

# Lafa – Laboratório de Fala

Carlos Meneses

M2A/ISEL – Grupo de Multimédia e Aprendizagem Automática  
Instituto Superior de Engenharia de Lisboa  
Rua Conselheiro Emídio Navarro nº1, 1950-062 Lisboa, Portugal  
cmeneses@deetc.isel.ipl.pt  
+351.218317224

## **Resumo**

*Lafa – Laboratório de Fala é uma aplicação gráfica de análise de sinais de fala, desenvolvida na plataforma MATLAB, com objectivos pedagógicos no âmbito de uma disciplina de “Processamento Digital de Fala”, de modo a assistir os alunos que se iniciam nesta área, mas também auxiliar na exposição de conceitos em sala de aula. No Lafa é possível analisar diversas características do sinal de fala: detecção de vozeamento e estimação da frequência fundamental; predição linear; estimação dos coeficientes LSF; estimação de formantes; e análise cepstral, incluindo mel-cepstra e cepstra de predição linear. Permite ainda, embora de um modo rudimentar, reconhecer vogais e sintetizar sons sustentados. O sinal de entrada pode estar gravado em formato wav ou ser adquirido do microfone.*

## **1. Introdução**

Lafa – Laboratório de Fala é uma aplicação gráfica de análise de sinais de fala, desenvolvida na plataforma MATLAB, com objectivos pedagógicos no âmbito da disciplina de opção do 5º ano “Processamento Digital de Fala”, do ISEL-DEETC. Esta disciplina tem sido frequentada por alunos da Licenciatura em Engenharia de Sistemas das Telecomunicações e Electrónica.

No Lafa é possível realizar diversas operações do sinal de fala: detecção de vozeamento e estimação da frequência fundamental; predição linear; estimação dos coeficientes LSF; estimação de formantes; e análise cepstral, incluindo mel-cepstra e cepstra de predição linear. É ainda possível visualizar o espectrograma de um sinal. Embora de um modo rudimentar, reconhece vogais baseado no “triângulo das vogais” e sintetiza sons sustentados. O sinal de entrada pode estar gravado em formato wav ou ser adquirido directamente do microfone.

O Lafa foi desenvolvido de modo a assistir os alunos que se iniciam nesta área e auxiliar na exposição de conceitos em sala de aula. É ainda possível elaborar um conjunto de questões a ser resolvidas em laboratório, comparando diversos métodos de análise ou de síntese.

Este software está acessível sem custos a partir do sítio<sup>1</sup> da disciplina de processamento digital de fala do ISEL-DEETC. Optou-se ainda por não se apresentar referências específicas, apresentado apenas o próprio sítio, que contém uma sebenta e as apresentações em sala de aula.

---

<sup>1</sup> <http://www.deetc.isel.ipl.pt/comunicacoes/disciplinas/pdf/>

## Análise de sinais de fala

A título de exemplo das capacidades de análise do LAFA é possível verificar a posição dos pólos do filtro de predição linear e comparara-los, como mostra a figura 1 em cima à esquerda, com os valores dos coeficientes LSF, apresentados sobre o círculo unitário, sendo evidente a atracção dos pólos sobre os coeficientes LSF, e a sua relação com os formantes. À direita em cima, compara-se a resposta em frequência do filtro de predição linear com o espectro da trama que lhe deu origem, emergindo naturalmente o conceito de envolvente espectral. No mesmo gráfico são apresentadas as posições dos formantes e da frequência fundamental, tornando estes conceitos intuitivos. Logo abaixo é mostrada a relação entre a autocorrelação da trama de entrada e da resposta impulsiva do filtro, mostrando que coincidem até à ordem de predição (10 neste exemplo). Logo abaixo é comparada a resposta impulsiva directamente com a trama, modelando um período fundamental.

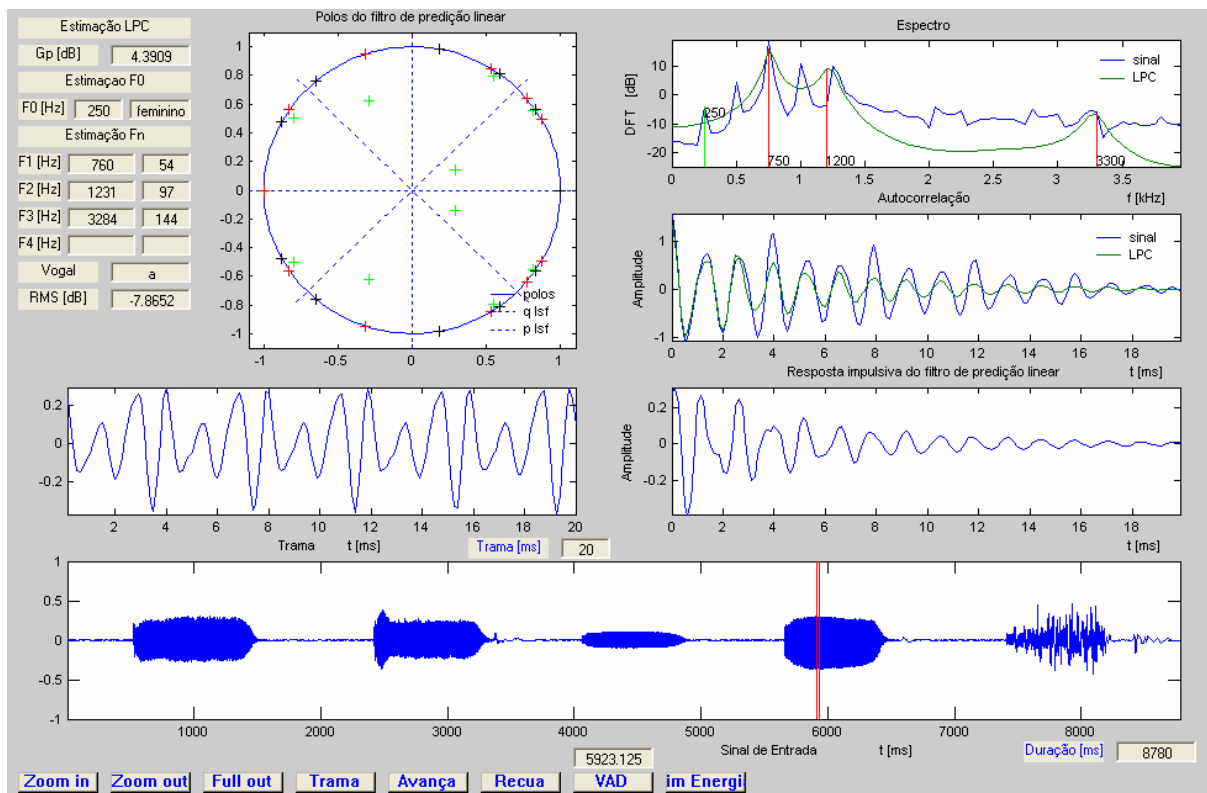
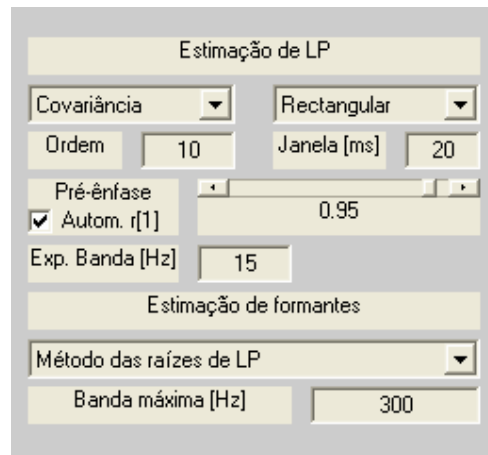


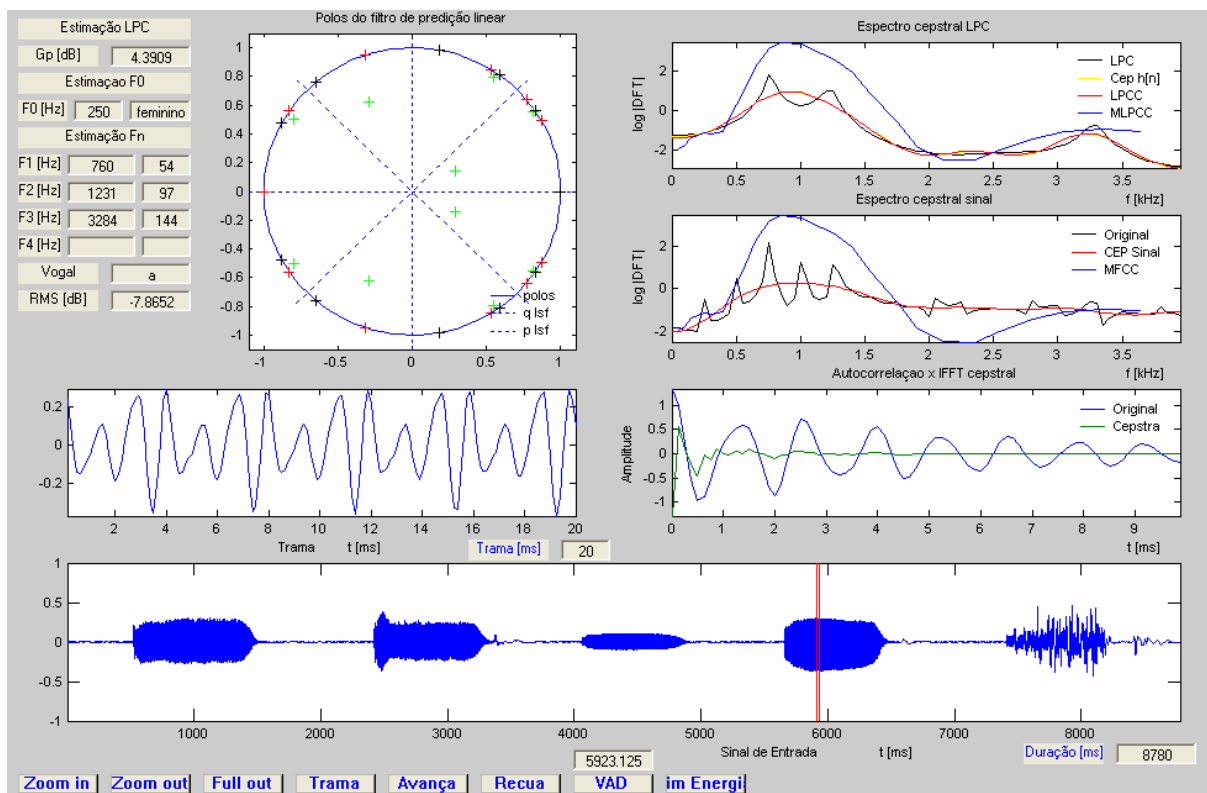
Figura 1: Análise de Predição Linear de uma trama.

O utilizador tem capacidade de alterar as preferências de modo a comparar métodos de estimação. No exemplo da figura 2 para as preferências de análise de predição linear, pode-se escolher entre o método de estimação (covariância/autocorrelação), ordem de predição, tipo de janela de análise (rectangular/hamming/hanning), pré-ênfase, expansão de banda, método de estimação de formantes (raízes de LP/pólos de LP e banda mínima).



**Figura 2: Janela de preferências da análise LP.**

Na figura 3 é ilustrada a análise *cepstral*. Os dois gráficos em cima à direita comparam vários métodos de estimação (*cepstra* de sinal com e sem filtro de MEL e *cepstra* de LPC) sendo também ilustrado o espectro do sinal original. Logo abaixo é comparada a autocorrelação do sinal com a transformada inversa do *cepstra* de sinal.

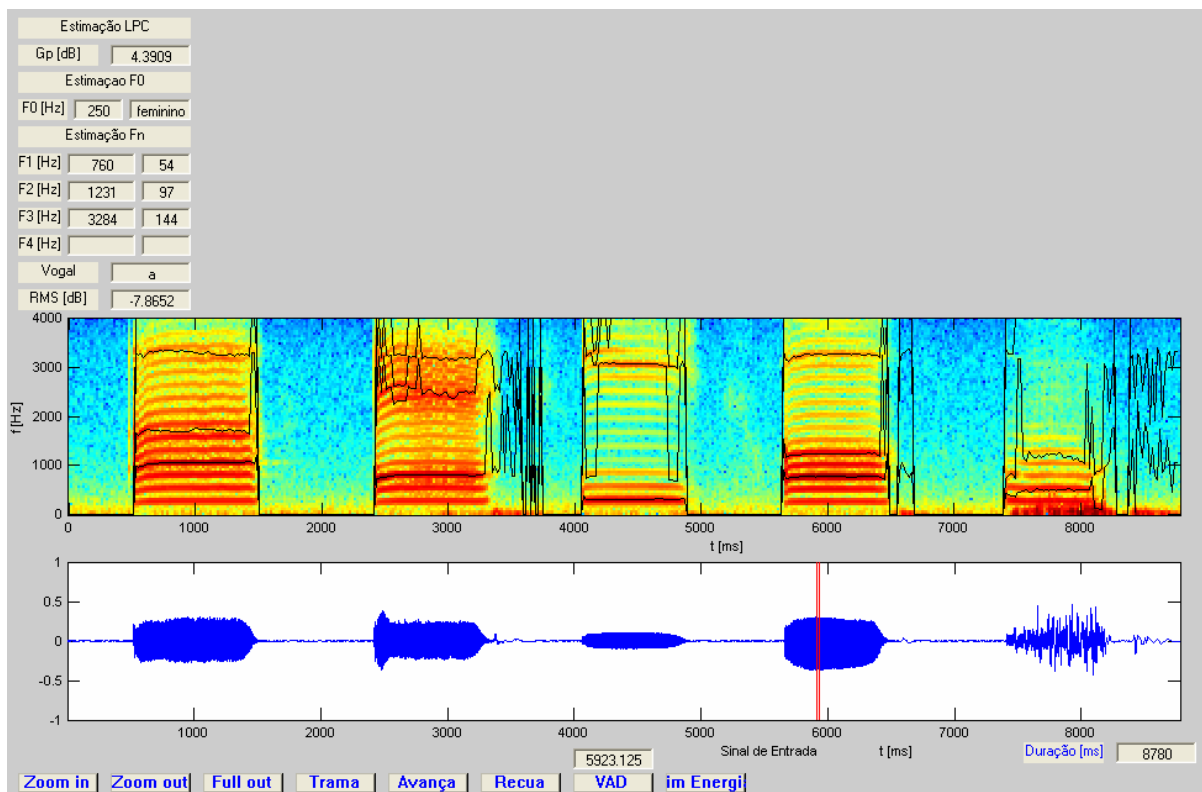


**Figura 3: Análise cepstral de uma trama**

O LAFA tem ainda a capacidade de exibir a evolução de um conjunto de parâmetros ao longo do tempo, (Espectrograma/formantes/LSF/frequência fundamental/RMS) alinhando-os com a forma de onda do sinal.

Na figura 4 é mostrado um exemplo da forma de onda e espectrograma tendo sobreposto os formantes. É possível verificar alguns erros, nomeadamente trocas entre o segundo e o terceiro formante.

A análise só é efectuada em zonas de actividade de voz. Para isso foi implementado um VAD (*voice activity detector*) que se baseia no valor do RMS trama a trama, mas também em informação temporal [2].



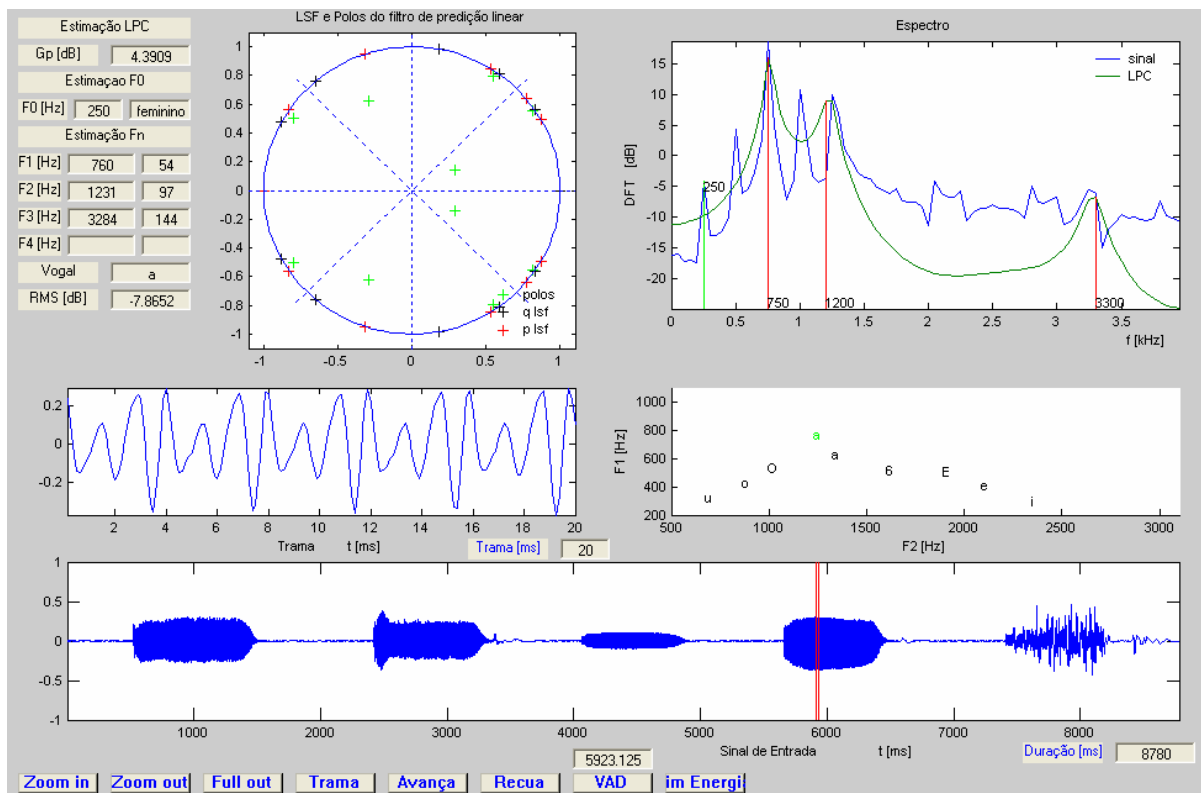
**Figura 4: Espectrograma com sobreposição de formantes, alinhado com a forma de onda**

## Reconhecimento de vogais

No âmbito do reconhecimento, com base na estimativa dos formantes detectam-se vogais. Da frequência fundamental é ainda possível revelar o género do orador ou determinar se é a voz de uma criança.

A figura 5 ilustra a análise de formantes, que inclui quer a detecção de vogais quer a detecção do género do orador. Esta informação é mostrada na tabela localizada em cima à esquerda, obtendo-se neste exemplo uma frequência fundamental de 250 Hz, a que corresponde um orador feminino.

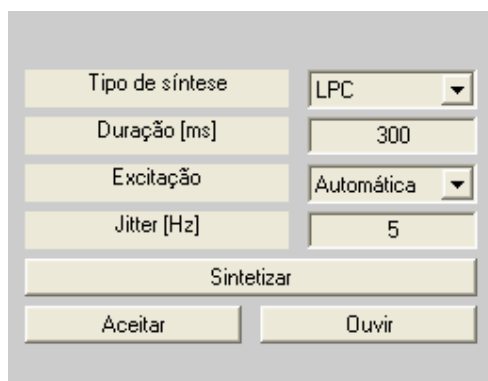
Quanto à detecção de vogais, é ilustrado na figura ao meio do lado esquerdo o triângulo das vogais, base do reconhecimento através de um critério de vizinho mais próximo ou máxima verosimilhança. Os valores dos formantes tomados como referências para cada vogal podem ser treinados, substituindo directamente os valores estimados, ou forçados por inserção dos valores dos formantes, bem como da respectiva vogal, na tabela localizada em cima à esquerda. Neste exemplo os dois primeiros formantes tomam os valores de 760 e 1231 Hz, ponto inserido a verde e com a identificação da vogal reconhecida. (a – “a aberto”).



**Figura 5: Análise de Formantes de uma trama e correspondente identificação da vogal e do género do orador**

## Síntese de sons sustentados

Da análise de uma trama ou forçando valores de parâmetros, é possível sintetizar sons sustentados com diversos modelos: síntese de formantes, de predição linear ou sinusoidal. O tipo de excitação pode ser periódica (diracs ou modelo da glote) para sintetizar sons vozeados, ruído branco para sintetizar sons não vozeados, ou detecção automática do tipo de excitação. O sinal sintetizado pode ser ouvido e revertido a sinal principal, tornando-se o sinal objecto de análise. A figura 6 ilustra a janela de preferências na síntese.



**Figura 6: Preferências de síntese de sons sustentados.**

## Conclusões

Foi apresentado uma aplicação de análise de sinais de fala com capacidades gráficas, em que é possível observar o efeito dos principais métodos de caracterização de sinais de fala. Tem ainda capacidades rudimentares de reconhecimento e síntese, mas que dão uma primeira impressão sobre estes procedimentos.

Desenvolvido na plataforma MATLAB, algumas das funções utilizadas fazem parte da biblioteca do próprio MATLAB (*e.g.* LPC, *cepstra*). Outras foram desenvolvidas intencionalmente para esta aplicação (*e.g.* LSF, *mel-cepstra*, *pitch*), de um modo modular, ficando disponíveis e tendo já sido utilizadas para outras aplicações de processamento de fala, tais como o desenvolvimento de codificadores e de reconhecedores. Esta modularidade tem permitido a criação de novas funções e a sua fácil integração.

O LAFA tem sido ajuda preciosa para os alunos que se iniciam nesta área, mas também na exposição de conceitos em sala de aula ou em investigação em processamento de fala na análise de casos particulares de alguns sinais de fala.

## Referências

- [1] Carlos Meneses, “Sítio da disciplina de Processamento Digital de Fala”, ISEL-DEETC-SCPS  
<http://www.deetc.isel.ipl.pt/comunicacoesep/disciplinas/pdf/>
- [2] Lori F. Famel, Lawrence R. Rabiner, Aarong E. Rosenberg e Jay G. Wilpon, “An Improved Endpoint Detector for Isolated Word Recognition”, Agosto 81