

Codificação LPC: análise e re-síntese do sinal. Manipulação intonacional do sinal de erro artificial

*5º Trabalho Laboratorial de Áudio Computacional, DEEC-FEUP

André de Azevedo Barata
Dep. de Engenharia Eletrotécnica e de Computadores
Faculdade de Engenharia da Universidade do Porto
Porto, Portugal
up201907705@edu.fe.up.pt

André Nogueira Soares
Dep. de Engenharia Eletrotécnica e de Computadores
Faculdade de Engenharia da Universidade do Porto
Porto, Portugal
up201905318@edu.fe.up.pt

Abstract—Este relatório explora a aplicação dos coeficientes LPC, analisando a quantidade necessária para obter um sinal de qualidade e a influência da pré-ênfase. São abordadas também as abordagens com sinais de ruído originais e sintéticos. Cada experimento é detalhadamente apresentado, acompanhado dos resultados obtidos e das conclusões decorrentes.

Index Terms—Coeficientes LPC, Predição Linear, Pré-Ênfase

I. INTRODUÇÃO

Neste 5º Trabalho Laboratorial de Análise Computacional, focaremos na Codificação LPC para análise e re-síntese de sinais de áudio usando o ambiente Matlab. Os objetivos passam por explorar a influência da ordem do modelo LPC e do tipo de segmento do sinal de áudio na codificação LPC deslizante, avaliar o sinal de erro, e re-sintetizar o sinal com o erro original e posteriormente re-sintetizar um excerto do sinal utilizando um sinal de erro artificial.

II. FUNDAMENTOS TEÓRICOS

Este capítulo visa estabelecer os conceitos fundamentais necessários à compreensão do método LPC (Linear Predictive Coding) e seus elementos constituintes.

Coeficientes LPC:

Os coeficientes LPC representam a base do processo de predição linear. Esses coeficientes são derivados a partir de uma análise de predição que visa modelar as características espectrais do sinal de áudio. A ordem do modelo LPC (p) desempenha um papel crucial na precisão da representação preditiva.

Resíduos - Sinal de Erro:

Os resíduos, ou sinal de erro, no contexto LPC, representam as discrepâncias entre a predição linear e o sinal original. Esses resíduos são essenciais para entender as variações não capturadas pelo modelo preditivo. Analisar o sinal de erro é crucial para avaliar a eficácia da predição e refinamento do modelo LPC.

Predição Linear:

A predição linear é um método que visa antecipar a amostra atual de um sinal como uma combinação linear das amostras passadas. Neste contexto, essa abordagem é aplicada para modelar as propriedades espectrais do sinal, permitindo uma representação compacta e eficiente.

Papel da Pré-ênfase:

A pré-ênfase é uma técnica que envolve a aplicação de um filtro passa-alto ao sinal de áudio antes da análise LPC. Essa etapa tem como objetivo realçar as componentes de alta frequência, melhorando assim a eficiência do processo de predição linear.

Importância da Ordem do Modelo LPC:

A ordem do modelo LPC (representada por " p ") é um fator determinante na qualidade da predição linear. Valores adequados de " p " são essenciais para equilibrar a complexidade do modelo e sua capacidade de representar as características espectrais do sinal. Uma ordem muito baixa pode resultar em submodelagem, enquanto uma ordem muito alta pode levar à sobreajuste.

Introdução de Sinal de Erro Artificial Monótono:

A incorporação de um sinal de erro artificial monótono é uma estratégia projetada para simular cenários específicos, adaptando a característica do sinal de erro. Em vez de empregar o sinal de erro gerado convencionalmente, torna-se viável adotar impulsos com frequência f_0 para sons vocais, ruído gaussiano para sons não vocais e uma combinação de impulsos e ruído gaussiano para sons mistos.

Conclusões:

Ao explorar esses fundamentos teóricos, torna-se possível compreender a base conceitual do método LPC, preparando o terreno para a implementação prática e análise dos resultados no contexto do trabalho laboratorial.

III. METODOLOGIA

Este capítulo explora detalhadamente a metodologia aplicada para atingir os objetivos propostos no 5º Trabalho Laboratorial, centrado na Codificação LPC.

Análise:

O principal objetivo desta fase consiste no cálculo dos Coeficientes LPC. Inicialmente, foi adotado o valor de 12 para o número de coeficientes ($p=12$). Contudo, para analisar quantos coeficientes são necessários para obter a máxima performance em sons voiced e unvoiced, foram utilizados diferentes valores de p .

Para calcular os coeficientes LPC de forma deslizante, foram criadas janelas de duração fixa (Hamming, 30 ms, sobreposição 2/3), com reamostragem prévia do sinal para $f_s=16\text{kHz}$. Para além disso, foram realizadas análises tanto com o sinal original como com pré-ênfase, utilizando $\alpha = 0.98$ no segundo caso.

A determinação do sinal de erro e dos coeficientes LPC foi realizada através do método LPC. Após a obtenção dos valores, foram conduzidas análises subjetivas e objetivas do Sinal de Erro, incluindo uma avaliação objetiva considerando timbre e periodicidade.

Re-Síntese:

Para a re-síntese, foram empregados dois modos principais: o uso do sinal de erro original, tanto com pré-ênfase como sem pré-ênfase, e a geração de um sinal de erro artificial. Após a re-síntese, uma análise comparativa entre a forma de onda original e a re-síntese foi conduzida.

Conclusões:

Este conjunto de etapas compõe a metodologia adotada para realizar a análise e re-síntese dos sinais de áudio, destacando as particularidades do método LPC e a influência da pré-ênfase nos resultados obtidos. No próximo capítulo, os resultados serão apresentados e discutidos, evidenciando as conclusões extraídas deste processo metodológico.

IV. RESULTS

Vamos subdividir esta secção com base no trabalho realizado e nas análises efetuadas. Para as subsecções A, B e C, foi empregado o script denominado *tp5_analise_resintese_erro_original.m*. Já para a subsecção D, foram utilizados os scripts *tp5_frameCompression.m* e *tp5_plotFrames.m*.

A. Análise do Sinal e Re-Síntese sem Pré-Ênfase

No processo de análise sem pré-ênfase, o primeiro passo consistiu na leitura e re-amostragem do sinal de áudio do arquivo 'ASRF20.wav', ajustando sua taxa de amostragem para 16 kHz. Em seguida, foram determinados os parâmetros para a análise LPC, incluindo a ordem $p=12$, a duração do

segmento de análise de 30 ms com sobreposição de 2/3 e a aplicação da janela de Hamming.

O sinal de áudio foi dividido em frames e, para cada segmento, os coeficientes LPC foram calculados. O sinal de erro foi obtido pela diferença entre o sinal original e o sinal predito pelos coeficientes LPC. O sinal re-sintetizado sem pré-ênfase foi gerado aplicando-se novamente os coeficientes LPC ao sinal de erro. Os resultados estão representados na Figura 2.

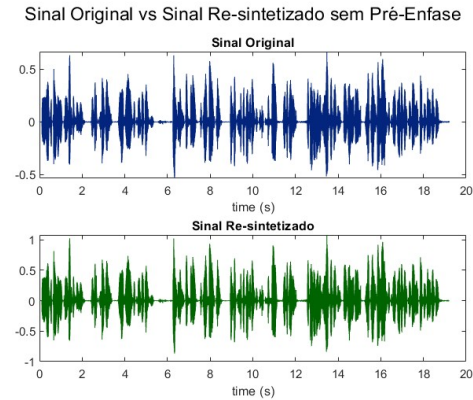


Fig. 1: Original vs Re-sintetizado sem Pré-Ênfase

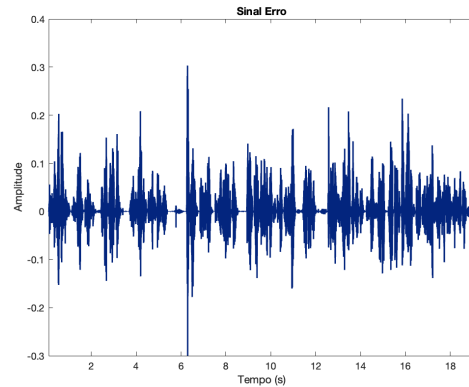


Fig. 2: Sinal Erro

O sinal reconstruído é bastante similar ao original, contudo, audivelmente é perceptível a diferença (apesar de ligeira) entre ambos quando são usados os coeficiente e os resíduos para reconstruir o sinal original. Eventualmente, utilizando um filtro de ordem maior, esta diferença poderia ser reduzida. Para além disso, é preciso ter em consideração que o LPC faz uma modulação aproximada que tem em conta as amostras passadas e pode originar pequenas discrepâncias entre o sinal original e o reconstruído.

Relativamente ao sinal reconstruído com apenas os resíduos, este sim apresenta diversas distinções relativamente ao original, parecendo um som mais abafado e com perda de algum carácter vozeado. Na verdade, quanto mais complexo for o sinal, maior é a perda de informação dos resíduos.

Para além disso, foi gerado um ficheiro áudio com o sinal re-sintetizado sem pré-ênfase (*Sinal_Re_sintetizado_sem_Pré_Enfase.wav*) e outro com o sinal de erro (*Sinal_Erro.wav*).

B. Análise do Sinal e Re-Síntese com Pré-Ênfase

No procedimento de análise com pré-ênfase, inicialmente, o sinal de áudio foi lido e re-amostrado para uma taxa de amostragem de 16 kHz. Posteriormente, os parâmetros para a análise LPC foram configurados, incluindo uma ordem $p=12$, uma duração de segmento de 30 ms com sobreposição de 2/3 e a aplicação da janela de Hamming. Além disso, foi introduzida a pré-ênfase no sinal original com um coeficiente de ênfase $\alpha = 0.98$.

O sinal de áudio pré-enfatizado foi dividido em frames, e para cada um, foram calculados coeficientes LPC. O sinal de erro foi obtido pela diferença entre o sinal original e o sinal predito pelos coeficientes LPC. O sinal re-sintetizado com pré-ênfase foi gerado aplicando-se novamente os coeficientes LPC ao sinal de erro. Os resultados encontram-se na Figura 3.

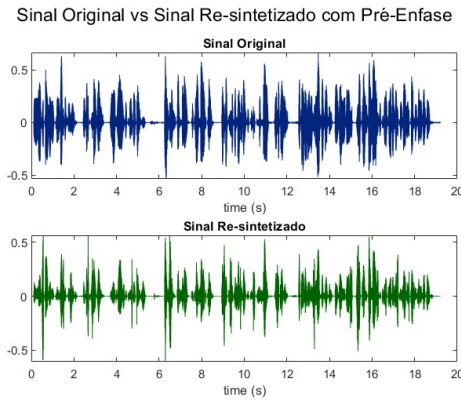


Fig. 3: Sinal Original vs Sinal Re-sintetizado com Pré-Ênfase

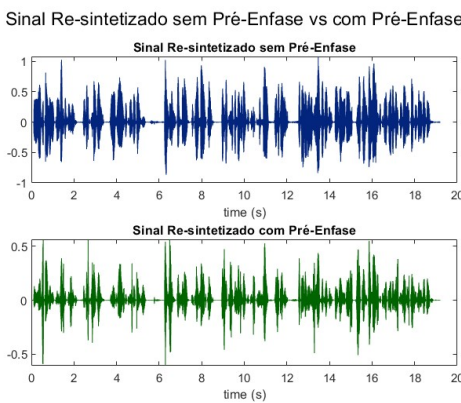


Fig. 4: Re-sintetizado sem Pré-Ênfase vs com Pré-Ênfase

A diferença quando é aplicado um filtro pré-ênfase é clara e resulta na perda de qualidade do sinal. Isto porque o

filtro remove as baixas frequências, o que retira caráter à voz do locutor que é uma voz grave, geralmente associado a frequências mais baixas, originando uma mudança visível no timbre. Poderia ser útil caso fosse necessário remover conteúdos de baixas frequências. Esta diferença pode ser observada visualmente na Figura 4, visto que o sinal após o filtro pré-ênfase tem menor amplitude e as transições deixam de ser tão suaves, especialmente em zona vozeadas.

Adicionalmente, foi gerado ficheiro de áudio representando o sinal re-sintetizado com pré-ênfase (*Sinal_Re_sintetizado_com_Pré_Enfase.wav*).

C. Análise do Sinal com variação de coeficientes LPC (p)

Conforme discutido nas aulas teóricas, foi constatado que o número de Coeficientes de Predição Linear (LPCs) necessários para alcançar a máxima performance é menor em sons não-vozeados (unvoiced) do que em sons vozeados (voiced), conforme evidenciado na Figura 5.

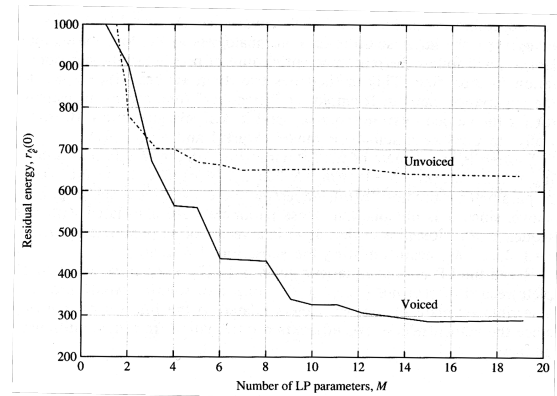


Fig. 5: Comparação do número de LPCs necessários para Vozeado e Não-Vozeado

Para analisar esse efeito, foram selecionados dois pequenos excertos do sinal, um vozeado e outro não-vozeado. De seguida, o valor de p foi variado de 1 a 24 de modo a mostrar que a voz não vozeada tem melhor resultados com menos coeficiente que a vozeada. A análise dos dados foi realizada em segmentos muito curtos, sendo 17.30s a 17.733s para o vozeado e 17.72s a 17.74s para o não-vozeado. Os resultados obtidos estão ilustrados na Figura 6.

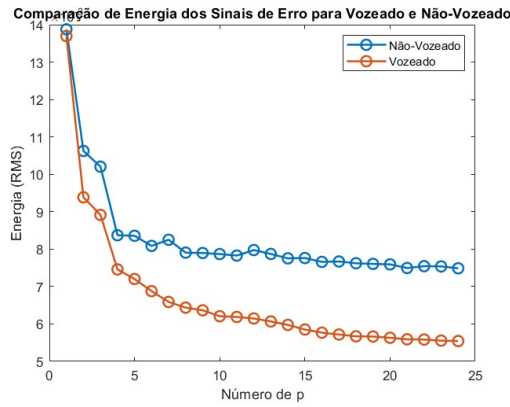


Fig. 6: Comparação entre diferentes valores de p para Vozeado e Não-Vozeado

Como evidenciado, para sons não-vozeados, o número de LPCs necessário para alcançar a máxima performance é inferior ao número necessário para sons vozeados.

D. Análise do Sinal e Re-Síntese com sinal de erro artificial monótono

Este método consiste em analisar o carácter da voz (se é vozeado, não vozeado ou misto) e substituir estes segmentos por sinais pré definidos. Por exemplo, neste contexto, pegamos num excerto do áudio original e substituímos os segmentos vozeados por um sinal de impulsos unitários à frequência fundamental do segmento, segmentos misto foram substituídos da mesma forma que os vozeados, mas com a adição de ruído branco e os segmentos não vozeados foram substituídos por ruído branco. Isto permite comprimir o sinal passando apenas a informação da caracterização de cada frame, por exemplo: frame 776 -z ruído branco com média 0.2 e desvio padrão de 0.1. Em sistemas atuais, são guardados características de cerca de 60 sinais e é possível caracterizar uma grande diversidade de ficheiros áudios apenas com a informação das características dos sinais guardados.

No projeto desenvolvido, foi escolhido um excerto entre os 17,30 e 17,85 segundos, devido à sua riqueza em termos fonéticos. Neste intervalos o locutor diz o seguinte: "Que os homens", no entanto o locutor juntar as palavras resultando na seguinte construção fonético "cuzomens". Deste modo, podemos caracterizar esta expressão em termos fonéticos, sendo "Cu" vozeado, "z" misto, "omen" vozeado e "s" não vozeado. A reconstrução utilizando os coeficientes e resíduos do filtro LPC encontra-se na figura 7

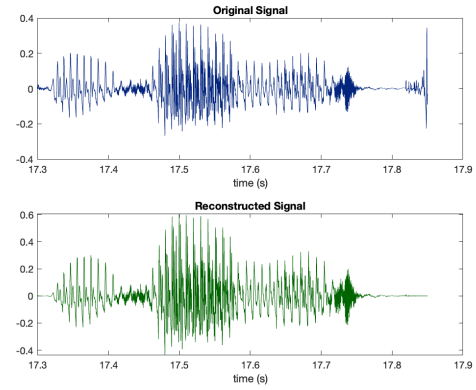


Fig. 7: Reconstrução do excerto usando resíduos

Na Tabela I estão representados os tipos de fonética e timestamps dos frames a contabilizar, assim como a frequência fundamental, quando aplicável, ou seja, quando se trata de um frame misto ou vozeado. A figura 8 representa os frames do excerto escolhido para uma melhor observação dos frames com segmentos com fonéticas diferentes. Para obter informação de f0, foi utilizado o software PRAAT para obter uma lista dos valores da frequência fundamental do excerto, e os valores foram mapeados para cada frame e feita a média de f0 de cada frame, que são os valores apresentados na tabela. As janela utilizadas nesta avaliação são de 100 ms, o que não é ideal, mas facilita bastante a análise dos frames, visto que reduz o número de frames a analisar em 33%. Caso este aumento não fosse feito teriam de ser analisado $14 * 3,3 \approx 47$ frames, o que seria uma tarefa temporalmente exigente.

Frame	Tipo	Substituição	F0 (Hz)
1 - 17.300s a 17.400s	Voiced	Impulsos	103.46
2 - 17.333s a 17.433s	Voiced	Impulsos	100.78
3 - 17.366s a 17.466s	Mix	Impulsos + Ruído	99.72
4 - 17.400s a 17.500s	Mix	Impulsos + Ruído	101.46
5 - 17.433s a 17.533s	Voiced	Impulsos	102.69
6 - 17.466s a 17.566s	Voiced	Impulsos	100.00
7 - 17.500s a 17.600s	Voiced	Impulsos	96.00
8 - 17.533s a 17.633s	Voiced	Impulsos	93.28
9 - 17.566s a 17.666s	Voiced	Impulsos	91.58
10 - 17.600s a 17.700s	Voiced	Impulsos	90.73
11 - 17.633s a 17.733s	Voiced	Impulsos	89.90
12 - 17.666s a 17.766s	Unvoiced	Ruído	-
13 - 17.700s a 17.800s	Unvoiced	Ruído	-
14 - 17.733s a 17.833s	Unvoiced	Ruído	-

TABLE I: Análise por frames e classificação tendo em conta voiced, unvoiced e mix

V. CONCLUSÕES

Este trabalho abordou a aplicação prática dos coeficientes LPC na análise de sinais de áudio. Destacamos a influência da ordem do modelo LPC, a importância da pré-ênfase e exploramos a utilização de sinais de erro, tanto originais quanto artificiais.

Os resultados revelaram que a pré-ênfase pode afetar a qualidade do sinal, e a escolha adequada da ordem do modelo LPC é crucial. A introdução de um sinal de erro artificial mostrou-se promissora para simular cenários específicos, apesar de possíveis degradações na qualidade ao utilizar frames mais longos.

Em resumo, este trabalho consolidou os nossos conhecimentos teóricos sobre a aplicação dos coeficientes LPC, contribuindo para a compreensão dos desafios e possibilidades na análise e ressíntese de sinais de áudio.

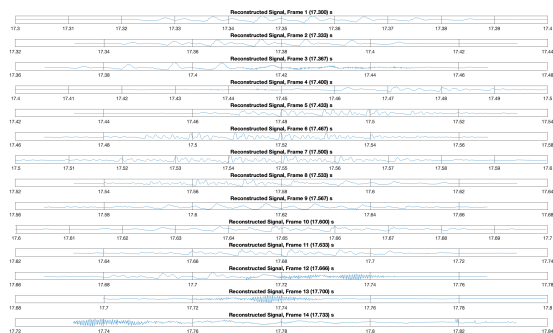


Fig. 8: Frames utilizados na codificação do excerto

Assim sendo, para cada frame, foram obtidos os coeficientes da mesma utilizando a função LPC do MATLAB, no entanto o resíduo foi uma das três funções mencionadas anteriormente (impulso periódicos, ruído branco ou impulso periódicos + ruído branco). O sinal resultante está representado na figura 9. Esta figura releva as limitações desta metodologia, na medida em que o sinal obtido é bastante distinto do sinal original. Contudo, a implantação aplicada é bastante rudimentar, já que apenas consiste num banco de 3 sinais, sem nenhuma alteração na amplitude nem espectral.

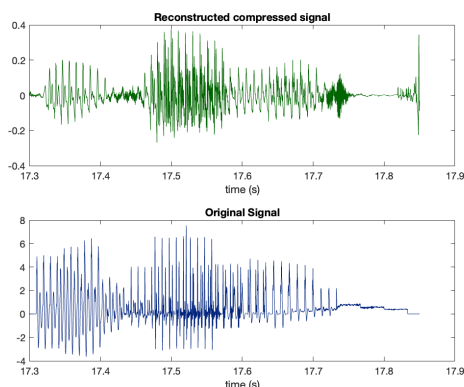


Fig. 9: Reconstrução do sinal usando erro artificial

Auditivamente, percebe-se que o som gerado com o sinal de erro artificial apresenta características significativamente distintas, é possível identificar que reproduz o mesmo excerto. A degradação na qualidade do som pode dever-se a diversos fatores, sendo que consideramos crucial destacar a utilização de frames de 100 ms em vez dos tradicionais 30 ms frequentemente empregados para sinais de voz, o que pode ter contribuído para essa degradação.

Foram gerados dois ficheiros de áudio para se conseguir ouvir o sinal original segmentado e o sinal de áudio produzido utilizando o sinal de erro sintético. O nome dos ficheiros é *Sinal_Re_sintetizado_com_Ruído_Sintético.wav* e *Sinal_Original_para_comparar_com_Ruído_Sintético.wav*.