



UNIVERSIDADE FEDERAL DO PARANÁ

CENTRO POLITÉCNICO – CURITIBA

SETOR DE TECNOLOGIA

DISCIPLINA: PROCESSAMENTO DIGITAL DE SINAIS – TE072

PROFESSOR: MARCELO ROSA

CURSO DE ENGENHARIA ELÉTRICA

REGULARIZED LINEAR PREDICTION OF SPEECH

NOME: KLEITON CHOCHI ZEMBOVICI

Nº: GRR20044805

CURITIBA

2008

SUMÁRIO

SUMÁRIO.....	2
1. Introdução.....	3
1.1 Propriedades da Voz.....	3
2. Linear Prediction.....	4
3. Teoria do LP Regularizado.....	5
3.1 Linear Prediction.....	5
3.2 Regularização.....	5
4. Seleção de λ	6
4.1. λ Constante	9
4.2. λ Adaptativo.....	10
5. Validação dos Resultados.....	11
6. Conclusão	13
7. Referências Bibliográficas.....	14

1. Introdução

Linear Prediction (LP, ou Predição Linear em português) é uma forma de codificar a voz digitalmente, já o *Regularized Linear Prediction of Speech*, é uma variação desse método. Para entender como funcionam essas técnicas, é necessário antes conhecer alguns aspectos da propriedade da voz.

1.1 Propriedades da Voz

O aparelho fonador humano produz diversos sons, considerando diferenças perceptuais e espectrais, eles podem ser classificados em duas categorias: vozeados e não-vozeados.

O som vozeado é formado quando o fluxo de ar dos pulmões pressiona as cordas vocais, forçando-as a abrir e fechar, gerando uma excitação periódica, ou muito próxima à periódica. A frequência com que ocorre a abertura e o fechamento das cordas é que determina a frequência fundamental da excitação. Essa pode assumir valores que variam de 50 a 400 Hz dependendo do tamanho das cordas vocais. Já a frequência fundamental percebida pelo nosso cérebro é denominada *pitch*.

Como o *pitch* e o formato do trato vocal se alteram ao longo do tempo, não se pode dizer que o som vozeado veja verdadeiramente periódico. Suas características espectrais e estatísticas variam ao longo do tempo. Porém, pode-se dizer que esse sinal é aproximadamente estacionário quando considerado um trecho de 10 a 30 ms. Dessa forma, para explorar essa característica do sinal da fala, a análise do mesmo deve ser feita em intervalos periódicos de pequena duração.

Os sons não-vozeados, chamados fricativos, não são gerados pela vibração das cordas vocais e, portanto, não apresentam a mesma periodicidade encontrada na estrutura dos sinais vozeados. As fricativas são formadas quando o ar dos pulmões é forçado através das cordas vocais abertas e de uma “brecha” do trato vocal, gerando um som semelhante ao ruído. No domínio do tempo, esses sons perdem periodicidade e seu espectro de potência é aproximadamente “achatado”, não apresentando tão claramente os picos de ressonância encontrados em sinais vozeados.

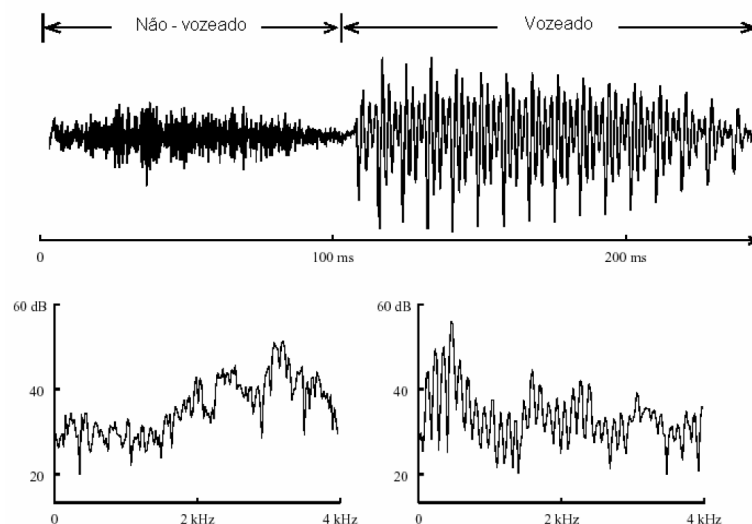


FIGURA 1 – Exemplos de sons não-vozeados e vozeados e seus respectivos espectros de potência.

2. Linear Prediction

Essa técnica de codificação consegue operar a taxas baixas quando comparado ao codificador PCM, enquanto preserva a qualidade do sinal através da exploração de redundâncias do sinal de voz e de limitações perceptuais do ouvido humano. As redundâncias ocorrem devido a:

- em geral o espectro do sinal da voz se altera relativamente devagar permanecendo estável ao longo de intervalos de 10 a 30 ms;
- sucessivos períodos de *pitch* são, geralmente, similares em trechos vozeados;
- a envoltória espectral é relativamente suave, com a maior parte da energia concentrada em baixas frequências.

Já as limitações perceptuais estão relacionadas ao fato de o ouvido humano ser, nos sinais de fala, insensível à fase e mais sensível a baixas frequências do que as altas.

A redundância do sinal de voz leva à conclusão de que suas amostras são correlacionadas. A envoltória espectral corresponde às correlações de curto-termo (*short-term correlations*) e a estrutura de harmônicos corresponde às correlações de longo-termo (*long-term correlations*). Essas correlações podem ser exploradas através da técnica LP (*Linear Prediction*) para resultar em um codificador com baixa taxa de bits.

LP é geralmente usado para estimar parâmetros do modelo auto-regressivo (AR – *autoregressive*) que descreve a envoltória espectral do sinal. Um problema conhecido quando se usa as aproximações do LP, é que o resultado da envoltória espectral pode ter vários picos bem agudos, pois o método possui uma deficiência em separar os longos-termos dos curtos-termos, contaminando o espectro com harmônicas, em consequência distorcendo a voz original.

O modelo AR é também conhecido como filtro IIR ou ainda filtro só-polos, é definido pela equação abaixo:

$$x_t = \sum_{i=1}^N a_i \cdot x_{t-i} + \varepsilon_t$$

Onde a_i é o coeficiente auto-regressivo, x_t a série investigada, N a ordem do filtro e ε geralmente é um ruído branco.

O problema da envoltória espectral não ser suave, pode ser evitado de duas maneiras. A primeira é usando um outro método para estimar o espectro. A segunda é utilizando uma modificação heurística do LP, o qual inclui usar um filtro AR de baixa ordem, ou usar uma janela atrasada para a seqüência de auto-correlação, ou ainda, expandir a largura de banda através de alteração dos pólos do modelo AR. Porém nenhum desses métodos é excelente, pois não resolvem o problema das harmônicas indesejadas.

Utilizar uma janela atrasada (*lag window*) para a seqüência de auto-correlação, consiste em multiplicar a seqüência por $w_{lag}(m)$ para obter valores de auto-correlação modificados $r'(m) = w_{lag}(m) \cdot r(m)$. Esta aproximação é usada nos seguintes codificadores: ITU-T G.729, 3GPP AMR de banda estreita e 3GPP AMR banda larga. Onde todos utilizam uma janela exponencial.

Expandir a largura de banda pode ser realizado através de alteração dos pólos do modelo AR por um fator $\gamma < 1 \approx e^{-\pi f_0 / f_s}$. O coeficiente do LP é multiplicado por γ ($a'_i = \gamma^i \cdot a_i$). Esse método é usado nos seguintes codificadores: CELP (*Code Excited Linear Prediction*) no MPEG-4 que utiliza $\gamma = 0.9902$ para 8 kHz e $\gamma = 0.9883$ para 16 kHz de taxa de amostragem, o qual corresponde a uma expansão 25 e 60 Hz respectivamente, outro exemplo é o ITU-T G728 q usa $\gamma = 253/256$ em 8kHz de frequência de amostragem para obter 30 Hz de expansão de largura de banda.

A regularização da suavização da envoltória pode ser realizada por um modelo de derivada de primeira e segunda ordem, o qual controla o *tradeoff* entre a envoltória originada pelo processo de LP e o modelo suavizado. Porém a desvantagem desse processo é a alta complexidade quando comparado ao método de LP convencional.

3. Teoria do LP Regularizado

A principal idéia é criar um parâmetro λ que regularize a suavização da envoltória espectral, sem que aumente a complexidade de cálculo computacional, fazendo com que o resultado seja melhor que o método de LP convencional.

No LP os coeficientes do modelo AR são escolhidos para minimizar os erros de predição do sinal, esses erros vão ser chamados de D , enquanto as variações de erros gerado pelo LP regularizado irá ser R . A composição desses erros é dado por $D + \lambda R$, onde se λ tender a 0, o modelo volta a ser o LP convencional.

3.1 Linear Prediction

Para um modelo AR $1/A(z)$ de ordem M , onde $A(z) = 1 + a_1z^{-1} + \dots + a_Mz^{-M}$, a envoltória espectral é definido por:

$$S(\omega, a) = \frac{1}{A(e^{j\omega})^2} \quad (1)$$

Os coeficientes do modelo AR é representado por $a = [a_1, a_2, \dots, a_M]^T$. A partir desses coeficientes é determinado a minimização da variância dos erros no sinal gerado pelo LP, $e(n) = \sum_{k=0}^M a_k s(n-k)$, onde $s(n)$ são as amostras da voz. Logo D será:

$$D = E[e^2(n)] = a^T R a + 2a^T r + r_0 \quad (2)$$

R é a seguinte matriz:

$$\mathbf{R} = \begin{bmatrix} r_0 & r_1 & \dots & r_{M-1} \\ r_1 & r_0 & & r_{M-2} \\ & & \dots & \\ r_{M-1} & r_{M-2} & \dots & r_0 \end{bmatrix} \quad (3)$$

$$r = [r_1, r_2, \dots, r_M]^T \quad (4)$$

e $r_k = E[s(n)s(n-k)]$ são os valores auto-correlacionados com o sinal da voz. Minimizando (2) através dos coeficientes do modelo AR, chega-se a a_{opt} :

$$a_{opt} = -\mathbf{R}^{-1}r \quad (5)$$

que pode ser resolvido computacionalmente pelo algoritmo de Levinson-Durbin.

3.2 Regularização

Uma rápida mudança na envoltória espectral, faz com que a mesma seja deteriorada; essa deterioração pode ser medida pela seguinte expressão sugerida:

$$R(s(\omega, a)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{d}{d\omega} \log S(\omega, a) \right]^2 d\omega \quad (6)$$

Porém essa expressão é difícil de ser implementada em um método não-iterativo para encontrar a_{opt} ; então aproxima-se a mesma por:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{d}{d\omega} \log A(e^{j\omega}) \right|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A'(e^{j\omega})}{A(e^{j\omega})} \right|^2 d\omega \quad (7)$$

onde $A'(e^{j\omega})$ é a derivada em ω . O denominador de (7) ainda previne uma solução não-iterativa, e mais uma modificação na expressão é realizado:

$$\hat{R}(s(\omega, a)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A'(e^{j\omega})}{W(\omega)} \right|^2 d\omega \quad (8)$$

onde $1/|W(\omega)|^2$ serve como uma aproximação grosseira da envoltória espectral de $1/A(e^{j\omega})^2$. Desde que $A'(e^{j\omega}) = -\sum_{k=0}^M jka_k e^{jk\omega}$, a expressão pode ser escrita como:

$$\sum_{k=0}^M \sum_{l=0}^M ka_k la_l \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{-j\omega(k-l)}}{|W(\omega)|^2} d\omega \quad (9)$$

a integração de (9) é o inverso da transformada discreta no tempo de Fourier de $1/|W(\omega)|^2$, logo:

$$\hat{R}(s(\omega, a)) = \sum_{k=0}^M \sum_{l=0}^M ka_k la_l f(l-k) \quad (10)$$

$$f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{j\omega x}}{|W(\omega)|^2} d\omega \quad (11)$$

Sendo $1/|W(\omega)|^2$ a envoltória espectral e $f(m) = r(m)v(m)$ a seqüência de auto-correlação do sinal da fala com uma janela $v(m)$, isto é, $1/|W(\omega)|^2$ é uma convolução da potência espectral da fala, descrita pela seqüência $r(m)$ de auto-correlação de comprimento M, com uma janela $V(\omega)$.

A equação (10) pode ser expressa por:

$$\hat{R}(s(\omega, a)) = \mathbf{a}^T \mathbf{D} \mathbf{F} \mathbf{D} \mathbf{a} \quad (12)$$

onde \mathbf{D} é a diagonal da matriz em que cada elemento da diagonal consiste no número da linha, e \mathbf{F} é a matriz Toeplitz (ou diagonal-constante) de auto-covariância para a janela da seqüência de auto-correlação $f(m)$.

Voltando a expressão de composição dos erros $(D + \lambda R)$, porém usando agora o \hat{R} , temos:

$$D + \lambda \hat{R} = \mathbf{a}^T (\mathbf{R} + \mathbf{D} \mathbf{F} \mathbf{D}) \mathbf{a} + 2\mathbf{a}^T \mathbf{r} + r_0 \quad (13)$$

e chega-se a seguinte solução:

$$a_{opt} = -(\mathbf{R} + \lambda \mathbf{D} \mathbf{F} \mathbf{D})^{-1} \mathbf{r} \quad (14)$$

que pode ser resolvida pelo algoritmo de decomposição de Cholesky.

4. Seleção de λ

Basicamente existem dois tipos de parâmetros de regularização de LP:

- 1) λ constante: nesse caso, λ é mantido sempre constante. A implementação e uso do mesmo é simples e praticamente não requer maior esforço computacional quando comparado ao LP convencional. A única preocupação é achar o melhor valor de λ .

- 2) λ *adaptativo*: o parâmetro λ varia a cada *frame*, sendo escolhido da melhor forma possível para se adaptar à condição momentânea.

Para verificar as diferenças entre o método LP convencional e suas variações, vários experimentos foram realizados. Como referência, foi utilizado exemplos de sons extraídos de “*Acoustic-phonetic continuous speech corpus*” (DARPA-TIMIT, NIST Speech Disc 1-1.1). Amostrados em 16 kHz, o experimento usará uma voz feminina de 9354 *frames* e uma voz masculina de 7498 *frames*, e para validação, foi utilizada uma outra voz feminina de 3929 *frames* e outra masculina de 4430 *frames*. As amostras foram processadas por uma janela tipo Hamming de 20ms antes das análises. O desempenho foi analisado para os modelos de ordem 10, 16, 22 e 28, e os *frames* amostrados tanto em 8 quanto a 16 kHz.

O desempenho do método de estimação de envoltória espectral pode ser medido de várias maneiras. Um critério importante para avaliar é medir a qualidade da distorção espectral nas harmônicas entre a potência espectral em cada *frame* e a envoltória espectral do sinal estimado, isto pode ser calculado por:

$$SD = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} [10 \log S_{\text{int}}(\omega) - 10 \log S(\omega, a)]^2 d\omega} \quad (15)$$

onde $S_{\text{int}}(\omega)$ é a envoltória de referência e $S(\omega, a)$ é o modelo AR estimado.

Os diferentes métodos analisados são encontrados na tabela abaixo.

Constantes	LP	Método de <i>Linear Prediction</i> convencional (referência).
	R _C	LP Regularizado com λ constante, otimizado para baixa distorção.
	BE _C	LP com largura de banda expandida (<i>Bandwidth-expanded</i>) com γ constante, otimizado para baixa distorção
Adaptativos	R _A	LP Regularizado com λ adaptativo, sendo escolhido por função baseado na frequência de <i>pitch</i> .
	R _{opt}	LP Regularizado com λ mudando <i>frame a frame</i> , sendo escolhido para a menor distorção possível para cada <i>frame</i> em particular.
	BE _{opt}	LP com largura de banda expandida (<i>Bandwidth-expanded</i>) com γ mudando <i>frame a frame</i> , sendo escolhido para a menor distorção possível para cada <i>frame</i> em particular.

TABELA 1 – Relação dos métodos utilizados para as análises.

Utilizando uma envoltória com as amostras da voz feminina, foi possível realizar uma rápida comparação o LP convencional de ordem 22 e LP regularizado com uma constante de $\lambda = 3,28 \times 10^{-3}$.

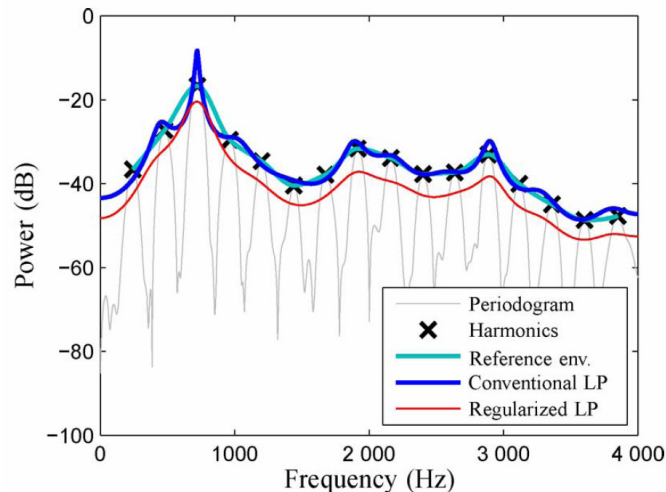


GRÁFICO 1 – Comparação da envoltória de referência e os métodos de LP convencional e regularizado.

É possível analisar que no gráfico 1 a principal diferença entre os métodos de predição de linear é a suavidade da curva vermelha (LP regularizado), quando comparado a curva azul (LP convencional) que apresenta muitos picos espectrais.

Para verificar que a envoltória de referência usada é uma boa aproximação do real, através do método de LP com largura de banda expandida, aplicada a cada *frame* de uma fala real, foi gerado um conjunto de modelos AR com envoltória suavizada. Utilizando trens de pulso para uma frequência de *pitch* particular, foi criada uma fala sintética a partir da qual, através de um processo AR conhecido, a envoltória espectral resultante foi considerada como uma envoltória de fala verdadeira. Convertendo a fala sintética para o domínio do tempo, foi possível utilizar os diferentes métodos de LP para estimar uma nova envoltória espectral, então é calculada a distorção espectral entre o modelo AR constante e o variável. Esse processo foi repetido para diferentes frequências de *pitch* (de 100 a 450 Hz, com variação de 10 Hz a cada repetição). Calculando a distorção espectral média de todos os métodos de LP, considerando as diferentes ordens, frequências de *pitch*, amostragem e gênero (voz masculina e feminina), é possível gerar o gráfico 2.

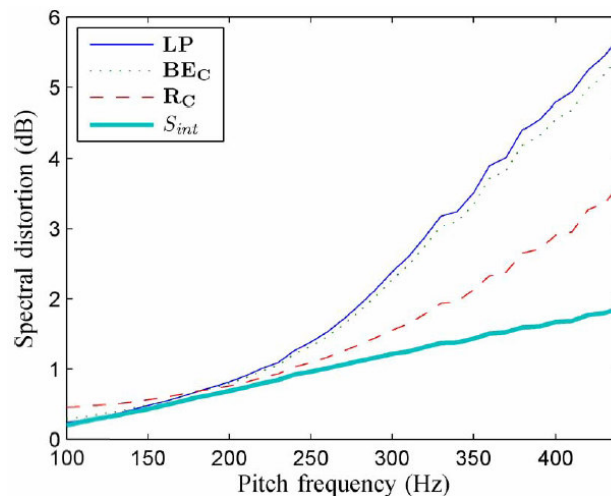


GRÁFICO 2 – Comparação da distorção espectral média entre os diferentes métodos de LP. Onde a curva em verde (S_{int}) representa a interpolação (*cubic spline interpolation*) entre as harmônicas. É o melhor método, porém impraticável.

O gráfico 2 deixa claro que os métodos de LP com largura de banda expandida e também o LP regularizado geram menos distorção espectral do que o método convencional. Porém ainda falta descobrir um λ ótimo.

4.1. λ Constante

Uma simulação foi realizada utilizando a voz feminina de 9354 *frames* como exemplo, e a cada teste, tanto a constante λ do método de LP regularizado, como a constante γ do método de LP com largura de banda expandida, são alterados e verificados a envoltória espectral resultante, bem como sua distorção. É possível verificar as curvas obtidas através do gráfico 3.

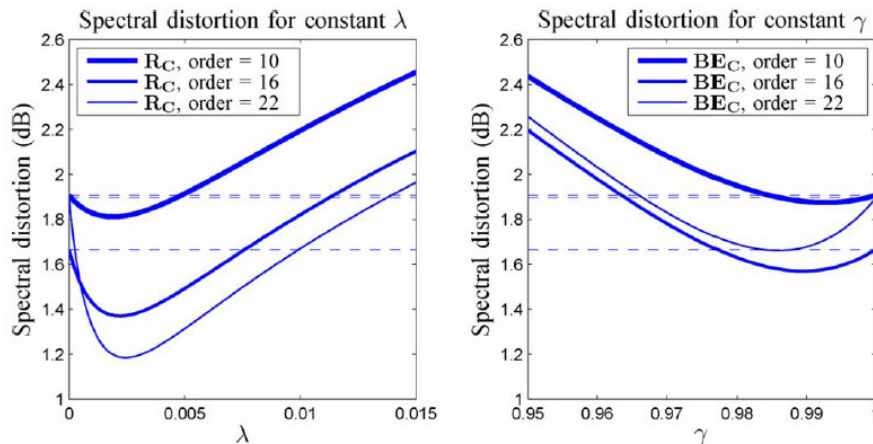


GRÁFICO 3 – Distorção espectral média sobre todos os 9354 *frames* amostrados a 8 kHz. Foram utilizados os modelos de ordem 10, 16 e 22. Vale ressaltar que quando $\lambda \rightarrow 0$ ou quando $\gamma \rightarrow 1$, a variação do método LP é nula, ou seja, retorna ao método de LP convencional.

Observando o gráfico 3, fica fácil perceber qual das constantes utilizadas oferece menor distorção espectral para esse exemplo específico, também é possível verificar que quanto maior a ordem do modelo utilizado, melhor o resultado final, e por último, que o modelo de LP regularizado é melhor que o modelo com largura de banda expandida por apresentar menor distorção.

Do mesmo modo que foi realizada a simulação para a voz feminina de 9354 *frames*, também foi feito para os diferentes tipos de amostragens, gênero (masculino feminino) e ordens. O resultado obtido das constantes que apresentaram menor distorção espectral é mostrado no gráfico 4, enquanto a diferença em dB entre os métodos é mostrado na tabela 2.

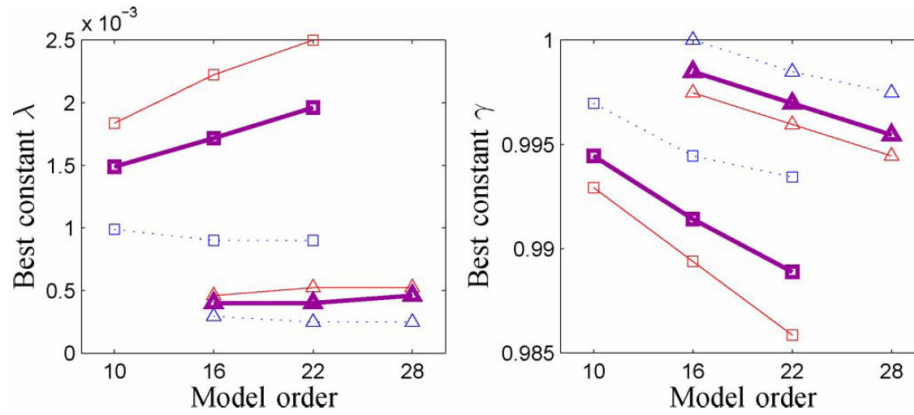


GRÁFICO 4 – Melhores constantes (λ e γ) encontradas na simulação, onde as linhas contínuas e vermelhas representam a voz feminina enquanto as linhas pontilhadas e azuis as masculinas. A taxa de amostragem é 8 kHz (quadrados) e 16 kHz (triângulos). As linhas em roxo representam os coeficientes usados pelo método adaptativo.

		$f_{\text{amostragem}} = 8 \text{ kHz}$			$f_{\text{amostragem}} = 16 \text{ kHz}$		
		10	16	22	16	22	28
Feminina	DS de LP	1,91	1,67	1,89	2,42	1,96	1,82
	Dif. de BE_{opt}	0,050	0,124	0,274	0,017	0,041	0,074
	Dif. de BE_C	0,032	0,096	0,233	0,007	0,026	0,056
	Dif. de R_{opt}	0,129	0,350	0,784	0,060	0,104	0,181
	Dif. de R_C	0,095	0,296	0,709	0,039	0,076	0,145
	Dif. de R_A	0,102	0,318	0,747	0,043	0,084	0,157
Masculina	DS de LP	2,44	1,99	1,85	3,22	2,63	2,44
	Dif. de BE_{opt}	0,031	0,057	0,079	0,016	0,035	0,052
	Dif. de BE_C	0,007	0,023	0,040	0,000	0,004	0,012
	Dif. de R_{opt}	0,076	0,101	0,150	0,045	0,059	0,072
	Dif. de R_C	0,035	0,051	0,093	0,016	0,025	0,033
	Dif. de R_A	0,040	0,058	0,101	0,018	0,028	0,037

DS = Distorção Espectral
Dif. = Diferença (para menos) das distorções espectrais quando comparado ao DS de LP.

TABELA 2 – Distorção espectral em dB para o método de LP convencional e as diferenças na queda da distorção para cada caso analisado entre os diferentes métodos modificados de LP.

4.2. λ Adaptativo

Uma simulação foi realizada utilizando como exemplo uma voz feminina amostrada em 8 kHz e modelo de ordem 16. Para cada frame, o melhor λ foi encontrado e marcado no gráfico 5. Após completar todo o processo, a curva que melhor se ajustou nos pontos é proposta pela seguinte relação:

$$\lambda(f_p) = ae^{bf_p} \quad (16)$$

onde f_p representa a frequência de *pitch* e $a > 0$.

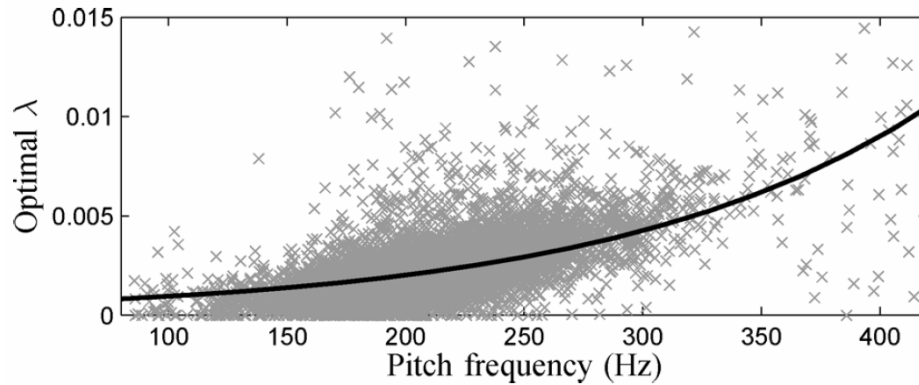


GRÁFICO 5 – λ ótimo em relação a cada frame da voz feminina amostrada em 8 kHz e modelo de ordem 16. A curva traçada através do modelo exponencial representa o melhor ajuste entre os pontos.

O processo de simulação foi repetido para todos os outros casos (taxa de amostragem, ordem e gênero), então, foram encontradas as melhores constantes λ e parâmetros para o modelo exponencial. O resultado está mostrado na tabela 3. A partir dos dados obtidos, chegou-se a conclusão que o valor ótimo de λ está relacionado praticamente a frequência *pitch* e não ao gênero de voz (masculina ou feminina), logo os dados apresentados vale para ambos os casos, uma vez que a diferença entre λ eram muito pequenos.

Ordem	$f_{\text{amostragem}} = 8 \text{ kHz}$			$f_{\text{amostragem}} = 16 \text{ kHz}$		
	10	16	10	16	10	16
$\lambda_{RC} (x 10^{-3})$	1,49	1,72	1,96	0,40	0,40	0,46
$a (x 10^{-4})$	5,73	4,46	4,55	1,49	1,12	1,07
$b (x 10^{-3})$	5,73	7,50	8,00	5,61	7,24	7,96

TABELA 3 – Melhores constantes λ e parâmetros para o modelo exponencial.

5. Validação dos Resultados

Para verificar se os resultados obtidos são realmente eficientes como mostrado anteriormente, a partir da referencia utilizada de outros exemplos de sons (extraídos de “*Acoustic-phonetic continuous speech corpus*”), foi analisado as distorções espectrais entre o método de LP convencional e o regularizado.

Utilizando os métodos R_C , R_A e BE_C com as melhores constantes λ e γ , uma nova simulação foi realizada para verificar as distorções espectrais de cada método.

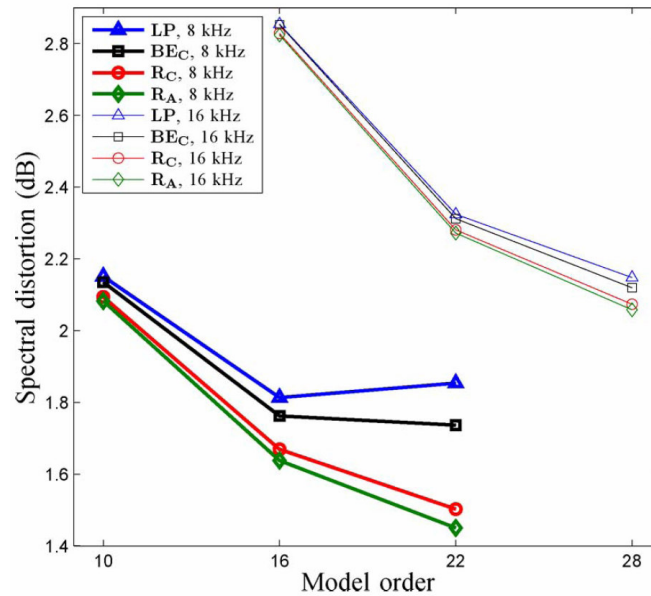


GRÁFICO 6 – Comparação das distorções espectrais médias que cada método gera.

Como é possível observar no gráfico 6, realmente o método que possui menor distorção em ambas taxas de amostragem utilizada, é o LP regularizado adaptativo. Outro ponto a ser validado é a dependência do modelo regularizado à frequência de *pitch*.

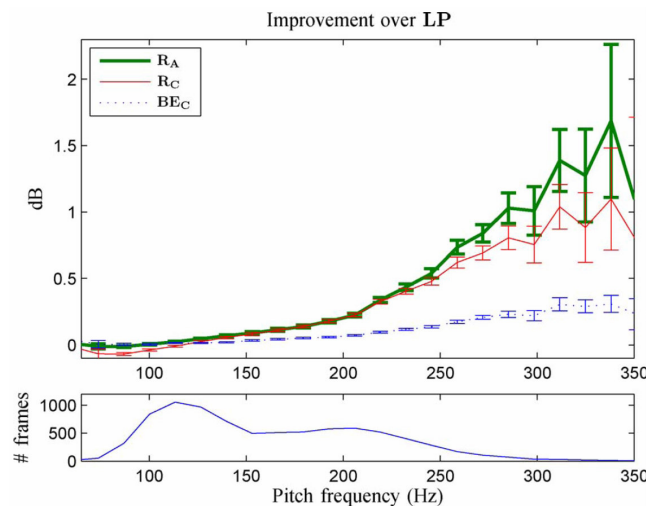


GRÁFICO 7 – Diminuição da distorção espectral sobre o LP para diferentes frequências de *pitch* e para diferentes métodos analisados. Foi utilizada a distribuição de probabilidade estatística T (*T de Student*) com intervalo de confiança de 95%. A curva azul representa a distribuição dos *frames* sobre as frequências de *pitch*. A taxa de amostragem utilizada é de 8 kHz e o modelo possui ordem 16.

Foi verificado também se o LP regularizado resultava em grandes erros do espectro, em baixas frequências quando comparado ao método de LP com largura de banda expandida. Utilizando o método de trens de pulso para a criação de fala sintética com frequência de *pitch* variável, foi realizada a validação do erro de distorção espectral em uma vasta faixa de frequência. O cálculo de distorção é dado por:

$$\|SD\|_{\infty} = \max_{\omega} |10 \log S(\omega, a) - 10 \log S(\omega, \hat{a})|$$

onde $S(\omega, a)$ é a envoltória verdadeira e $S(\omega, \hat{a})$ a envoltória baseada nos métodos de LP.

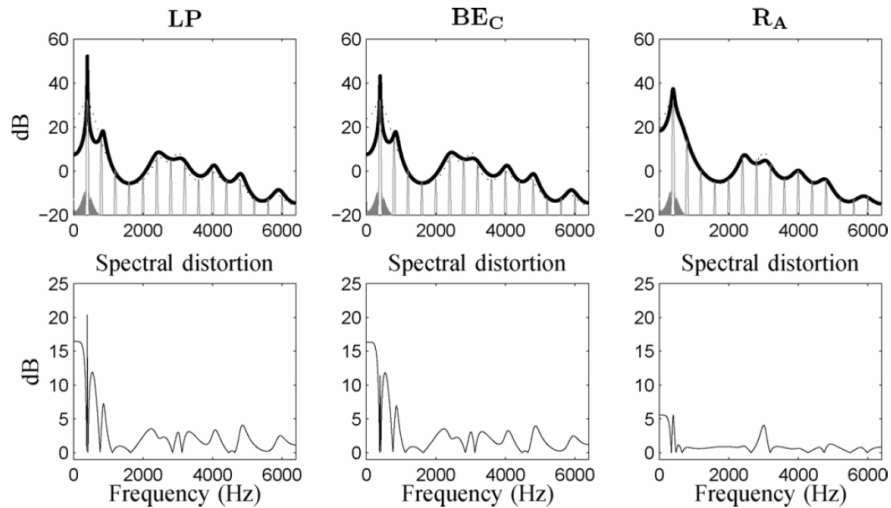


GRÁFICO 8 – Os 3 primeiros gráficos mostram a envoltória espectral resultante de cada método, já os 3 últimos gráficos representam a distorção espectral. O exemplo é de um *frame* com frequência de *pitch* de 400 Hz e taxa de amostragem de 12,8 kHz.

Através do gráfico 8 é possível verificar que o método que possui menor distorção em baixas frequências é o LP regularizado, mesmo que as falas sintéticas seja gerado a partir do método de LP com largura de banda expandida.

6. Conclusão

Existem vários métodos baseados no LP convencional, um dos mais utilizados atualmente é o LP com largura de banda expandida, pois com uma leve modificação nos coeficientes do modelo AR é possível obter uma envoltória espectral mais suave que no método convencional e isso gera um esforço computacional adicional praticamente nulo, porém diminui significativamente as distorções harmônicas espectrais.

Este trabalho apresentou uma nova variação do método convencional, o LP regularizado, onde com um parâmetro λ , a envoltória espectral é suavizada sem que aumente a complexidade de cálculo computacional. O objetivo foi atingido e esse novo método mostrou-se mais eficiente que os demais, pois além de diminuir a distorção espectral, se aproximou muito das simulações com coeficientes ótimos variáveis, porém os métodos de R_{opt} e BE_{opt} não são aplicáveis na prática, visto que a complexidade de análise é muita alta. Ainda vale ressaltar que o LP regularizado é predominantemente dependente da frequência de *pitch* e muito pouco dependente do gênero (voz masculina, feminina, entre outros), o que se aplica muito bem, visto que na maioria dos casos esse gênero é desconhecido para o programa.

Com esse novo método, as vantagens sobre o LP convencional são muitas, como modificação e codificação da fala e outras aplicações onde se deseja maior fidelidade a envoltória original.

7. Referências Bibliográficas

1. Regularized linear prediction of speech, L. Ekman, W. Kleijn e M. Murthi, IEEE transactions on audio, speech and language processing, vol 16, no. 1, 2008, pp 65-73.
2. Linear prediction: A tutorial review, John Makhoul, Proceedings of the IEEE, vol. 63, no. 4, 1975, pp 561-580.
3. Processamento Digital de Sinais, Marcelo de Oliveira Rosa, UFPR, versão 1.0.3, 2007.
4. AutoRegression Analysis (AR),
<<http://local.wasp.uwa.edu.au/~pbourke/other/ar/>>, acesso em 07/06/2008.
5. The Autoregressive Model and Formant Analysis,
<<http://cnx.org/content/m14203/latest/>>, acesso em 07/06/2008.
6. Análise Preliminar da Robustez do Codificar de Voz CELP, Natasha da Rocha Moura, UFRJ, 2005, pg 7-14.