Identificação da História de Compressão em sinais de Áudio nos Formatos WAV e MP3 utilizando o classificador Máquina de Vetor de Suporte

Rodrigo Cenachi Araujo[1], Flávia Magalhães Ferreira[1] e Zélia Myriam Peixoto[1]

[1] Programa de Pós Graduação em Engenharia Elétrica - PUC Minas
Belo Horizonte, Minas Gerais, 30535-901, Brasil
rodca1986@gmail.com, {flaviamagfreitas,assiszmp}@pucminas.br

Abstract. *Music on the Internet is available on MP3 audio compression format in high bit-rate. The double compression of MP3, achieved by decompressing and recompressing audio to a different compression ratio, is a typical manipulation of audio for malicious purposes. In this context, this research will approach the evaluation and identification of audio whose quality differs from the CD standard and evaluation of audio quality without any prior knowledge of the original audio. To summarize, this work aims to analyze and improve the existent methods of history compression on WAV and compressed MP3 format. Additionally, it was possible to achieve a detection rate of* ***99,9%*** *on the test compressed versus uncompressed audio.*

Resumo. *Músicas na Internet são disponibilizadas no formato de compressão de áudio MP3, em altas taxas de bits. A dupla compressão do MP3, a partir da descompressão e recompressão do áudio com diferentes taxas, é uma manipulação típica nos sinais digitais de áudio para propósitos ilícitos. Neste contexto, este trabalho tratará da avaliação e identificação da qualidade de áudio não correspondente ao áudio de CD e avaliação da qualidade de áudio sem nenhum conhecimento prévio do áudio original. Em resumo, o trabalho visa à análise e melhoria dos métodos já existentes da história de compressão nos formatos WAV e comprimido MP3. Obteve-se uma taxa de detecção de até* ***99,9%*** *no teste áudio comprimido versus não comprimido.*

1. Introdução

Algoritmos de compressão de áudio geram sinais de áudio com alta fidelidade e taxa de bits reduzida, para aplicações em armazenamento, transmissão em tempo real pela internet e radiodifusão [Thiagarajan e Spanias 2011].

Atualmente, há lojas de música *online* que fazem venda pela Internet. Frequentemente, essas músicas encontram-se no formato MPEG-1 *Audio Layer* 3 (MP3), em altas taxas de bits. Nesse caso, o custo de aquisição da música varia de acordo com a taxa de bits [Yang et. Al 2009] [Liu et. Al 2010].

Nos últimos anos, a pesquisa em multimídia forense começou a considerar conteúdos de áudio para investigar sua origem e autenticidade. Em particular, similarmente ao estudo de campo forense de imagem, a análise de artefatos devido à dupla compressão vem recebendo grande destaque [Bianchi et. Al 2014]. A dupla compressão do MP3, obtida pela descompressão e recompressão com diferentes taxas de bits, é uma manipulação típica do áudio para propósitos maliciosos [Qiao et. Al 2013]. Paralelamente, sinais de áudio também são muitas vezes armazenados em

formato WAV, sem nenhum conhecimento de operações de compressão anteriores [Luo et. Al 2014].

Esta pesquisa baseia-se, fundamentalmente, na estimação da qualidade do áudio, por meio da identificação da história de compressão do áudio. Em termos mais específicos, a empregabilidade e importância desse estudo refere-se à identificação de CDs com qualidade de áudio falsa e avaliação da qualidade do áudio sem nenhum conhecimento prévio do áudio [Luo et. Al 2014]. Nesse trabalho foi realizado a extração dos coeficientes MDCT e MFCC no áudio em formato WAV e utilizado o classificador SVM para a realização dos testes finais. A Seção 2 exemplifica com maiores detalhes essas técnicas.

2. Referencial teórico

2.1 Compressão de Dados

A compressão de dados é classificada em duas categorias principais: sem perdas (*lossless*) e com perdas (*lossy*). A compressão sem perdas produz a cópia exata do arquivo original depois de realizada a descompressão enquanto na com perdas o resultado pode ser praticamente indistinguível do original ou somente audível [Jacaba 2001].

Comprimir imagens e áudio através do formato sem perdas não é tão eficiente, uma vez que a informação nesse tipo de dados é redundante, o que justifica o emprego da compressão com perdas. Na aplicação de imagens, tem-se como exemplo o formato JPEG e, em áudio, a codificação MP3, WMA (*Windows Media Audio*) e AAC (*Advanced Audio Coding*). O formato MP3 é baseado, principalmente, na psicoacústica que considera o comportamento da percepção do ouvido humano [Jacaba 2001].

2.2 O MP3

Algoritmos MPEG-1/2 envolvem três camadas distintas para a compressão. A camada 1 forma o algoritmo de compressão mais básico (codificação de sub-bandas simples) enquanto as camadas 2 (banco de filtros com baixo atraso) e 3 (banco de filtros híbrido) são melhorias que usam alguns elementos da camada 1. Cada sucessiva camada melhora o desempenho de compressão, mas ao custo de uma complexidade maior do codificador e decodificador [Britanak 2011].

Essencialmente, a camada 3 do algoritmo MPEG-1/2, conhecido como padrão MP3, tornou-se a tecnologia chave para realizar a decodificação de áudio para várias plataformas: distribuição de música pela Internet, *players* de MP3 portáteis e sistemas multimídia. [Britanak 2011].

A arquitetura do codificador MP3, opera com *frames* que consistem de 1152 amostras de áudio. Cada *frame* é dividido em 2 *subframes* de 576 amostras, chamados grãos (*granules*) [Thiagarajan e Spanias 2011] [Jacaba 2001].

2.3 O Banco de Filtros Híbrido e a MDCT

O banco de filtros inclui segmentação adaptativa (bloco longo ou curto) e consiste de filtros de sub-banda seguidos pela MDCT (*Modified Discrete Cosine Transform*). O banco de filtros e a MDCT realizam a análise tempo-frequência com resolução adaptativa (dependente da análise psicoacústica humana), consistindo de filtros de 32

canais com largura de banda fixa, seguidos pela MDCT [Thiagarajan e Spanias 2011] [Jacaba 2001].

A saída do banco de filtros é processada usando a MDCT. Os dados são segmentados e processados com blocos de sobreposição de 50%. Na camada 3, existem dois tamanhos possíveis de blocos para a MDCT chamados, bloco curto (12 amostras) e bloco longo (36 amostras) [Thiagarajan e Spanias 2011] [Britanak 2011].

A MDCT é uma transformada com sobreposição que possui como saídas metade dos valores referentes ao número de entradas. Ela é baseada na Transformada Discreta de Cossenos (DCT – *Discrete Cosine Transform*) tipo IV, com 50% de sobreposição entre as janelas adjacentes de tempo. Desta forma, a transformada MDCT se estende através de dois blocos no tempo, eliminando os artefatos entre blocos. Apesar da sobreposição de 50%, a MDCT é amostrada criticamente e somente *M* amostras são geradas a cada *2M* amostras do bloco de entrada. Portanto são produzidos 18 componentes de frequência a cada 36 amostras no domínio do tempo, obtendo assim no formato MP3 um frame com 576 coeficientes de frequência [MDCT/IMDCT 2014] [Thiagarajan e Spanias 2011] [Jacaba 2001].

É possível observar, na Equação 1, a expressão matemática da MDCT *X(k)* de um sinal de entrada *x(n)*, no domínio do tempo, onde *h(n)* é a resposta ao impulso da janela (longa ou curta) e *M* número de sub-bandas.

$$X(k) = \sqrt{\frac{2}{M}} \sum_{n=0}^{N-1} x(n)h(n)\cos\left[\left(n + \frac{M+1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{M}\right]; onde, k = 0,1,\ldots,\frac{N}{2} - 1 \quad (1)$$

Os módulos MDCT empregam blocos curtos (melhor resolução de tempo) para transientes rápidos e blocos longos (melhor resolução de frequência) para sinais com variação lenta. Para evitar transições rápidas, janelas intermediárias, longa pra curta e curta pra longa, são fornecidas pelo padrão [Thiagarajan e Spanias 2011].

Em resumo, para obter os coeficientes MDCT do arquivo WAV a ser analisado os seguintes passos são realizados:

(1) Divisão em *frames* de 1152 amostras com 50% de sobreposição.
(2) Para cada *frame*, as amostras de áudio são separadas em 32 sub-bandas pelo banco de filtros de análise, adiante a janela MDCT divide cada uma das 32 sub-bandas em 18 sub-bandas (janela longa) ou 6 sub-bandas (janela curta). Portanto 18 coeficientes podem ser obtidos. É importante destacar que 3 janelas curtas serão combinadas juntas.
(3) Finalmente, um total de 576 (32x18=576) coeficientes MDCT para cada *frame* pode ser obtido.

As operações abordadas anteriormente são exatamente as mesmas no processamento da compressão MP3 antes da quantização dos coeficientes e da codificação [MP3Standard]. Na Figura 1, é possível observar exatamente esse ponto de extração dos coeficientes MDCT.

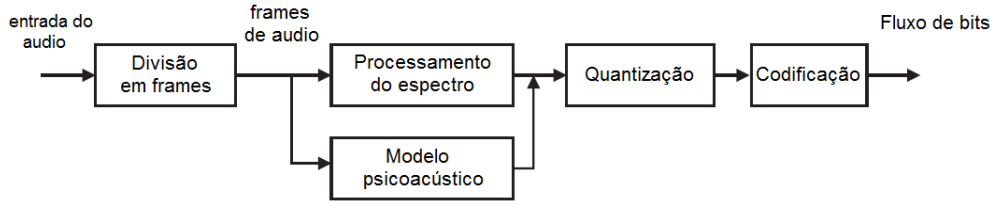

Figura 1 – Diagrama do esquema de compressão do formato MP3.

A extração desses coeficientes MDCT é utilizado como parte do trabalho que fará a realização do estudo da história de compressão do arquivo de áudio.

2.4 Mel-Frequency Cepstral Coefficient (MFCC)

O coeficiente cepstral na escala de frequência Mel (MFCC – *Mel Frequency Cepstral Coefficient*) é uma DCT de espectro modificado, no qual a frequência e amplitude são escaladas de forma logarítmica. A distorção de frequência é realizada de acordo com as bandas críticas da audição humana [Terasawa 2009].

Um banco de filtros de 32 canais, com espaço e largura de banda que assemelha aproximadamente ao sistema crítico de bandas auditivo faz a distorção da frequência linear [Terasawa 2009].

Aplica-se ao sinal de áudio, o banco de filtros com resposta em frequência triangular. Então a energia total em cada canal de frequência, Fi, é integrada para obter a saída do banco de filtros, no qual i é o número do canal no banco de filtros, Hi(f) é a resposta do filtro do canal i, e S(f) é o valor absoluto da transformada de Fourier do sinal. Observa-se esses termos na Equação 2 [Terasawa 2009] [Prahallad 2015].

$$F_i = \int |H_i(f) \cdot S(f)| df \quad (2)$$

O vetor de coeficientes MFCC, denominado Ci na Equação 4, é calculado tomando a transformada discreta de cossenos (DCT) da saída do banco de filtros na escala logarítmica, como mostra a Equação 3 [Terasawa 2009] [Prahallad 2015].

$$L_i = \log_{10}(F_i) \quad (3)$$

$$C_i = DCT(L_i) \quad (4)$$

Para essa pesquisa, o vetor MFCC é utilizado como composição de atributos para a construção de vetores com informações sobre a história de compressão do arquivo de áudio. Utilizou-se a biblioteca [VoiceBox] para a extração dos coeficientes MFCC do arquivo de áudio.

2.5 Aprendizado de Máquina (*Machine Learning*)

Na área de extração de coeficientes MDCT e MFCC de áudios é comum o uso dos classificadores *Linear Discriminant Analysis, Dynamic Evolving Neural-Fuzzy* e *Support Vector Machine* (Máquina de Vetor de Suporte). Justifica-se o uso do SVM por ter opções de *kernel* bem distintos e flexíveis, ser comumente utilizado e obter precisão condizente com os melhores classificadores do estado da arte [Ben-Hur et. Al 2010].

O SVM caracteriza-se como uma técnica de classificação binária que realiza a separação ótima entre duas classes distintas por meio de um hiperplano de separação [Vapnik 1995]. O algoritmo trabalha com dados linearmente separáveis. No entanto, existe a possibilidade de adaptação para conjuntos não lineares através das funções *kernel* não lineares. Por meio da função *kernel* RBF (*Radial Basis function*), é possível resolver problemas não linearmente separáveis, através da projeção do problema para um espaço de alta dimensão [Oliveira Junior 2010]. Segundo Hsu et. Al (2010), usuários do classificador SVM não necessitam conhecer toda a teoria por traz do algoritmo, mas sim algumas premissas básicas para realizar o procedimento da classificação.

Neste trabalho utilizou-se para a classificação do áudio, o *kernel* RBF. No emprego da classificação em várias classes (*multi-class classification*) foi utilizado a

aproximação *one-against-one*. A biblioteca SVM de Chang e Lin (2011) foi usada em todos os modelos de classificações.

3. Trabalhos Relacionados

Em Luo et. Al (2014) é investigado como identificar um arquivo de áudio descomprimido que foi comprimido anteriormente pelos esquemas de compressão MP3, WMA e AAC. O artigo Luo et. Al (2014) assemelha-se muito aos objetivos da pesquisa presente, utiliza-se da técnica de aprendizado de máquina SVM e é um dos trabalhos mais completos em relação à área de pesquisa. Pode-se enfatizar como diferença que o presente trabalho está focado nos formatos WAV original e MP3 e possui como destaque principal fazer a detecção de arquivos MP3 com taxa de bits de valor mais alto.

Outros trabalhos relacionados são Yang et. Al (2009), Liu et. Al (2010), Bianchi et. Al (2014) e Qiao et. Al (2013), que fazem a análise do áudio utilizando somente os coeficientes MDCT e aplicam técnicas estatísticas distintas. O trabalho de Yang et. Al (2009) utiliza como fator determinante um *threshold* dos coeficientes MDCT de baixo valor e não faz uso de classificadores. Em Liu et. Al (2010) é utilizado o valor absoluto dos coeficientes MDCT em cada banda de frequência e aplicado o SVM. Já em Bianchi et. Al (2014) é calculado a distância *chi-square* de histogramas de coeficientes MDCT e não faz uso de aprendizado de máquina. Em Qiao et. Al (2013) utiliza-se como parâmetro a distribuição de valores discretos dos coeficientes MDCT e realizado a comparação dos resultados em dois classificadores: DENFIS (*Dynamic Evolving Neural-Fuzzy)* e SVM.

4. Implementação e Validação

A extração dos coeficientes do áudio é realizada nesse trabalho sempre em um áudio em formato WAV. Na Figura 2, ilustra-se que mesmo para a análise de um arquivo MP3 deve-se descomprimi-lo para o formato WAV.

Figura 2 – Diagrama do método de análise do arquivo de áudio.

Para realizar a extração dos atributos MDCT do arquivo de áudio WAV, foi utilizado o codificador MP3 LAME [LAME MP3 Encoder] com os parâmetros padrões.

Conforme proposto por Luo et. Al (2014), após obter esses coeficientes o padrão estatístico especificado na Figura 3 foi realizado. Em resumo foi obtido um *frame* (576 coeficientes) que representa o valor médio absoluto de todos *frames*. Esse *frame* foi dividido em 24 caixas não sobrepostas e realizado o valor médio de cada pedaço. Obteve-se assim, (24 = 576 / 24) caixas sendo que as últimas 4 caixas foram descartadas pois são iguais a zero. Outra informação armazenada foi o valor médio de coeficientes MDCT iguais a zero por *frame*.

Para realizar esse trabalho e validá-lo foram coletados 2000 arquivos WAV aleatórios de CDs de áudio originais. A base de arquivos de áudio criada inclui 53 gêneros musicais diferentes distribuídos igualmente entre si. Os arquivos WAV foram convertidos para o formato mono e cada arquivo possui tempo total igual a 5 segundos.

A divisão da base ficou na proporção 50% (1000 arquivos para treinamento e 1000 para teste).

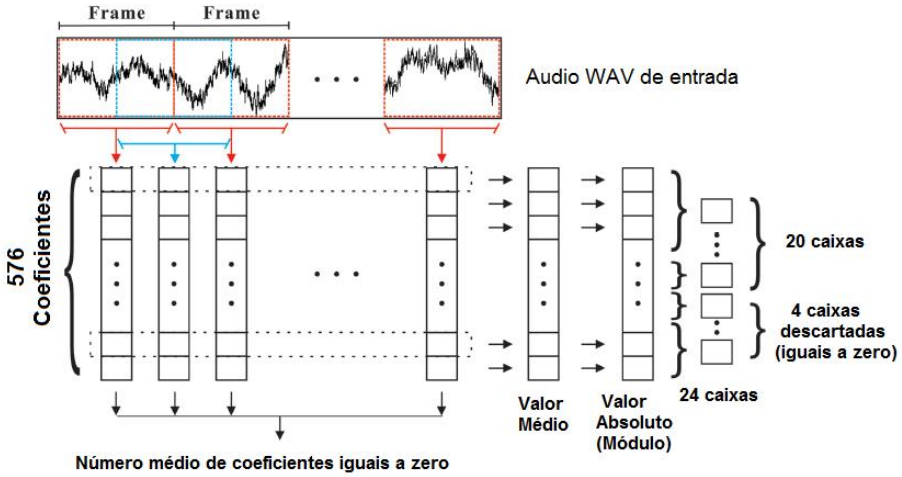

Figura 3 – Ilustração da extração do padrão estatístico aplicado aos coeficientes MDCT. Fonte: Adaptado de Luo et. Al (2014).

Após a criação da base foram realizadas as conversões para o formato de áudio MP3 em estéreo nas taxas de bits (32, 64, 96, 128, 192 e 256 kbps) e realizada a descompressão para WAV em formato mono.

Para cada arquivo de áudio WAV foi criado um vetor com 75 coeficientes (1 + 20 + 54 = 75). O valor 1, representa o número médio de coeficientes MDCT iguais a zero por *frame*. O 20 é representado pelas vinte caixas obtidas no *frame* de valor médio absoluto. O valor 54 representa os 18 coeficientes MFCC originais e suas respectivas primeira e segunda derivadas (18+18+18=54).

5. Resultados Computacionais

Na Figura 4 do lado esquerdo, é possível observar o gráfico *boxplot* com os resultados do valor médio de coeficientes MDCT iguais a zero por *frame* para o arquivo WAV original e os MP3 comprimidos nas taxas (32, 64, 96, 128, 192 e 256k). Pode-se perceber, que essa medição ajuda a diferenciar se um arquivo de áudio é um WAV original sem compressão, ou se sofreu compressão no formato MP3. Uma vez que essa detecção não é 100% confiável, os demais coeficientes MDCT e MFCC são necessários para realizar uma classificação com alta precisão.

No lado direito da Figura 4, observa-se o gráfico do vetor MDCT com os coeficientes de suas 20 caixas para um arquivo de áudio de gênero rock, conforme mencionado na seção 4. São demonstrados no gráfico os formatos WAV original, e MP3 32k, 64k e 128k. Essa diferença dos coeficientes MDCT em cada formato, será uma das bases para o classificador SVM diferenciar cada classe de áudio.

Na tabela 1, obtém-se a porcentagem de detecção obtida na identificação de arquivos de áudio WAV não comprimidos versus arquivos MP3 descomprimidos em uma taxa de bits fixa e aleatória. Pode-se observar que as taxas de detecção foram excelentes, chegando ao valor de até 99,9%.

Tabela 1 – Porcentagem de detecção obtida na identificação de arquivos de áudio WAV não comprimidos versus arquivos MP3 descomprimidos.

	32k	64k	96k	128k	192k	256k	Aleatório

MP3	99,9	99,9	99,85	99,85	99,55	99,1	99,65

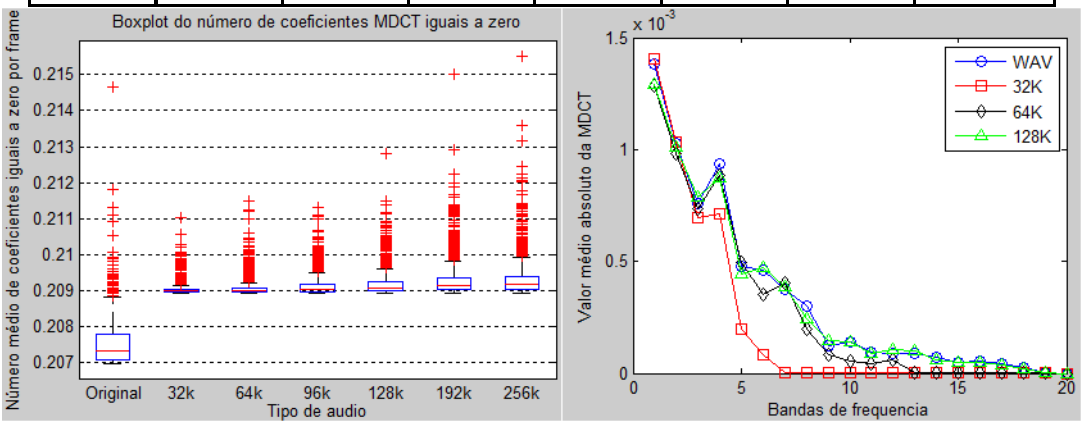

Figura 4 – Boxplot do valor médio de coeficientes MDCT iguais a zero por *frame* e gráfico do vetor MDCT de um arquivo de áudio com suas 20 caixas.

A tabela 2 apresenta em porcentagem como resultado, a matriz de confusão na identificação do áudio original e do áudio MP3 em diferentes taxas de bits. É possível observar que em taxas de bits maiores, obtém-se taxas menores de detecção. Justifica-se essa ocorrência, devido ao fato de que quanto maior a taxa de bits, menor o número de artefatos e mais difícil fazer detecção dessa classe.

Tabela 2 – Matriz de confusão em porcentagem da identificação do áudio MP3 em diferentes taxas de bits.

	Original	32k	64k	96k	128k	192k	256k
Original	98,30	0,00	0,00	0,00	0,10	0,10	1,50
32k	0,00	99,90	0,10	0,00	0,00	0,00	0,00
64k	0,00	0,30	99,30	0,10	0,10	0,00	0,20
96k	0,00	0,10	0,90	94,90	3,80	0,20	0,10
128k	0,00	0,10	0,50	8,40	82,00	7,00	2,00
192k	0,00	0,00	0,30	0,80	8,90	72,40	17,60
256k	0,20	0,10	0,50	0,60	2,10	18,70	77,80

6. Conclusão

Nesse trabalho foi proposta a identificação da história de compressão do arquivo de áudio em formato WAV, além de potenciais aplicações dessa pesquisa. O método utilizado, estatísticas dos coeficientes MFCC e MDCT aplicados em um classificador SVM, é semelhante ao proposto em [Luo et. Al 2014]. As contribuições principais dessa pesquisa constitui a realização da detecção do áudio MP3 em taxas de bits maiores (192k e 256k) e a aplicação da técnica em uma base de dados de áudio com gama de variação maior (53 gêneros musicais).

Pode-se constatar que no teste áudio comprimido versus não comprimido obtém-se uma taxa de detecção acima de 99% mesmo para um áudio MP3 comprimido em altas taxas de bits. No teste da identificação se o áudio pertence à classe original ou em qual taxa de bits ele foi comprimido, obtém-se taxas de detecção acima de 72,4% em todos os casos. Como trabalho futuro, deseja-se aplicar ou criar um método de escolha de coeficientes MDCT e MFCC, que possa selecionar melhor esses coeficientes, a fim de obter taxas de detecção acima de 90% em todas as taxas de compressão do formato MP3.

Referências Bibliográficas

Thiagarajan, J. , Spanias, A. Analysis of the MPEG-1 Layer III (MP3) Algorithm Using MATLAB, vol.3, pp.1-129 , Morgan & Claypool Publishers, November 2011

Yang, R., Shi Y., and Huang, J. Defeating fake-quality MP3. In Proceedings of the 11th ACM workshop on Multimedia and security ACM, New York, NY, 2009

Liu, Q., Sung, A., Qiao, M., Detection of Double MP3 Compression, Cognitive Computation, Dec 2010

Bianchi T., Rosa A., Fontani M., Rocciolo G., and Piva A. Detection and classification of double compressed MP3 audio tracks. EURASIP, 2014.

Qiao M.; Sung, A.H.; Liu Q., Improved detection of MP3 double compression using content-independent features, ICSPCC, 2013 IEEE, Aug. 2013

Luo, Da; Luo Weiqi; Yang Rui and Huang Jiwu. Identifying Compression History of Wave Audio and Its Applications. ACM Trans. Multimedia Comput. 2014

Jacaba, Joebert S., Audio compression using Modified Discrete Cosine Transform: the MP3 coding standard, Research paper, University of the Philippines Diliman, Oct 2001

Britanak, Vladimir, A survey of efficient MDCT implementations in MP3 audio coding standard: Retrospective and state-of-the-art, ISSN 0165-1684, April 2011

MDCT/IMDCT- Properties and Applications, Discrete Transforms and their Applications, Department of Electrical Engineering, the University of Texas, 2014

MP3Standard. Information technology - coding of moving pictures and associated audio for digital storage media up to about 1.5 mbit/s.

Terasawa Hiroko, A hybrid model for timbre perception: quantitative representations of sound color and density, a Dissertation submitted to the Department of Music, Stanford University, December 2009

Prahallad Kishore, Speech Technology: A Practical Introduction Topic: Spectrogram, Cepstrum and Mel-Frequency Analysis, Carnegie Mellon University & International Institute of Information Technology Hyderabad, 2015

Voicebox. http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

Ben-Hur, A., and Weston, J. A user's guide to support vector machines data mining techniques for the life sciences. Methods in molecular biology (Clifton, N.J.), 2010

Vapnik, V.N. The Nature of Statistical Learning Theory. 2. ed. New York: Springer-Verlag, 332p, 1995

Oliveira Junior, G. M. Máquinas de Vetores Suporte: Estudo e Análise de Parâmetros para Otimização de Resultado. Universidade Federal de Pernambuco, 2010.

Hsu, C.W.; Chang, C.C; Lin, C.J. A Practical Guide to Support Vector Classification, National Taiwan University (Technical report), 2010.

Chang C.-C. and Lin C.-J. LIBSVM: A library for support vector machines. ACMTrans. Intell. Syst. Technol. 2, 27:1–27:27, 2011.

Lame MP3 Encoder. Disponível em http://sourceforge.net/projects/lame/