

Natanael M. Gomes

Universidade de Taubaté

Marcio Abud Marcelino

Universidade Estadual Paulista

Universidade de Taubaté

Francisco José Grandinetti

Universidade Estadual Paulista

Universidade de Taubaté

Correspondência/Contato

UNIVERSIDADE DE TAUBATÉ
Departamento de Engenharia Mecânica

Rua Daniel Danelli, s/n, Jd. Morumbi
Taubaté - SP
CEP 12060-440
Fone (12) 3625-4193

Editores responsáveis

Prof. Dr. Luiz Eduardo Nicolini do P. Nunes
luiz.nunes@unitau.com.br

Profa. Dra. Valesca Alves Correa
valesca.correa@unitau.com.br

DETERMINAÇÃO DO GÊNERO DO LOCUTOR USANDO A TRANSFORMADA RÁPIDA DE FOURIER

RESUMO

Este trabalho apresenta uma técnica de determinação do gênero do locutor utilizando a Transformada Rápida de Fourier. A distinção entre a voz de um homem e a de uma mulher se dá pela frequência fundamental da voz, sendo que na masculina está compreendida entre 80 e 160 Hz e, na feminina, entre 160 e 150 Hz. Para determinar a frequência fundamental é utilizada a Transformada Rápida de Fourier (FFT) para obter o espectro de frequência do sinal capturado. O algoritmo desenvolvido foi implementado em um *kit* de desenvolvimento da Analog Devices e obteve resultados que permitiram a distinção na quase totalidade dos ensaios.

Palavras-chave: Processamento digital de sinais, voz humana, Transformada Rápida de Fourier.

1. INTRODUÇÃO

Nesse trabalho é apresentado uma forma de determinar o gênero do locutor utilizando a Transformada Rápida de Fourier em um processador digital de sinais. Também são apresentados a base do processamento digital de sinais, uma breve descrição do processador e do programa utilizados e os resultados obtidos.

Para determinar o gênero do locutor é realizada uma análise para delimitar a frequência fundamental do sinal da voz. A frequência fundamental da voz é a frequência que contém maior energia em todo o espectro que compreende o sinal da voz. A voz masculina tem frequência fundamental compreendida entre 80Hz e 160Hz e a voz feminina entre 160Hz e 250Hz [5].

2. PROCESSAMENTO DE SINAIS

Para utilizar processadores na análise de sinais é necessária a digitalização do sinal analógico, e nesse processo deve-se respeitar a Teoria da Amostragem [3]. Determinando assim a frequência de amostragem e a frequência de corte do filtro de entrada, a frequência máxima do sinal, determinada pelo filtro *antialiasing*, deve ser menor ou igual à metade da frequência de amostragem.

O ouvido humano é capaz de captar sons de até 20kHz, assim uma frequência de amostragem de 40kHz garante que não haverá perda na faixa audível. Porém para a voz humana uma faixa de frequência de 4kHz garante que 99% de inteligibilidade da conversação, e é muito comum utilizar frequência de amostragem de 8kHz para aplicações que envolvam processamento de voz.

3. TRANSFORMADA RÁPIDA DE FOURIER

Para obter o espectro de frequência da voz é utilizada a Transformada Rápida de Fourier, que é um algoritmo melhorado para executar a Transformada Discreta de Fourier, *Discrete Fourier Transform* (DFT). Todas as operações são realizadas no domínio complexo.

A Transformada Discreta de Fourier é dada pela Equação 1:

$$\tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n]W_N^{kn} \quad (1)$$

Onde,

$$W_N = e^{-j(2\pi/N)} \quad (2)$$

A equação 1 é chamada equação de síntese da DFT e tem custo computacional para solução da ordem de N^2 , onde N é o número de amostras para computar a DFT. Os algoritmos mais eficientes pertencem a um grupo conhecido como FFT, ou Transformada Rápida de Fourier. Uma FFT é capaz de reduzir o número de operações da ordem de N^2 para ordem de $N \log N$, tornando possível a utilização da Transformada de Fourier em aplicações de tempo real [4].

A Figura 1 mostra uma comparação entre ordem de N^2 e $N \log N$.

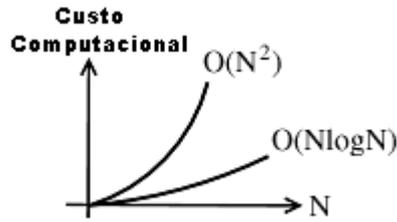


Figura 1. Custo computacional.

A FFT utiliza as características de simetria e periodicidade no tempo para resolver a DFT através de DFTs menores, chamada de decimação, que pode ser no tempo ou na frequência. O processo de decimação consiste em quebrar uma DFT em duas DFTs menores e repetir o processo até obter um vetor unitário. Assim a FFT sempre terá como entrada e saída vetores de tamanho 2^n .

Na Figura 2 é mostrado uma etapa de decimação no tempo para uma amostra de 8 pontos, reduzindo de uma DFT de 8 pontos para duas DFTs de 4 pontos cada e mais algumas operações, observa-se que existe um padrão de execução da decimação. Ao realizar mais uma decimação este padrão também é observado.

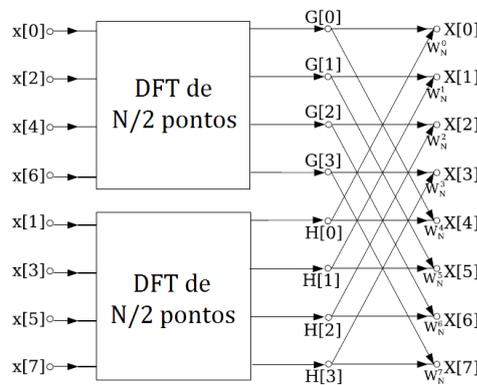


Figura 2. Decimação no tempo.

A estrutura básica de uma FFT é a chamada de borboleta, já que a representação gráfica se assemelha ao inseto, e é mostrada na Figura 3.

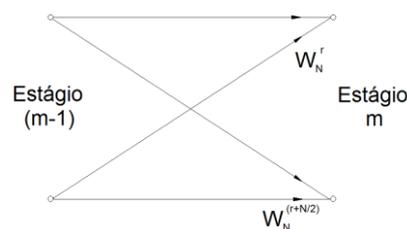


Figura 3. Estrutura básica da FFT

A Figura 3 mostra que a borboleta utiliza duas variáveis do estágio $m-1$, soma e multiplica por uma constante complexa, representada na notação da equação 2. Esta variável pode ser estocada em uma tabela para economizar custo computacional. A estrutura é repetida nos vários estágios do programa mudando-se apenas os

índices, sendo executado como uma função e os índices ou a fase de execução são dados de entrada para essa função.

Um algoritmo que utilize a estrutura básica como na figura 1 é chamado FFT de raiz de 2 mas existem algoritmos de raiz 4, que são mais complexos mas que obtém soluções ainda mais rápidas que algoritmos raiz de 2 [4].

4. IMPLEMENTAÇÃO DO FILTRO

Este trabalho foi implementado em um processador de sinais ADSP-2101 da Analog Devices de 16 bits com capacidade máxima de 25 MIPS (Milhões de Instruções por Segundo). Com uma placa de desenvolvimento EZ-ICE equipada com o CODEC TP3054, que realiza as funções de A/D e D/A, e tem comunicação compatível com μ -law [2].

A comunicação com o CODEC é feita através da porta serial SPORT do processador, sempre que o CODEC tem um novo dado disponível, esse dado é transferido pela porta serial e o processador recebe uma interrupção de dados na SPORT, quando um vetor de dados está disponível é chamada a função FFT. A solução da FFT foi realizada com a chamada computação no lugar (*in place*), isto é, os dados de entrada são processados, e disponibilizados como saída, nas mesmas posições de memória [2] [4].

Vale ressaltar que todos os dados são complexos e para realizar as transformadas os vetores de dados são constituídos de amplitude e fase.

A frequência de amostragem utilizada foi de 8kHz, permitindo capturar a maior parte da informação contida na voz, com uma Transformada Rápida de Fourier raiz de 4 de 256 pontos, conseguiu se uma resolução do espectro de frequência de 31,25Hz. As amostras de frequências na região de estudos ficam espaçadas como apresentado na Tabela 1.

Tabela 1. Amostras da região de estudo

Índice	1	2	3	4	5	6	7	8
Freq. (Hz)	31,25	62,5	93,75	125	156,25	187,5	218,75	250

De acordo com a Tabela 1, caso as amostras de índice 3, 4 ou 5 tiverem maior amplitude, o algoritmo determina que a voz é masculina. Caso as amostras de índice 6, 7 ou 8 tiverem maior amplitude determina que a voz é feminina.

Na Figura 4 é mostrado um sinal de voz masculina composto por 256 amostras, nestas figuras é mostrada apenas a amplitude dos sinais.

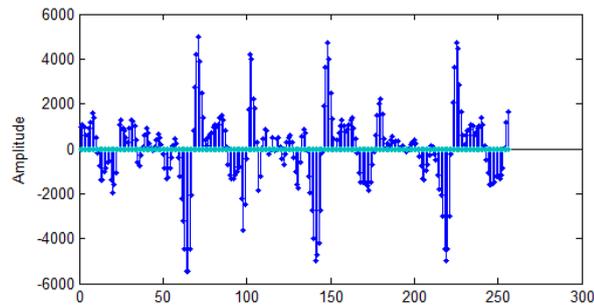


Figura 4: Sinal de voz masculina.

Na Figura 5 é mostrada a FFT realizada através do software Matlab e na Figura 6 a FFT realizada pelo programa desenvolvido para o DSP.

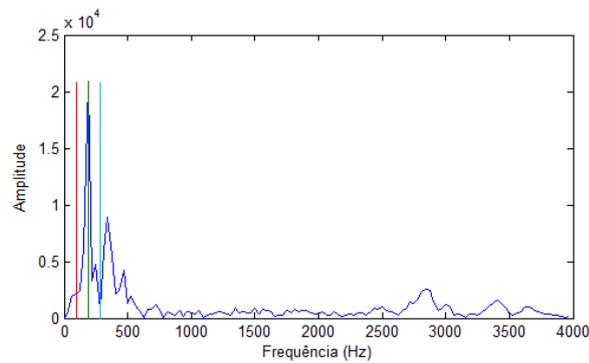


Figura 5. FFT realizada pelo Matlab.

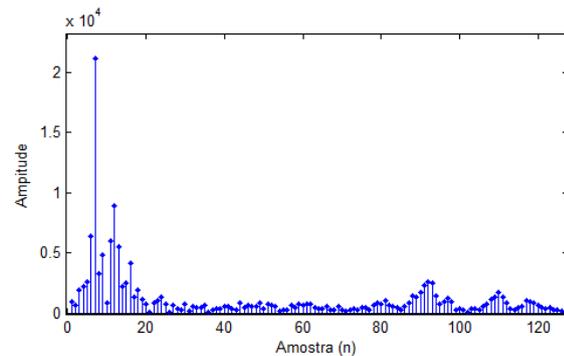


Figura 6. FFT realizada no DSP.

Observa-se que a FFT realizada pelo programa no DSP é fiel à realizada pelo Matlab, utilizado para a comparação por ser um software profissional. Na Figura 5 foram acrescentadas linhas para melhor observar os limiares de decisão do programa.

Durante a execução, o programa analisa um total de 10 amostras e decide pela maioria, com 5 votos para um tipo de voz ele dá a saída. A saída utilizada pelo programa é um LED que pisca com frequências diferentes para cada caso, 1Hz para voz masculina e 8Hz para voz feminina.

5. RESULTADOS

Os resultados foram satisfatórios, reconhecendo o gênero do locutor na maioria dos casos, com grande repetitividade. Em poucos casos percebeu-se que o sistema identificou erroneamente locutores do gênero feminino, e após analisar esses casos concluiu-se que são locutoras com frequências fundamentais da voz próximas do limite de decisão. Analisando a Tabela 1, nas amostras de índice 5 e 6, observa-se um erro de até 7,4% no limiar de decisão, tendendo a indicar erroneamente voz masculina.

O som ambiente e a utilização incorreta do microfone também geram resultados errôneos, porém são problemas mais simples de se resolver, utilizando uma sala silenciosa e a distância correta para utilização do microfone.

Um fator que degradou os resultados é a frequência de corte de 80 Hz do filtro A/D do CODEC utilizado, com atenuações abaixo de 200Hz. Assim foi necessário compensar a atenuação durante o processamento.

6. CONCLUSÃO

A identificação do gênero do locutor utilizando a Transformada Rápida de Fourier em um processador digital de sinais é um trabalho complexo que é o início para uma análise mais completa como o reconhecimento de voz. Neste trabalho foi utilizado um processador de pequeno porte e ainda assim obteve-se resultados satisfatórios, para análises mais complexas faz-se necessário o uso de um processador com maior capacidade de processamento e resolução de bits.

REFERÊNCIAS

Analog Devices, "ADSP-2100 Family EZ Tools Manual"; 1994.

Analog Devices, "ADSP-2100 Family User's Manual"; 1994.

HAYES, M. Schaum's Outline of Digital Signal Processing. New York: McGrawHill. 1998.

OPPENHEIM, Alan, and Ronald Schafer. Digital Signal Processing. Englewood Cliffs, N. J.: PrenticeHall. 1989.

RABINER, L. R. et al. A comparative performance study of several pitch detection algorithms, IEEE Transactions on Acoustics, Speech and Signal Processing, VOL. ASSP-24, N. 5, 1976.