fonte-filtro, da mesma forma com $y[n]$ como a correlação do sinal $x[n]$ com o filtro $h[n]$, tem-se a seguinte relação cepstral:

$$y[n] = x[n] * h[n]$$

$$Y[k] = X[k] \cdot H[k]$$

¹ Naturalmente trata-se da transformada discreta de Fourier definida da forma $X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}}$, onde k é a frequência discreta.

Em seguida, tomando o módulo e o logaritmo tem-se:

$$|Y[k]| = |X[k]| \cdot |H[k]|$$

$$\log_{10}(|Y[k]|) = \log_{10}(|X[k]|) + \log_{10}(|H[k]|)$$

Por fim, tomando a transformada inversa de Fourier:

$$\mathcal{C}\{y[n]\} = \mathcal{F}^{-1}\{\log_{10}(|X[k]|)\} + \mathcal{F}^{-1}\{\log_{10}(|H[k]|)\}$$

$$Y[q] = X[q] + H[q]$$

É importante salientar que o cepstrum real não permite a reconstrução do sinal pois a tomada do módulo da transformada de Fourier provoca a perda da informação de fase.

Outros Cálculos Cepstrais

Outros autores propuseram o cálculo cepstral de forma diferente, por exemplo [Brigham \(1988\)](#) define a transformada cepstral como a transformada de Fourier do logaritmo natural da transformada de Fourier de um sinal, da forma:

$$Y[q] = \mathcal{C}\{y[n]\} = \mathcal{F}\{\ln(\mathcal{F}\{y[n]\})\}$$

Nesta definição tem-se uma variação da variável independente q denominada por *quefrequency*.

Outra definição, o kepsrum, é oriundo da Série de Potências da Equação de Kolmogorov (*Kolmogorov equation power series* - KEPS, ou kepsrum) e foi proposta por [Silvia e Robinson \(1978\)](#), em sua própria notação, da seguinte forma:

$$\beta_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(|B(\omega)|) e^{j\omega k} d\omega \quad k = 0, \pm 1, \pm 2, \pm 3, \dots$$

Onde a série kepsrum é composta pela série β_k que é correspondente ao cepstrum real.

A escala Mel

A escala mel, com nome derivado da palavra melodia (*melody*), é uma escala logarítmica perceptual definida por ([STEVENS; VOLKMANN; NEWMAN, 1937](#)) que tem por objetivo manter os tons de frequência equidistantes tomando como referência 40 dB acima do limite de percepção humana em 1000 Hz. Se observarmos a anatomia da cóclea humana é possível constatar que a sensibilidade da membrana basilar às frequências segue uma escala logarítmica (vide figura 1).

Entretanto a transformação da escala linear de frequência para a escala mel não é única, sendo a fórmula mais popular definida por [O'Shaughnessy \(1987\)](#) pela equação a seguir e possui uma curva de relação de frequência linear e componente mel conforme a figura 2.

$$m(f) = 1127 \cdot \ln\left(1 + \frac{f}{700}\right)$$

Em consequência tem-se a equação para o cálculo em hertz da escala mel.

$$f(m) = 700 \cdot \left(e^{\frac{m}{1127}} - 1\right)$$

Desta forma os componentes mel cepstrais de um sinal são definidos a partir da soma da potência cepstral de uma faixa espectral dentro da escala mel. Em termos práticos para obter-se um componente mel-cepsral de um sinal é realizado o somatório da potência cepstral, no domínio da quefrequência, de um sinal filtrado em uma determinada faixa de acordo com a escala mel.

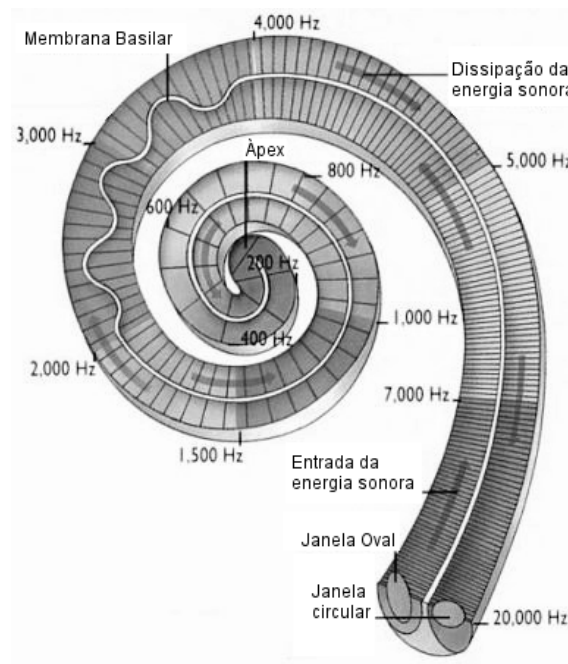


Figura 1 – Diagrama apresentando a sensibilidade de frequência da cóclea humana.

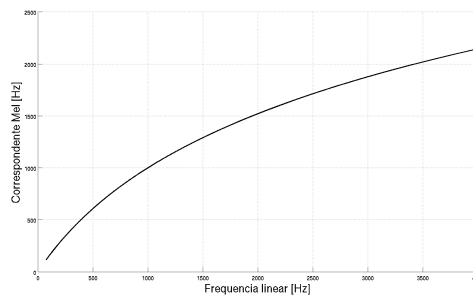


Figura 2 – Gráfico de relação entre frequência linear e componente mel

Calculo das componentes Mel-Cepstrais

A componente mel cepstral pode ser definida como a potencia cepstral de uma faixa de frequência na escala mel. Uma das abordagens mais difundidas é descrita por [Togneri e Pullella \(2011\)](#). Cada uma das C componentes mel cepstrais será calculada, em cada frame de N pontos, a partir de uma faixa espectral obtida pela aplicação de um filtro centralizado em uma frequência na escala mel (geralmente triangular) sobre o módulo da transformada de Fourier. Em seguida é calculado o logaritmo do espectro filtrado e a transformada discreta cosseno do tipo 2 (DCT-II - *Discrete Cosine Transform - type II*) ([RAO; YIP; RAO, 1990](#)) como apresentado na figura 3.

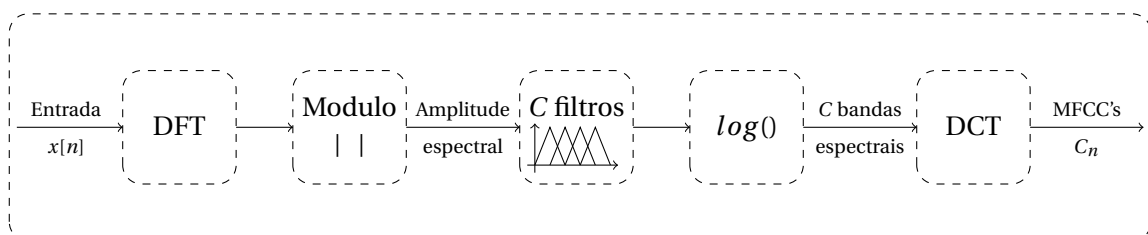


Figura 3 – Etapas do cálculo das componentes Mel Cepstrais

O banco de filtros em escala mel pode ser obtido de diferentes formas sendo as principais variações estão no número de filtros; no formato do filtro, triangular, quadrado ou de Schroeder (HERMANSKY, 1990); em suas posições, com ou sem sobreposição do espectro; com amplitude unitária ou área unitária. A abordagem de Cardoso (2009) sugere um banco de filtros de área unitária, sendo o primeiro terço dos filtros com largura constante até 1 kHz e os dois terços finais com largura logarítmica até o fim do espectro. A figura 4 apresenta algumas configurações de bancos de filtros que podem ser utilizados. O trabalho de Zheng, Zhang e Song (2001) discute sobre diferentes formas de obtenção dos MFCC's, explorando também a escala Bark (SMITH; ABEL, 1999), e concluem que o banco de filtros retangular com sobreposição e de mesma amplitude produz melhores resultados.

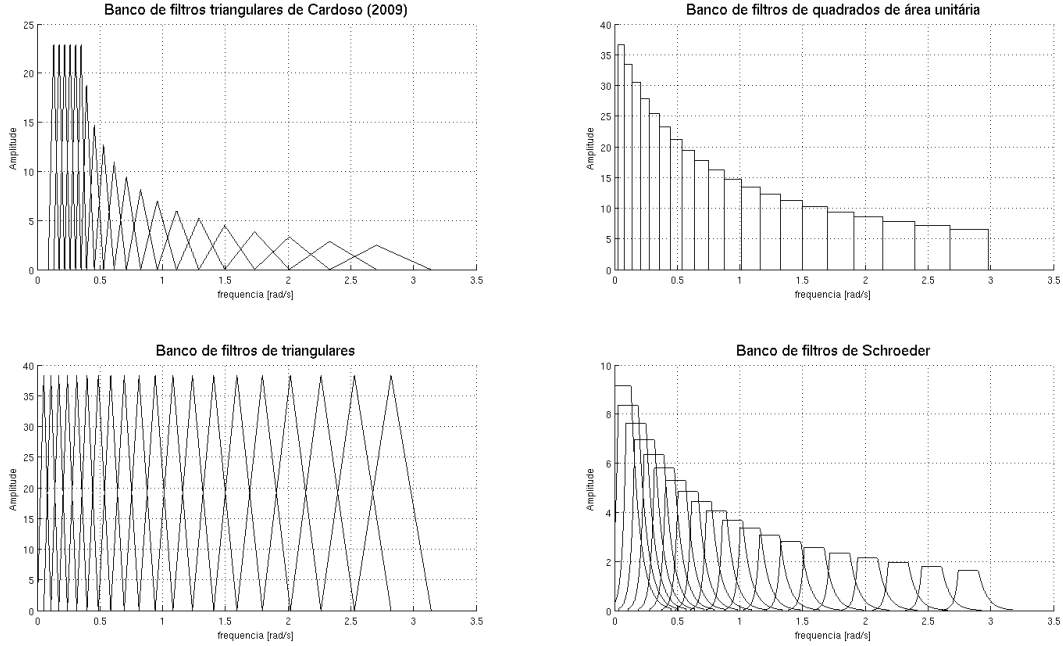


Figura 4 – Exemplos de Bancos de Filtros para obtenção de componentes mel cepstrais.

Tendo como $|X[k]|$ o módulo da transformada de Fourier do *frame* $x[n]$ calculado utilizando a janela espectral de Hanning² e $H_n[k]$ o n -ésimo dos C filtros na escala mel, obtém-se S_n como:

$$S_n = \sum_{k=0}^{\frac{N}{2}-1} |X[k]| \cdot H_n[k]$$

Assim, cada coeficientes mel cepstrais é obtido pela transformada discreta cosseno da forma:

$$c_n = \sum_{k=1}^K \log(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad \text{para } n = 1, 2, \dots, C.$$

Outras abordagens são sugeridas por Cuadros et al. (2007) para o cálculo da transformada discreta cosseno utilizando no somatório da equação anterior apenas o valor de k_i onde i é o índice do i -ésimo filtro.

² A janela de Hanning é definida como: $w[n] = 0.5 \left(1 - \cos \left(\frac{2\pi n}{N-1} \right) \right)$.

Considerações finais

Este pequeno resumo sobre componentes mel cepstrais não detalha e não aprofunda-se nas bases do MFCC, para tal uma abordagem mais profunda sobre a análise cepstral, tipos de escalas, diferentes tipos de bancos de filtros e também das diferentes abordagens para a transformada cosseno seriam necessárias para esgotar o assunto, isso sem citar uma vasta gama de diferentes características utilizadas para representar material acústico.

O objetivo deste texto foi cumprido apenas em trabalhar mais especificamente o algoritmo para computação de MFCC mais difundido e utilizado nos sistemas de automáticos de verificação de locutor.

Referências

- BOGERT, B. P.; HEALY, M. J.; TUKEY, J. W. The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In: CHAPTER. *Proceedings of the symposium on time series analysis*. [S.l.], 1963. v. 15, p. 209–243. Citado na página 1.
- BRIGHAM, E. O. *The Fast Fourier Transform and Applications*. [S.l.]: Englewood Cliffs, NJ, 1988. 448 p. Citado na página 3.
- CARDOSO, D. P. *Identificação de locutor usando modelos de mistura de gaussianas*. Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo, 2009. Citado na página 5.
- CHILDERS, D. G.; SKINNER, D. P.; KEMERAIT, R. C. The cepstrum: A guide to processing. *Proceedings of the IEEE*, IEEE, v. 65, n. 10, p. 1428–1443, 1977. Citado 2 vezes nas páginas 1 e 2.
- CUADROS, C. D. et al. Comparação entre as técnicas de mfcc e zcpa para reconhecimento robusto de locutor em ambientes ruidosos. *Rio de Janeiro, RJ*, 2007. Citado na página 5.
- HERMAN, H. Perceptual linear predictive (plp) analysis of speech. *the Journal of the Acoustical Society of America*, Acoustical Society of America, v. 87, n. 4, p. 1738–1752, 1990. Citado na página 5.
- OPPENHEIM, A. V.; SCHAFER, R. W. *Discrete-Time Signal Processing*. [S.l.]: Pearson, 2010. Citado na página 2.
- O'SHAUGHNESSY, D. *Speech communication: human and machine*. [S.l.]: Addison-Wesley Pub. Co., 1987. (Addison-Wesley series in electrical engineering). ISBN 9780201165203. Citado na página 3.
- RAO, K. R.; YIP, P.; RAO, K. R. *Discrete cosine transform: algorithms, advantages, applications*. [S.l.]: Academic press Boston, 1990. Citado na página 4.
- SILVA, M. T.; ROBINSON, E. A. Use of the kepstrum in signal analysis. *Geoprospection*, Elsevier, v. 16, n. 1, p. 55–73, 1978. Citado na página 3.
- SMITH, J. O.; ABEL, J. S. Bark and erb bilinear transforms. *Speech and Audio Processing, IEEE Transactions on*, IEEE, v. 7, n. 6, p. 697–708, 1999. Citado na página 5.
- STEVENS, S. S.; VOLKMAN, J.; NEWMAN, E. B. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 8, n. 3, p. 185–190, 1937. Citado na página 3.
- TOGNERI, R.; PULLELLA, D. An overview of speaker identification: Accuracy and robustness issues. *IEEE Circuits And Systems Magazine*, Second Quarter 2011. Citado na página 4.

ZHENG, F.; ZHANG, G.; SONG, Z. Comparison of different implementations of mfcc. *Journal of Computer Science and Technology*, Springer, v. 16, n. 6, p. 582–589, 2001. Citado na página [5](#).