



COBENGE 2005

XXXIII - Congresso Brasileiro de Ensino de Engenharia

"Promovendo e valorizando a engenharia em um cenário de constantes mudanças"

12 a 15 de setembro - Campina Grande - Pb

Promoção/Organização: ABENGE/UFCG-UFPE

UM AMBIENTE COMPUTACIONAL PARA AUXÍLIO AO ENSINO NA ÁREA DE PROCESSAMENTO DIGITAL DE SINAIS DE VOZ

Maí C. R. de Vasconcelos – maicrv@gmail.com

Departamento de Engenharia Elétrica.

Departamento de Sistemas e Computação

Universidade Federal de Campina Grande

Av. Aprígio Veloso, 882, Bodocongó

58.109-970 – Campina Grande – PB

Joseana M. Fehine – joseana@dsc.ufcg.edu.br

Benedito G. A. Neto – bganeto@dee.ufcg.edu.br

José L. do Nascimento – jluisn@dee.ufcg.edu.br

Mozart G. C. Pontes – mozart@dee.ufcg.edu.br

***Resumo:** Este trabalho trata do estudo, desenvolvimento e implementação de um software para auxílio ao ensino e à pesquisa na área de processamento digital de sinais de voz. A ferramenta em desenvolvimento consiste em um ambiente computacional construído em linguagem de programação C++, capaz de realizar tarefas essenciais para o pré-processamento e extração de parâmetros temporais da fala, comuns à inúmeras pesquisas no contexto da comunicação vocal homem-máquina. Além da produção do aplicativo, o trabalho aborda também o estudo de técnicas de pré-processamento de sinais de voz para implementação de algoritmos alternativos para detecção de início e fim de palavras.*

***Palavras-chave:** Ambiente Computacional para Auxílio ao Ensino, Processamento Digital de Sinais de Voz, Biblioteca de Funções para Pré-processamento de Sinais de Voz*

1. INTRODUÇÃO

A comunicação oral é, sem dúvida alguma, a forma mais natural de comunicação humana. Em virtude da interação homem-máquina se tornar cada vez mais comum, surge uma demanda natural por sistemas capazes de reconhecer o que está sendo dito, bem como quem está falando e responder ao que está sendo solicitado. O interesse nessa área se deve ao número de aplicações, bem como a existência de várias questões teóricas que ainda não foram respondidas[1, 2, 3, 4].

Na recente história do computador, a flexibilidade da chamada "Máquina Universal" foi de fundamental importância. Porém, o fato de os computadores só realizarem tarefas de processamento de dados, através de operações matemáticas, se tornou limitado. Hoje, acredita-se que os computadores necessitam perceber o mundo, através de sentidos similares aos nossos, como visão, audição, tato e, por que não, olfato e paladar? Além disso, esses equipamentos também necessitam de meios de expressão, como a voz, as mãos, e a habilidade de criar uma diversidade de formas para representar a informação.

Com a disponibilidade das linhas telefônicas, microfones e placas de som acopladas aos computadores, o custo da implementação de sistemas para possibilitar a comunicação vocal homem-máquina, está relacionado, basicamente, ao projeto de software.

Qualquer aplicação no contexto da comunicação vocal Homem-Máquina exige uma série de procedimentos padrão a serem realizados, a exemplo da necessidade de realização das etapas que constituem o pré-processamento digital do sinal de voz. Observou-se que diversos pesquisadores das principais subáreas da comunicação vocal Homem-Máquina (Sistemas de Resposta Vocal, Sistemas de Reconhecimento de Fala e Sistemas de Reconhecimento de Locutor) despendiam/repetiam esforços, isoladamente, na realização dessas etapas básicas.

Visto que o pré-processamento é fundamental a qualquer subárea da comunicação vocal homem-máquina, o intuito deste trabalho é desenvolver um aplicativo (ferramenta de software) para auxílio/otimização das pesquisas nessas subáreas, fornecendo uma biblioteca de funções para o pré-processamento digital de sinais de voz.

Além do foco principal do aplicativo ser o auxílio à pesquisa, este foi desenvolvido de uma forma bastante didática e com interface amigável para que pudesse ser utilizado em disciplinas de ensino de processamento digital de sinais de voz.

Na Figura 1 é ilustrado o contexto no qual o aplicativo se insere.

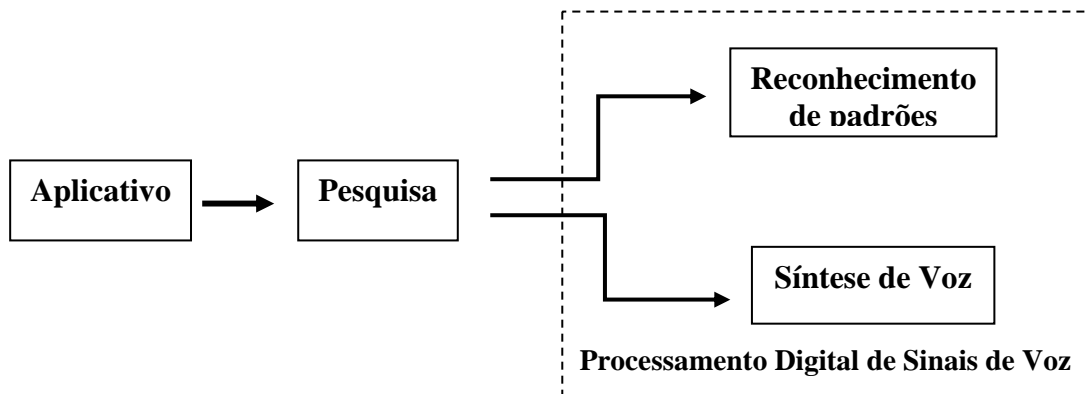


Figura 1 - Contexto do aplicativo.

2. COMUNICAÇÃO VOCAL HOMEM-MÁQUINA

Com o desenvolvimento tecnológico surgiram equipamentos eletrônicos com o objetivo de melhorar a qualidade de vida do homem moderno. Tais equipamentos, embora sofisticados, enfrentam ainda dificuldades quanto a sua utilização, devido à forma artificial com que o usuário deve interagir com os mesmos. Assim, parece claro que o desenvolvimento de uma interface vocal, tornaria mais fácil e produtiva a relação Homem-Máquina [1, 2, 5].

Em função disso e do crescente desenvolvimento tecnológico de *hardware* para o processamento digital de sinais, o meio de comunicação mais adequado seria a fala humana. Tal meio de comunicação proporciona uma cômoda adaptação do usuário e a capacidade de transmitir uma grande quantidade de informações com pouca interação.

A entrada vocal é bastante adequada para aplicações em que uma ou mais das seguintes condições se aplicam: as mãos do usuário estão ocupadas; mobilidade é exigida durante o processo de entrada de dados; os olhos do operador devem permanecer fixos sobre um *display*; um instrumento óptico ou algum objeto é rastreado; é inconveniente o uso do teclado em um ambiente, dentre outras. Por não requererem nem as mãos nem os olhos do usuário para sua operação, os sistemas de entrada vocal podem ser utilizados em diversas

aplicações, como por exemplo: controle de tráfego aéreo, auxílio a portadores de necessidades especiais, controle de qualidade e inspeção e controle de acesso a ambientes restritos[1-5].

A comunicação vocal entre pessoas e máquinas inclui síntese de voz, reconhecimento automático de voz e o reconhecimento de locutores a partir de suas vozes. Portanto, a comunicação vocal homem-máquina se divide nas seguintes subáreas principais [1, 5]: Sistemas de Resposta Vocal, Sistemas de Reconhecimento de Fala e Sistemas de Reconhecimento de Locutor.

Os sistemas de resposta vocal são projetados para responder a um pedido de informação utilizando mensagens faladas. Assim, a comunicação de voz em sistemas de resposta vocal se faz em uma única direção, da máquina para o homem [1, 6].

A tarefa básica no reconhecimento de fala é reconhecer uma determinada elocução de uma sentença ou “entender” um texto falado [1, 2].

Em um sistema de reconhecimento de locutor, o objetivo é verificar se o usuário é quem alega ser (verificação de locutor) ou identificar um dado locutor em um conjunto de locutores [1, 6]. Para o reconhecimento de fala e de locutor, a comunicação se faz do homem para a máquina.

Atualmente, os estudos pertinentes ao processamento digital de sinais, com vistas ao desenvolvimento de aplicações na área da comunicação vocal homem-máquina, evoluíram muito graças aos avanços tecnológicos como o computador e as placas de som, as quais tornaram possível a conversão do som em dados digitais. As técnicas de processamento de sinais permitem a extração de características que realmente mereçam destaque, pois atuam no sentido de fornecer não somente a informação de interesse ao processamento de determinada amostra de som, como também ocasionar uma redução considerável na quantidade de informações a serem processadas. Tais informações serão responsáveis pela produção de padrões a serem identificados numa comparação entre determinada referência registrada e a apresentação de uma nova amostra para teste. Então, o papel do sistema será validar ou não determinada amostra, dependendo do tipo e funcionalidade do sistema de reconhecimento em questão. Portanto, o desenvolvimento de aplicações na área da comunicação vocal homem-máquina consiste em uma tarefa de reconhecimento de padrões da fala [1, 2].

O pré-processamento de sinais é uma etapa importante na realização da tarefa de reconhecimento de padrões. Esta etapa é responsável por extrair do sinal de voz os dados relevantes e menosprezar a informação redundante, com a finalidade de repassar a informação de interesse à fase de comparação dos padrões.

3. PRÉ-PROCESSAMENTO DO SINAL

O pré-processamento do sinal de voz inclui a aquisição, pré-ênfase e janelamento, podendo ser incluída a detecção automática de início e fim de palavra [1-6].

Na etapa de aquisição do sinal de voz, a informação analógica captada no microfone é convertida em sinais digitais.

Após a aquisição do sinal de voz, é realizado um pré-processamento nas amostras a fim de prepará-las para a extração de suas características [1, 7, 8]. Tais características são utilizadas, posteriormente, em um algoritmo para reconhecimento de padrões da fala.

3.1 Detecção de Início e Fim de Palavras

A detecção de início e fim de palavras tem por objetivo eliminar intervalos de silêncio ou de ruído ambiente indesejáveis. As amostras sem informação relevante devem ser

eliminadas de forma precisa, de maneira que seja considerada apenas a informação útil de forma a otimizar os cálculos nas demais etapas.

Tradicionalmente, algoritmos de detecção de início e fim utilizam a energia do bloco de amostras como parâmetro principal, sendo o mesmo adotado como parâmetro suficiente para a detecção. Porém, essa metodologia leva a uma decisão bastante radical, tendo em vista que utiliza um limiar pré-determinado de energia. O início é detectado quando uma seqüência sucessiva de blocos com energia acima do limiar é encontrada e o fim quando uma seqüência com energia abaixo do limiar é encontrada após o início [1, 8].

Para suavizar a decisão, um algoritmo foi proposto utilizando 2 limiares de energia e o parâmetro da taxa de cruzamento por zero como auxiliar ao da energia. Essa metodologia implica ainda numa otimização da detecção para as sentenças que possuem início ou fim com sons surdos, visto que, dependendo das condições do ambiente de gravação estes sinais se confundem com um ruído em termos de níveis de energia. Nesses casos, a taxa de cruzamento por zero foi bastante decisiva [1, 5, 8].

3.2 Pré-Ênfase

A pré-ênfase objetiva eliminar uma tendência espectral de aproximadamente -6dB/oitava na fala irradiada dos lábios. Essa distorção espectral não traz informação adicional e pode ser eliminada através da aplicação de um filtro, de resposta aproximadamente +6dB/oitava, que ocasionaria um nivelamento no espectro. Para um sistema digital, a pré-ênfase pode ser implementada como um circuito analógico, o qual precede o filtro e o amostrador, ou como uma operação digital no sinal amostrado, através de um filtro FIR de primeira ordem.

A equação de diferença (1) descreve a pré-ênfase do sinal [1, 2, 5, 6].

$$y(n) = x(n) - a.x(n-1) \quad (1)$$

para $1 \leq n < M$, em que M é o número de amostras do sinal, $x(n)$, e a constante “ a ” é escolhida entre 0.9 e 1.

3.3 Janelamento

Em todas as aplicações práticas de processamento de sinais, é necessário trabalhar com "pequenos segmentos" ou *frames* do sinal, a não ser que o sinal seja de curtíssima duração. Isso é verdade especialmente se estiverem sendo utilizadas técnicas de análise convencionais de sistemas lineares invariantes no tempo. A segmentação consiste em particionar o sinal de voz em segmentos, selecionados por janelas ou quadros de duração perfeitamente definida. Os segmentos são escolhidos dentro dos limites de estacionariedade do sinal (duração média de 16 a 32 ms) [1, 6].

Os tipos básicos de janela são: Janela Retangular, Janela de Hamming e Janela de Hanning. A janela de Hamming é normalmente utilizada por apresentar características espectrais úteis e por atenuar a transição entre *frames* adjacentes. Sua descrição matemática pode ser vista em [1, 6]. As janelas usualmente são sobrepostas entre si, para que a variação dos parâmetros entre janelas sucessivas seja mais gradual [1, 2, 5, 6].

4. PARÂMETROS TEMPORAIS DO SINAL DE VOZ

O gráfico amplitude-*versus*-tempo de um sinal permite a avaliação de muitas características que possibilitam uma completa descrição do mesmo. A partir do uso de parâmetros temporais, torna-se possível identificar características básicas da fala. Para as mais diversas aplicações na área de processamento digital de sinais de voz, os parâmetros temporais do sinal são bastante importantes, pois caracterizam o bloco de amostras, sendo úteis para aplicações de reconhecimento de padrões e síntese de voz.

Dentre esses parâmetros, destacam-se: a Energia do Sinal, a Taxa de Cruzamento por Zero, o Coeficiente de Correlação Normalizado, o Número Total de Picos, dentre outros [1, 2, 5, 6]. A obtenção dos parâmetros é levada a efeito após o pré-processamento do sinal, com as formas de obtenção descritas a seguir.

4.1 Energia do Segmento

A energia por segmento, E_{seg} , é definida pela Equação (2) [1]:

$$E_{seg} = N_A \cdot E\{[s(n) - \mu_{s(n)}]^2\} \quad (2)$$

Para sinais ergódicos e estacionários no sentido amplo, com média nula, como o sinal de voz, E_{seg} é definida pela Equação (3) e Equação (4) [1, 2]:

$$E_{seg} = N_A \cdot E\{[s(n)]^2\} = \sum_{n=0}^{N_A-1} [s(n)^2] \quad (3)$$

$$E_{seg}(dB) = 10 \cdot \log[E_{seg}] \quad (4)$$

em que $s(n)$ é o sinal de voz, $\mu_{s(n)}$ a média de $s(n)$ e N_A o tamanho da janela (bloco de amostras do sinal em análise). A energia é obtida, portanto, simplesmente, somando-se os quadrados das amplitudes das N_A amostras do sinal contido na janela em análise, devendo refletir as variações de amplitude do sinal de voz entre intervalos ou janelas.

4.2 Taxa de Cruzamento por Zero

A taxa de cruzamento por zero, TCZ, é outro parâmetro bastante utilizado em aplicações de processamento digital de sinais de voz, que utilizam métodos de análise no domínio do tempo. Esse parâmetro indica o número de vezes que as amostras de um sinal, em um determinado segmento, cruzam o zero (limiar) tomado como referência.

Esse parâmetro é, geralmente, definido pela Equação (5) [1, 2]:

$$E_{seg} = N_A \cdot E\{\text{sgn}[s(n)] - \text{sgn}[s(n-1)]\} = \sum_{n=1}^{N_A-1} |\text{sgn}[s(n)] - \text{sgn}[s(n-1)]|, \quad (5)$$

Sendo,

$$\text{sgn}[s(n)] = \begin{cases} 1, & \text{se } s(n) \geq 0 \\ -1, & \text{se } s(n) < 0 \end{cases}$$

Isto significa que a contagem é efetuada sempre que amostras sucessivas do sinal tiverem polaridades contrárias.

Taxas de cruzamento por Zero elevadas caracterizam os sons surdos e taxas mais reduzidas indicam a presença de sons sonoros.

4.3 Coeficiente de Correlação Normalizado

Este parâmetro tem bastante utilidade na diferenciação de sons surdos dos sonoros. Para sons sonoros, seu valor é muito próximo da unidade devido à concentração de energia nas baixas frequências do espectro. Já para os sons surdos, o valor do parâmetro é próximo de zero.

A Equação (7) descreve a obtenção dos coeficientes [1, 2]:

$$Cor = \frac{\sum_{n=1}^N [S_n \cdot S_{n-1}]}{[(\sum_{n=1}^N S_n^2) \cdot (\sum_{n=0}^{N-1} S_n^2)]^{\frac{1}{2}}} \quad (7)$$

4.4 Número Total de Picos e Diferença entre os picos

Estes parâmetros medem o número de picos encontrados e a diferença entre picos positivos e negativos de um segmento de voz. Tradicionalmente não são muito utilizados, mas podem ser úteis na identificação de consoantes fricativas de pequena intensidade [1].

Diante do exposto acima, observa-se que o pré-processamento de sinais é uma etapa importante no estudo da comunicação vocal homem-máquina. As técnicas de pré-processamento permitem a extração de características que realmente merecem destaque, pois atuam no sentido de fornecer não somente a informação de interesse ao processamento de determinada amostra de som, como também ocasionar uma redução considerável na quantidade de informações a serem processadas.

5. METODOLOGIA

A metodologia empregada na produção do aplicativo consistiu no estudo, desenvolvimento, otimização e implementação em software de uma biblioteca de funções básicas para o pré-processamento e extração de características temporais de sinais de voz.

O procedimento utilizado foi baseado na execução de etapas, conforme descrição a seguir.

- Levantamento do Estado da Arte das tecnologias voltadas ao desenvolvimento de sistemas na área do processamento digital de sinais de voz.
- Estudo detalhado da documentação do padrão *Microsoft Wave File Format*.
- Implementação em linguagem C++ de um detector de início e fim de palavras. Tal etapa tem por objetivo eliminar os intervalos de silêncio existentes no início e no fim das locuções proferidas.
- Implementação em linguagem C++ das funções que executam a etapa de pré-processamento do sinal de voz (aquisição, pré-ênfase e janelamento).
- Implementação em linguagem C++ das técnicas básicas para a extração dos parâmetros temporais representativos da fala.

Para a execução de cada uma das etapas supracitadas, foi de fundamental importância o desenvolvimento de uma interface eficiente para as funções implementadas que permita ao aluno visualizar as saídas correspondentes a cada função no final de cada etapa, como por exemplo, a visualização e audição da forma de onda do sinal de voz original e do sinal após a pré-ênfase.

6. RESULTADOS

O ambiente computacional para a realização das etapas descritas acima foi desenvolvido utilizando a Linguagem de Programação C++ [10], tendo como premissa básica ser uma ferramenta de fácil utilização e que possa auxiliar o ensino na área processamento e transmissão digital de sinais de voz.

O ambiente computacional (ferramenta de software) possui uma interface bastante intuitiva mesmo para os menos experientes na área, conforme Figura 2.

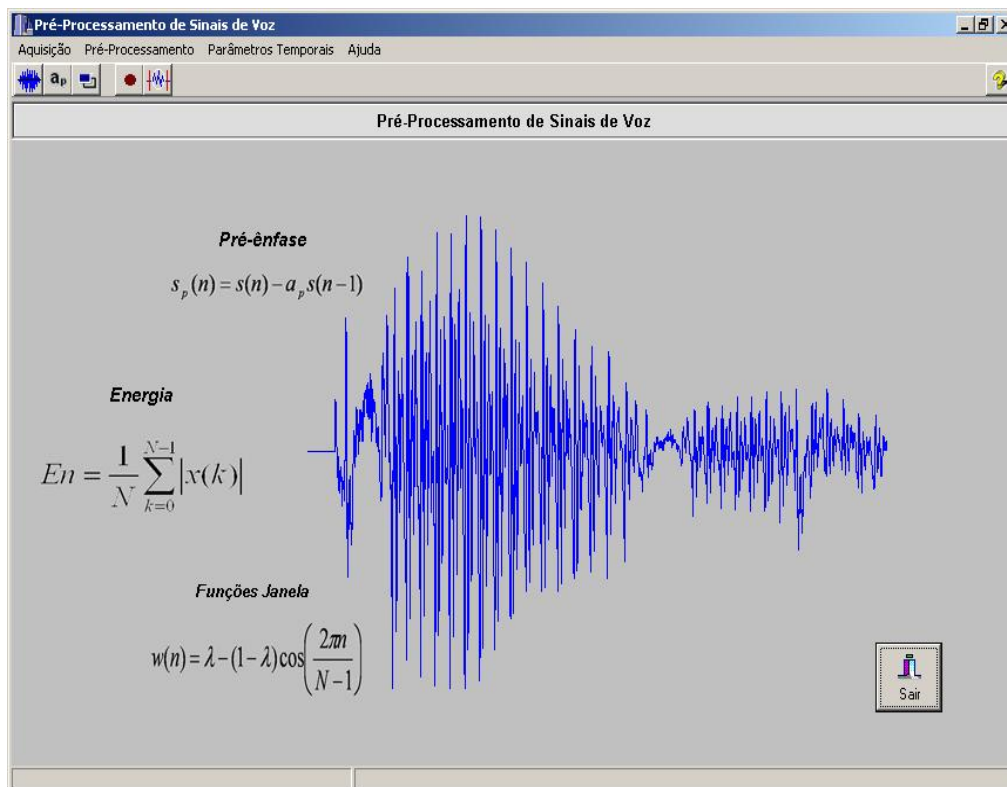


Figura 2 - Tela inicial do ambiente computacional.

A ferramenta também dispõe de um “help” que possui todas as informações necessárias para a operação correta das suas funcionalidades.

6.1 Leitura do Sinal

A operação de leitura de um arquivo de voz é apresentada na Figura 3. A forma de onda é apresentada, juntamente com o cabeçalho hexadecimal do arquivo de áudio e as

amostras do mesmo. As informações contidas no cabeçalho hexadecimal são decodificadas e apresentadas na tela.

Na etapa de leitura, é permitido ao usuário abrir um arquivo de áudio, sendo fornecidas na tela informações do tipo: taxa de amostragem do sinal, número de amostras, código de compressão, número de canais, duração, número de bits por amostra e, por fim, as amostras do sinal. A forma de onda do sinal poderá ser visualizada, como também poderá ser “tocada”.

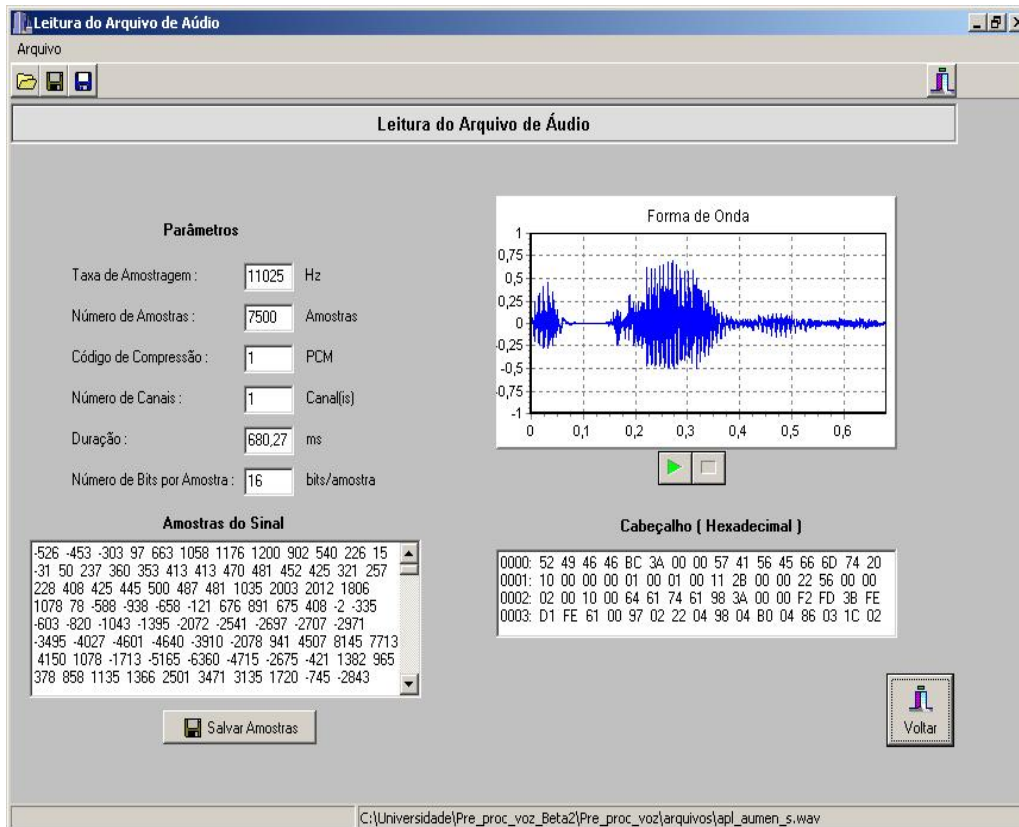


Figura 3 - Tela para leitura do arquivo de áudio.

O formato de gravação é o padrão Microsoft® WAV, importado do sistema operacional Windows o Microsoft® Gravador de som, que é utilizado como interface de aquisição de dados.

Na aquisição do sinal de voz é importante observar a relação sinal-ruído do dispositivo utilizado para garantir a qualidade dos resultados a serem obtidos no pré-processamento deste sinal

A manipulação de arquivos de áudio é realizada a partir de um conjunto de funções agrupadas em uma classe, denominada classe WAV [10].

Ao final desta etapa, as amostras do sinal (sem cabeçalho) são gravadas em um arquivo *.txt*.

6.2 Detecção de início e fim de palavras

A Figura 4 apresenta a tela que realiza a etapa de detecção. O software permite visualizar a forma de onda original, a forma de onda com o silêncio eliminado e a distribuição de energia do sinal original.

Nessa etapa, um arquivo de extensão *.wav*, é automaticamente gravado e as formas de onda podem ser armazenadas em formatos de imagens digitais.

O algoritmo de detecção utilizado buscou trabalhar com decisões suaves, sendo implementado utilizando como parâmetro principal a energia do sinal por segmento e como parâmetros auxiliares a taxa de cruzamento por zero e o número total de picos.

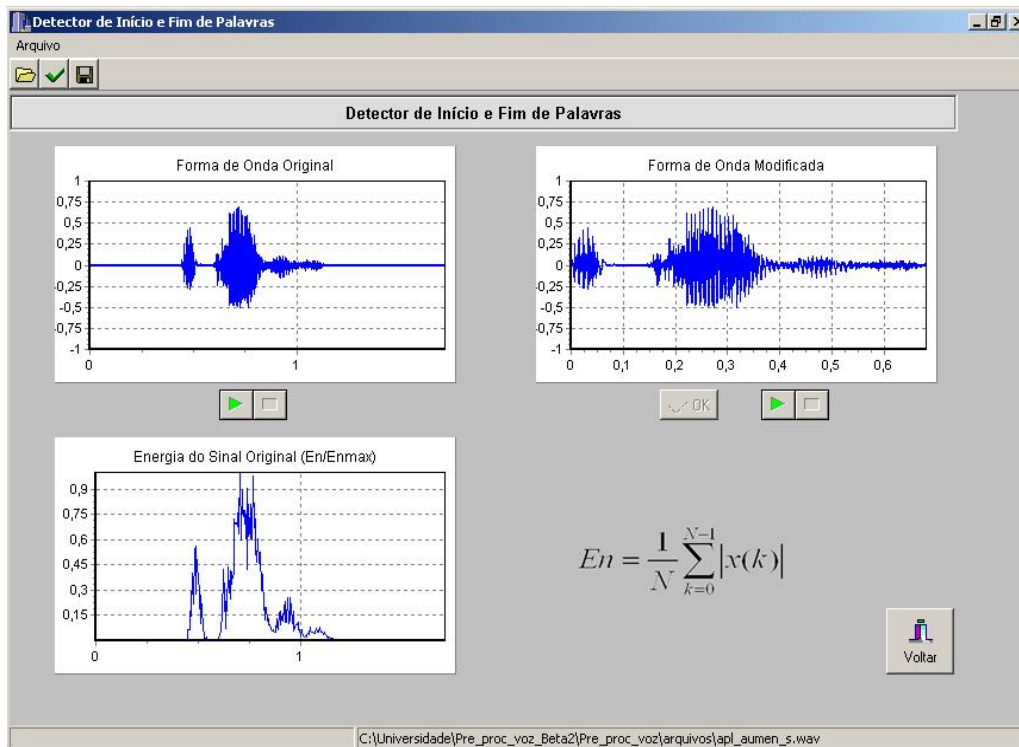


Figura 4 - Tela para detecção de início e fim de palavras.

6.3 Pré-ênfase do Sinal

Nesta etapa o usuário poderá realizar a pré-ênfase, podendo definir o fator de pré-ênfase a ser utilizado (Figura 5).

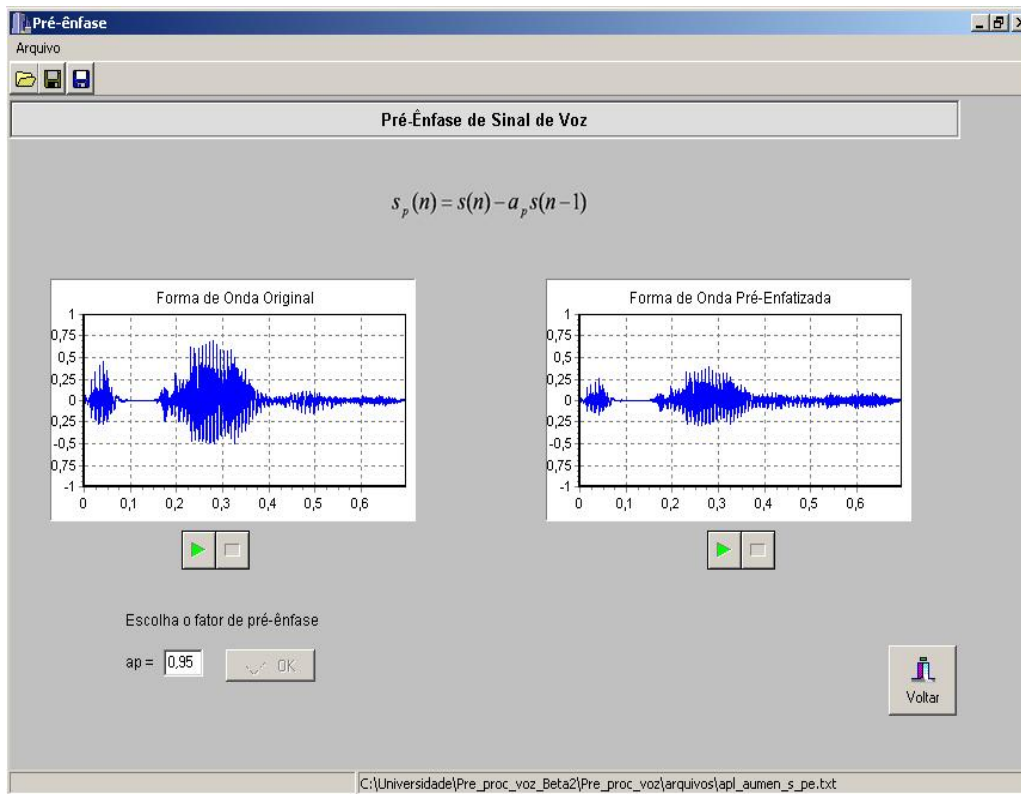


Figura 5 - Tela para realização da pré-ênfase.

A forma de onda lida e a forma de onda pré-enfatizada são apresentadas, podendo ser ouvidas, o que possibilita uma comparação entre o arquivo de áudio original e a sua reprodução após a aplicação do filtro.

Nessa etapa o arquivo de áudio pré-enfatizado é gravado e as amostras são disponibilizadas em um arquivo *.txt*.

6.4 Janelamento do Sinal

No aplicativo desenvolvido, é possível escolher o percentual de superposição das janelas, o tamanho do bloco (16 a 32ms) e o tipo de janela. As seguintes janelas podem ser aplicadas:

- Janela de Hamming.
- Janela de Hanning.
- Janela Retangular.

Após a escolha dos parâmetros, são apresentados a taxa de amostragem do sinal e o número de amostras por bloco.

Ao final do janelamento, um arquivo com extensão *.txt* é automaticamente gerado, o que possibilita a obtenção dos parâmetros temporais. Na Figura 6 é apresentada a etapa de realização do janelamento.

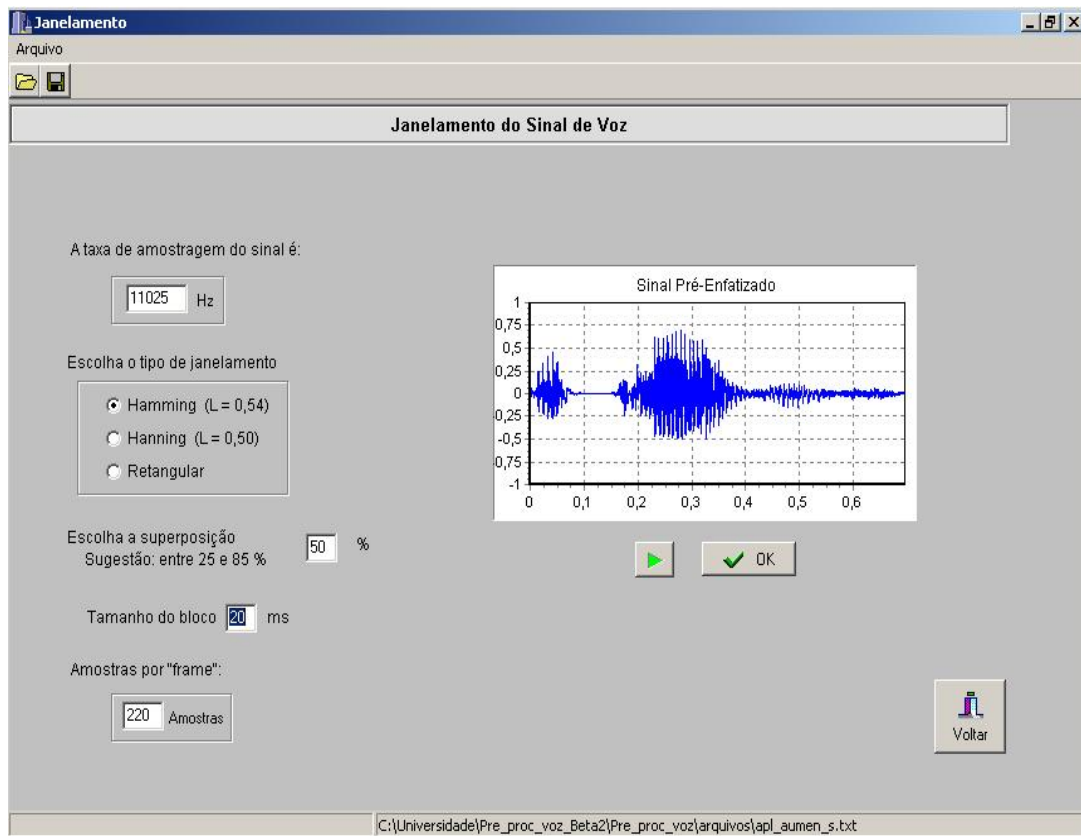


Figura 6 - Tela para realização janelamento do sinal de voz.

6.5 Análise dos Parâmetros Temporais

A análise no domínio do tempo é feita no menu de parâmetros temporais. Na Figura 7 é apresentada a tela de visualização dos parâmetros e dos correspondentes níveis de cada parâmetro por quadro.

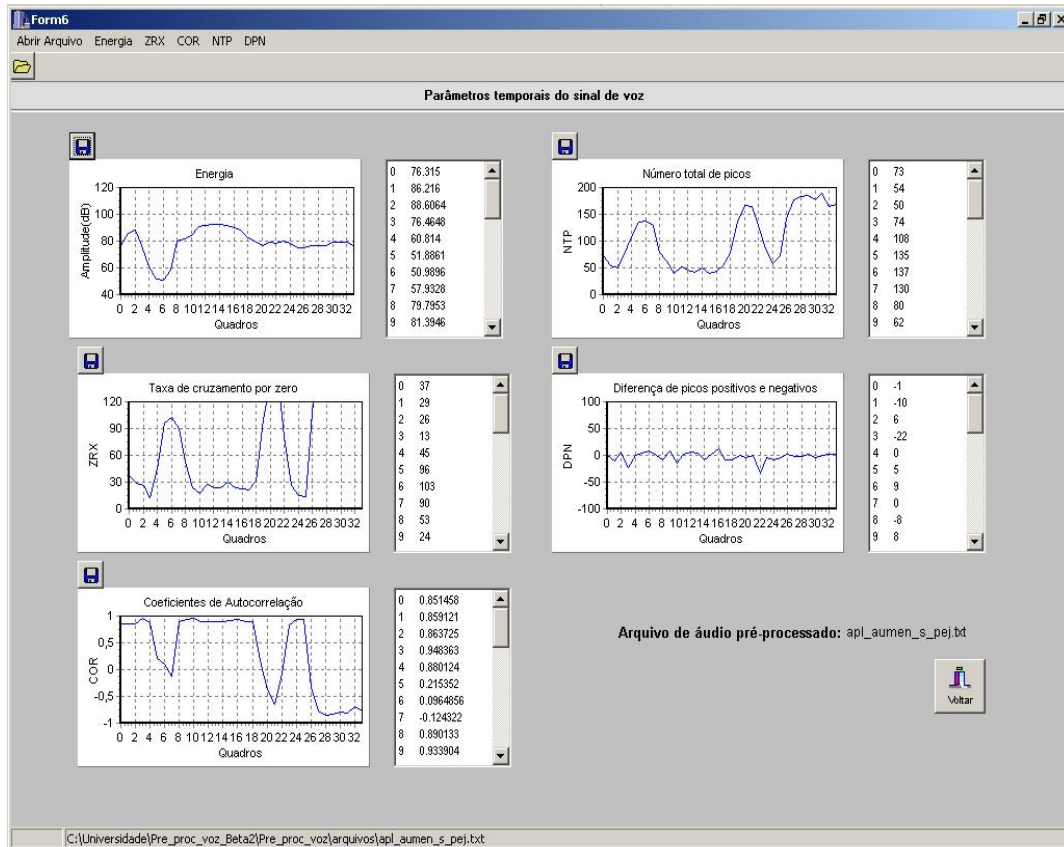


Figura 7 – Tela para análise de parâmetros temporais.

Os resultados obtidos, referentes à cada um dos parâmetros, podem ser gravados em um arquivo para posterior análise.

7. CONSIDERAÇÕES FINAIS

A comunicação vocal homem-máquina se caracteriza como uma área de pesquisa bastante promissora, cuja gama de aplicações tende a ampliar-se a cada dia, com o usuário buscando um meio de comunicar-se com a máquina que seja cada vez mais atrativo e eficiente. Dentro desse contexto, a utilização de um ambiente computacional que facilite o aprendizado de conceitos fundamentais, bem como possibilite auxiliar na realização de diversas tarefas na área da comunicação vocal homem-máquina, se constitui em uma ferramenta de ensino/aprendizagem bastante atraente para as disciplinas no contexto do processamento e transmissão digital de sinais de voz.

O ambiente computacional desenvolvido realiza as etapas de pré-processamento como também obtém informações importantes de um sinal de voz, a exemplo da extração de parâmetros temporais, que são fundamentais para diversas aplicações na área da comunicação vocal homem-máquina.

O interesse principal do aplicativo é o incentivo ao ensino/aprendizagem na área de processamento digital de sinais de voz, fornecendo uma ferramenta prática que possa permitir que estudantes e pesquisadores tenham um ambiente de desenvolvimento diretamente voltado a aplicações da área da comunicação vocal homem-máquina. Dentro dessa proposta, o ambiente ainda irá incorporar outras funcionalidades que permitirão a realização de tarefas básicas de reconhecimento e síntese de voz, de forma a se constituir em um software didático para disciplinas no âmbito de processamento digital de sinais.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] RABINER, L. R. ; SHAFER, R. W. **Digital Processing of Speech Signals**. New Jersey: Prentice Hall, 1978.
- [2] DELLER Jr. R. ; PROAKIS, J. G. ; HANSEN, J. H. L. **Discrete-time Processing of Speech Signals**. London: Macmillan Publishing Co., 1993.
- [3] CHU, M. S.; GOEL, V. E.; POTAMIANOS, M. G. **Efficient likelihood computation in multi-stream HMM based audio-visual speech recognition**. In: Proceedings of Proc. Int. Conf. Spoken Lang. Process., Jeju Island, Korea, 2004.
- [4] MARCHERET, E.; VISWESWARIAH, K.; POTAMIANOS, M. G. **Speech activity detection fusing acoustic phonetic and energy features**. In: Proceedings of Proc. Europ. Conf. Speech Comm. Technol. (Interspeech), Lisboa, Portugal, 2005.
- [5] FECHINE, J. M. **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística**. Tese de Doutorado em Engenharia Elétrica. Universidade Federal da Paraíba, Paraíba, 2000.
- [6] CAMPBELL, J. P., **Speaker Recognition: A Tutorial**. Proceedings of the IEEE, Vol. 85, 1997.
- [7] MADEIRO, F. ; VILAR, R. M. ; FECHINE, J. M. ; AGUIAR NETO, B. G., **A Self Organizing Algorithm for Vector Quantizer Design Applied to Signal Processing**, International Journal of Neural Systems, v. 9, n.2, p. 219-226, 1999.
- [8] PETRY, A.; ZANUZ, A.; BARONE, D. A. C. **Utilização de técnicas de processamento digital de sinais para identificação automática de pessoas pela voz**. Rio Grande do Sul, 2002.
- [9] VERGIN, R.; O'SHAUGHNESSY, D., **On the Use of Some Divergence Measures in Spekaer Recognition**. International Conference on Acoustics, and Signal Processing, 1999.
- [10] DEITEL, H. M. ; DEITEL, P. J. **C++ How to Program**. New Jersey: Prentice Hall, 2003.

A COMPUTATIONAL ENVIRONMENT FOR AID TO EDUCATION IN THE AREA OF DIGITAL PROCESSING OF SPEECH SIGNALS

Abstract: *This work deals with the study, development and implementation of software for aid to education and the research in the area of digital processing of speech signals. The tool in development consists in a constructed computational environment in C++ programming language, capable to carry through essential tasks of daily pay-processing and extraction of secular parameters of speaks, common to all research in the context of the vocal man-machine communication. Beyond the production of the applicatory one, the project also involves the study of the daily pay-processing techniques and processing of voice signals.*

Key-words: *Education and research in the area of digital processing of speech signals, daily pay-processing techniques.*