

# Redução de Ruído em Sinais de Fala por Subtracção Espectral

Carlos Teixeira e Isabel Trancoso

representando o consórcio SUNSTAR†

**Sumário** - Construiu-se um pré-processador para reconhecedores de fala, capaz de reduzir ruído estacionário ou quasi-estacionário. O sinal de fala a processar é recolhido através da rede telefónica pública, nas condições típicas (ruidosas) do quotidiano. O método escolhido para este fim é conhecido pela designação de Subtracção Espectral: a amplitude do espectro do ruído, estimada durante as pausas do sinal de fala, é subtraída à amplitude do espectro discreto do sinal original. Concebeu-se um algoritmo para execução deste método em tempo real. Posteriormente desenvolveu-se este algoritmo num processador digital de sinais (DSP32C). Num futuro próximo prevê-se a integração no mesmo pré-processador de um cancelador de eco.

## I - Introdução

A necessidade de generalização de novas aplicações a serviços de atendimento públicos (meteorologia, despertador, apostas, movimento de capitais, bases de dados específicas, espectáculos, resultados desportivos, marcação de reservas, etc.), aumentam cada vez mais as exigências no desempenho dos reconhecedores automáticos de fala.

O problema aqui focado surge no âmbito dos serviços oferecidos através da rede telefónica pública. As condições ambientais em que o utente consulta esses serviços introduzem invariavelmente ruído, quer dentro de edifícios (outros falantes, máquinas registadoras, impressoras, teclados, telefones, campainhas, electrodomésticos, etc.) quer no exterior (veículos motorizados). Além disso a rede telefónica introduz ruído eléctrico (bem como como a própria instalação do assinante) habitualmente designado como ruído de linha. No entanto, a distorção de amplitude e fase produzida pelo canal telefónico não foi ainda tida em conta neste trabalho.

O problema do ruído é aqui abordado com a utilização de um pré-processamento para a respectiva redução da componente ruidosa. Numa perspectiva mais vasta e integrada com o próprio reconhecedor, surgem alternativas como a da generalização dos 'Hidden Markov Model' (Varga, 1990) convencionais para uma decomposição óptima de processos simultâneos.

Contudo, a modularidade inerente ao pré-processamento representa uma vantagem importante no âmbito do projecto mais vasto em que este decorre. A redução de ruído é feita sobre o próprio sinal, dispondo-se indistintamente de saída analógica ou/e digital sem custos computacionais adicionais. Assim, qualquer reconhecedor

de fala já desenvolvido, pode ser utilizado sem mais modificações.

Uma das técnicas de pré-processamento mais conhecida baseia-se na utilização de mais que um microfone (Martins, 1990). Contudo, esta técnica revela-se inviável no contexto do nosso problema, uma vez que implicaria custos desmesurados para as alterações necessárias na rede telefónica pública.

O método escolhido é conhecido pela designação de Subtracção Espectral (Boll, 1979): a amplitude do espectro do ruído, estimada durante as pausas do sinal de fala, é subtraída à amplitude do espectro discreto do sinal original. Deste modo, consegue-se reduzir ruído estacionário ou quasi-estacionário no sinal de fala. Este método não pressupõe treino prévio, e são conhecidos de antemão bons resultados na bibliografia (Van Comperolle, 1987).

O algoritmo desenvolvido neste trabalho decorre da adaptação do método original, para processamento em tempo real.

Este artigo encontra-se estruturado em cinco secções: uma descrição muito breve do método de base conhecido; uma outra mais detalhada em que se define a estrutura do algoritmo para a sua execução em tempo real; outra ainda com os detalhes do desenvolvimento prático deste algoritmo; finalmente apresentam-se alguns resultados experimentais e, para concluir, apontam-se perspectivas das consequências futuras deste trabalho.

## II - O Método

O método da Subtracção Espectral assenta em quatro pressupostos (Boll, 1979):

- O ruído a cancelar foi somado acústicamente ou digitalmente ao sinal de fala.
- Esse ruído mantém-se localmente estacionário de tal modo que a sua amplitude, estimada num período de tempo exclusivamente com ruído, é igual à que poderia ser estimada durante a actividade de fala que sucede imediatamente a esse período.
- O ruído pode ser lentamente variável em termos de estacionaridade.
- A qualidade da redução é aceitável se apenas for retirado o efeito do ruído na amplitude espectral e não na fase.

Partindo destes pressupostos o método consiste em determinar o espectro e subtrair-lhe a estimativa da amplitude do espectro do ruído. Antes de reconstruir o sinal de saída, procede-se ainda a uma rectificação do sinal

† O trabalho no projecto SUNSTAR é desenvolvido no âmbito do programa ESPRIT e parcialmente financiado pela Comissão das Comunidades Europeias. O consórcio SUNSTAR é formado pelas seguintes companhias: Jydsk Telefon (DK) (à qual o INESC se encontra associado), Alcatel FACE Standard (I), Fraunhofer Gesellschaft (D) e Telefónica I+D (E).

de diferença, de forma a terem-se sempre amplitudes espectrais positivas.

No trabalho de S. Boll são ainda mencionados procedimentos adicionais para redução do ruído residual e atenuação dos segmentos de sinal apenas com ruído. Contudo, não foram referidos elementos essenciais ao processamento em tempo real, nomeadamente como se efectua a decisão de estimar ou actualizar a estimativa da amplitude espectral do ruído.

### III - Algoritmo para Tempo Real

A estrutura do algoritmo para processamento em tempo real está representada esquemáticamente na figura 1.

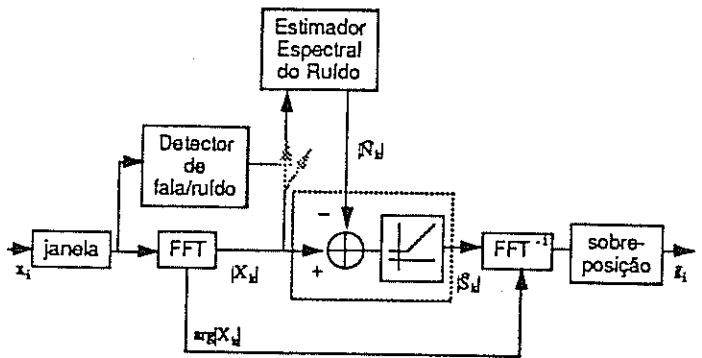


Figura 1: Estrutura do Algoritmo para Tempo Real.

Deste esquema sobressaem cinco tipos de blocos básicos: a subtracção das amplitudes espectrais e a respectiva rectificação; dois blocos para as transformadas directa e inversa de Fourier; mais dois blocos para memorizar os segmentos de sinal de entrada e de saída e finalmente os dois blocos mais críticos no desenvolvimento prático do método: o estimador da amplitude espectral do ruído e o detector de fala/ruído. Os blocos de retenção de sinal incluem a multiplicação por janela de Hamming dos segmentos à entrada, e a sobreposição a 50% dos segmentos de saída.

#### III.1. Detector de fala/ruído

O detector de fala/ruído decide, tal como o nome indica, se um segmento de sinal contém fala sobreposta no ruído ou apenas ruído. Uma vez que existem necessariamente segmentos limítrofes durante os quais se inicia ou termina a actividade da fala, esta decisão binária não pode, nestes casos, ser rigorosa. Contudo, um dimensionamento suficientemente baixo dos segmentos, permite que o erro cometido nestes casos não comprometa o desempenho global.

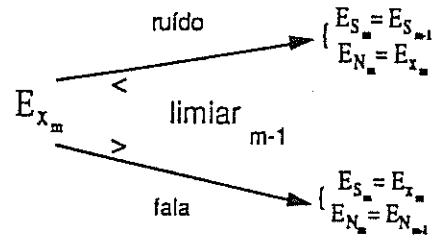
O detector de fala/ruído concebido para este trabalho, baseia a sua decisão para cada segmento de sinal na respectiva energia.

Duas estimativas da energia, uma para o sinal com fala, outra para o sinal só com ruído, concretizam medidas localmente estacionárias com um factor de esquecimento exponencial  $(1-\beta)$  ajustável. A grande vantagem destas estimativas consiste na facilidade da sua actualização recursiva.

Com estas duas estimativas constrói-se um limiar de decisão para a energia de cada segmento. Acima do limiar classifica-se como segmento de fala e abaixo deste, como segmento de ruído. O limiar é calculado acima da estimativa da energia do ruído de uma quantidade ajustável pelo parâmetro  $\gamma$ . Essa quantidade é proporcional à diferença entre a energia do sinal e a do ruído.

#### Detector de fala/ruído

$$E_{x_m} = \frac{\sum_{i=0}^{n_{amostr\_segm} - 1} x_i^2}{n_{amostr\_segm}} \quad \text{: energia no segmento "m"}$$



$$\hat{E}_{S_m} = (1-\beta) \sum_{i=0}^m \beta^{m-i} E_{S_i} \quad \text{estimativa da energia do sinal de fala}$$

$$\hat{E}_{N_m} = (1-\beta) \sum_{i=0}^m \beta^{m-i} E_{N_i} \quad \text{estimativa da energia do ruído}$$

$$0 < \gamma < 1 \quad (\gamma \cong 0.2)$$

$$\text{limiar}_m = \gamma (\hat{E}_{S_m} - \hat{E}_{N_m}) + \hat{E}_{N_m} \quad 0 < \beta < 1 \quad (\beta \cong 0.8)$$

Figura 2: Detector de fala/ruído.

#### III.2. O Estimador do Ruído

O estimador da amplitude espectral do ruído utiliza e beneficia da mesma técnica utilizada para as estimativas da energia da fala e do ruído. Cada elemento discreto da amplitude do espectro é actualizado recursivamente com os elementos dos segmentos classificados pelo detector como ruído. Utilizou-se igualmente um factor de esquecimento  $(1-\alpha)$  exponencial.

### IV - Desenvolvimento Prático

Nesta secção descrevem-se sumariamente detalhes do desenvolvimento e opções quer da programação quer do processador e da placa em que está inserido.

Existem basicamente dois tipos de procedimentos, um em tempo real e outro em tempo simulado para processamento dos ficheiros DOS residentes num IBM®PC.

#### IV.1. O Processador

Numa primeira versão dispunha-se de um processador DSP32 ("Digital Signal Processor") numa placa da Burr-Brown (B-B) a DSPeet™ - ZPB32 IBM®PC CARD que utiliza o codec T7520 numa outra placa da B-B, a ZPB100, para a conversão AD/DA. Na última versão final adaptaram-se as aplicações num DSP32C da Loughborough Sound Images, Ltd. (LSI), a 'DSP32C PC System Board'.

Só na placa da LSI é possível a execução em tempo real do algoritmo com sobreposição de segmentos a 50%. Efectivamente, esta sobreposição quase duplica o tempo de processamento, não sendo assim acessível ao DSP32 que suporta no máximo um ritmo de relógio de 16MHz. O DSP32C accionado a um ritmo de 50MHz, não só resolve este problema como ainda lhe sobra tempo para outros processamentos previstos no âmbito do projecto SUNSTAR (cancelador de ruído). Assim, enquanto na placa da B-B se consumia 80% do tempo disponível para a aplicação em tempo real sem utilizar sobreposição de segmentos de sinal, na placa LSI utiliza-se apenas 50% incluindo já a sobreposição de segmentos de sinal, com a multiplicação por janela de Hamming. Os segmentos são de 256 amostras sobrepostos em 128 amostras (16 ms) obtidas a um ritmo de 8 KHz. Esta sobreposição implica um atraso fixo de 48 ms entre a entrada e a saída do sinal.

Na aplicação em tempo real a conversão AD/DA é feita de modo diferente em cada uma das placas. A B-B utiliza o codec T7052 através do porto série acessível quer por rotinas de interrupção quer utilizando DMA ("direct memory access"). Na segunda utiliza-se uma técnica de "memory mapped" onde um endereço fixo de memória é lido e escrito a cada interrupção. Assim, o código de interface da placa LSI é completamente diferente do anterior e mesmo a própria rotina de subtração espectral teve de ser alterada. Este código é único quer para a comunicação com os AD/DA, quer com os ficheiros DOS, ao contrário do código desenvolvido no DSP32, e permite adicionalmente a comunicação com o ST-BUS. Este foi a norma escolhida para a comunicação digital de dados entre os diversos DSP's no projecto SUNSTAR.

O carregamento do código no DSP é feito a partir de ficheiros DOS no IBM®PC onde estão instalados. e ocupa cerca de 12 Kbytes (10 Kbytes de variáveis + 2 Kbytes de código executável).

#### IV.2. Programa

O código encontra-se dividido em três partes: o código de interface, as rotinas de retenção de amostras, multiplicação por janela de Hamming e sobreposição, e o algoritmo de subtração espectral propriamente dito.

Em primeiro lugar há a necessidade de inicializar as estimativas do algoritmo e o respectivo limiar. As estimativas de energia e da amplitude espectral do ruído são atribuídas às do primeiro segmento recolhido. O limiar é escolhido arbitrariamente grande, de modo a que o primeiro segmento seja classificado como ruído e não como fala (a hipótese mais provável e a mais simples de impôr).

O algoritmo de subtração espectral vem representado na figura 3.

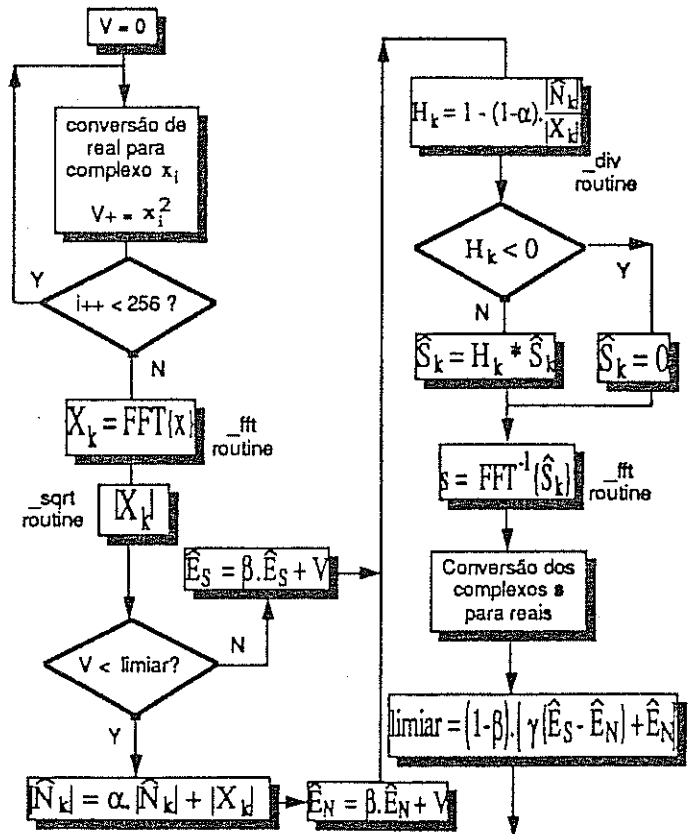


Figura 3: Fluxograma abreviado do algoritmo de Subtração Espectral.

Para as transformadas de Fourier directa e inversa usou-se uma rotina de FFT (Fast Fourier Transform). O processamento desta rotina representa mais de 60% do tempo consumido por segmento. Assinala-se que na figura 3, todos os blocos representados entre os blocos de FFT são repetidos para  $k=1, \dots, 256$ .

Além da rotina de FFT é necessária uma rotina para a raiz quadrada na determinação do módulo do espectro e uma de divisão para o cálculo dos factores  $H_k$ .

Assinala-se ainda a conveniência do uso da forma complexa polar que evita preocupações adicionais com a fase. O significado das variáveis no fluxograma coincide com os da figura 2.

#### V - Resultados

Para a avaliação dos resultados estão previstas duas fases. A primeira fase, da qual se apresentam resultados preliminares, inclui apenas testes auditivos informais e determinação do melhoramento na relação sinal ruído. A segunda fase está prevista para depois da referida integração com um cancelador de eco e consiste na determinação da influência do módulo de redução de ruído nas taxas de aceitação e rejeição atingidas por um reconhecedor de palavras isoladas baseado em modelos de Markov não observáveis.

Para a primeira fase de testes, adicionaram-se digitalmente sinais em ficheiros, de ruído aos de fala, por forma a obterem-se relações de sinal/ruído de cerca de 6 dB.

Nos resultados obtidos com ruído branco obtiveram-se incrementos de cerca de 10 dB nas relações de sinal/ruído.

As figuras 4 e 5 ilustram estes resultados com um sinal artificial na entrada - uma sinusóide pura (250 Hz) adicionada de ruído branco (relação de sinal/ruído de 6 dB). A primeira apresenta os sinais de entrada (linha seguida) e o de saída (linha pontuada), a segunda a amplitude dos respectivos espectros. Depois do processamento a fundamental mantém-se praticamente imperturbável enquanto algumas bandas de ruído chegam a ser atenuadas mais de 20 dB.

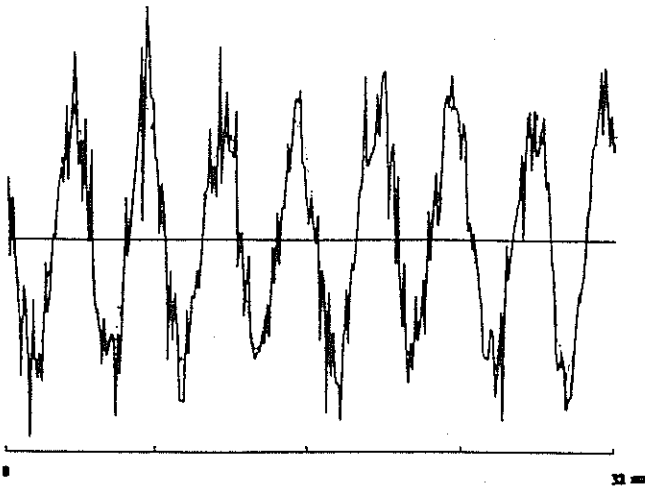


Figura 4: Sinal de entrada ( — ) e o de saída ( - - ).

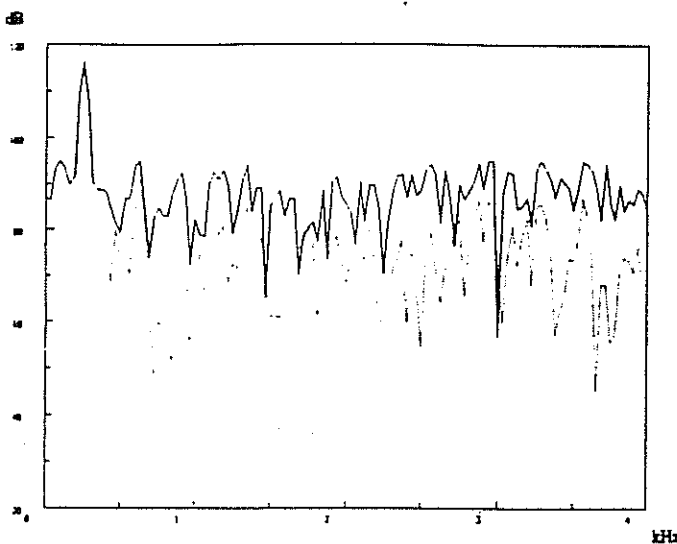


Figura 5: Amplitude do espectro de entrada ( — ) e o do de saída ( - - ).

## VI - Conclusões e Perspectivas Futuras

Desenvolveu-se um pré-processador de fala capaz de reduzir ruído estacionário e quase-estacionário, num ambiente compatível com as outras aplicações do projecto SUNSTAR, com vista a um demonstrador de serviços públicos automáticos na rede telefónica.

Actualmente está em curso a integração do módulo de redução de ruído com um outro, de cancelamento de eco, de modo a actuarem simultaneamente num único DSP32C. Este segundo módulo elimina ecos com atrasos até 32ms (não cancela os de longa distância).

Na fase seguinte, proceder-se-á a testes que avaliem a contribuição efectiva destes módulos no reconhecimento automático, com as bases de sinais recolhidas pelo consórcio SUNSTAR.

Em meados deste ano o mesmo consórcio apresentará um demonstrador integrando estes módulos com reconhecedores e aplicações desenvolvidos em processadores idênticos. Estes serão instalados em diversos IBM®PC's e geridos a partir de uma estação de trabalho SPARK. Este demonstrador permitirá atender a várias linhas telefónicas simultaneamente, possibilitando alguns serviços de interesse público.

O problema do ruído será ainda tido em consideração, numa perspectiva integrada, no futuro desenvolvimento de um reconhecedor de 2ª geração. Haverá então condições, para uma comparação experimental de desempenhos com a técnica aqui utilizada.

## Agradecimentos

Agradece-se ao Prof. Paul Dalsgaard e ao Prof. Børge Lindberg do 'Speech Technology Centre' da Universidade de Aalborg, os conselhos e experiência transmitida, bem como a cedência do código de interface para ficheiros DOS e para a placa LSI.

## Referências

- Boll S. (1979). Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Trans. Acoustic, Speech, and Signal Proc.*, Vol. ASSP-27, pp.113-120.
- Van Compernelle D. (1987). Increase noise immunity in large vocabulary speech recognition with the aid of spectral subtraction. *Proc. Int. Conf. Acoust., Speech and Signal Proc.*, pp. 1143-1146.
- Varga A.P., Moore R.K. (1990). Hidden Markov Model Decomposition of Speech and Noise. *IEEE Proc. ICASSP 90*, pp. 845-848.
- Martins C. R., Almeida T. M., Piedade M. S. (1990). An Adaptive Noise Canceller and its implementation on a DSP. *IEEE ISCAS 90*, New Orleans, USA, pp. 1939-1942.
- AT&T (1986). Application Software Library - Reference Manual, WE® DSP32-AL Digital Signal Processor. .
- AT&T (1988). Information Manual, WE® DSP32 and DSP32C Digital Signal Processor. .
- AT&T (1988). User Manual, WE® DSP32 and DSP32C C Language Compiler.
- Burr-Brown® (1988), User Manual, DSpeed™ - ZPB32 IBM®PC CARD.
- Loughborough Sound Images, Ltd. (1989). DSP32C PC System Board - User Manual, Issue 2.0.