MONITORAMENTO DA VELOCIDADE DA FALA EM PESSOAS PORTADORAS DE DEFICIÊNCIA AUDITIVA

Suzete CORREIA (1); Silvana C. COSTA (2); Fagner ASSIS (3); Hanniere FALCÃO (4); Náthalee ALMEIDA (5).

(1) Centro Federal de Educação Tecnológica da Paraíba-CEFET-PB,
Coordenação de Telecomunicações, Av. Primeiro de Maio, 720 – Jaguaribe – João Pessoa - PB
Fone: (83) 3208-3055, e-mail: suzete@cetefpb.edu.br
(2) e-mail: silvana@cetefpb.edu.br
(3) e-mail: fagnertelecom@gmail.com
(4) e-mail: hanniereheim@yahoo.com.br
(5) e-mail: nathalee.telecom@gmail.com

RESUMO

Pessoas portadoras de deficiência auditiva apresentam alterações na fala em aspectos como: intensidade, freqüência fundamental, ressonância, velocidade e ritmo, entre outros. Técnicas de processamento digital de sinais têm sido usadas para avaliar o comportamento das vozes de pessoas com deficiência auditiva e proporcionar ferramentas de apoio ao melhoramento da qualidade vocal. Devido à falta de uma realimentação auditiva, uma realimentação visual computacional pode ser fornecida para monitoramento da fala em terapias vocais, com auxílio de um especialista. Neste trabalho, propõe-se a implementação de um algoritmo para detecção de início e fim de palavras, baseado em medidas de entropia do sinal de voz. O algoritmo implementado será utilizado para medir a velocidade da fala em pessoas com deficiência auditiva. Espera-se, assim, fornecer uma ferramenta de apoio computacional à melhoria da qualidade vocal, contribuindo com o processo de inclusão social destes indivíduos.

Palavras-chave: entropia, deficiência auditiva, processamento digital de sinais de voz.

1. INTRODUÇÃO

Em um mundo cada vez mais dinâmico onde a comunicação surge como uma necessidade primária, a fala exerce um papel fundamental.

No âmbito social, a comunicação verbal é uma via de inclusão, canal de comunicação eficiente e de entendimento entre as pessoas. A fala é, por assim dizer, a ponte por onde o entendimento vai e vem.

Muitas são as formas de comunicação e todas elas trouxeram grandes contribuições para o desenvolvimento da humanidade ao longo dos tempos. Contudo, a fala ainda é considerada a expressão mais completa e particular entre todas as formas de comunicação humana.

Geralmente, pessoas com deficiência auditiva apresentam complicações na fala devido à falta de percepção e monitoramento do som por elas produzido.

A surdez pode provocar no indivíduo um grave bloqueio comunicativo, prejudicando a sua integração com a sociedade. A criança surda sofre dificuldades escolares e o adulto surdo encontra grandes obstáculos ao tentar se inserir no mercado de trabalho. Esses indivíduos necessitam ainda de uma terapia que melhore a qualidade da fala que, em um deficiente auditivo, apresenta algumas características diferentes de um ouvinte em condições normais (COSTA, 2004).

Alguns estudos sobre o tema mostram que parâmetros importantes na produção da fala, tais como, velocidade, ritmo, freqüência fundamental (correlato perceptual- *pitch*), intensidade, articulação, respiração, ressonância e inteligibilidade da voz, podem ser afetadas se o indivíduo não possuir uma realimentação auditiva adequada.

Técnicas de processamento digital de voz são recursos bastante relevantes que podem ser utilizados na modelagem acústica do sinal de voz. A partir do modelamento obtêm-se recursos para o monitoramento acústico do sinal de voz.

A idéia básica dos sistemas de apoio ao deficiente auditivo é fornecer-lhe algum tipo de realimentação visual e/ou tátil de forma a suprir a falta de realimentação auditiva, que dificulta a sua oralização (BARROS, 2004 apud COSTA, 2004).

Este trabalho trata da implementação de um algoritmo para monitoramento da velocidade da fala baseado em medidas de entropia (medida de informação contida em uma mensagem) do sinal de voz, que visa oferecer mais uma ferramenta computacional de apoio à terapia vocal, para melhoria da qualidade vocal em pessoas portadoras de deficiência auditiva. O emprego da entropia tem a vantagem de ser mais robusto às influências do ruído, quando comparado aos métodos mais tradicionais que utilizam a energia do sinal.

Métodos baseados em medidas de energia do sinal de voz sofrem as variações ambientais. O nível de ruído do ambiente onde o software é utilizado pode não ser o mesmo do ambiente de implementação, provocando mudanças no desempenho. Torna-se necessário, portanto, ajuste dos limiares de energia para adaptar o algoritmo às mudanças ambientais. Já os métodos baseados na medida de entropia do sinal de voz dispensam esses ajustes, sendo mais robusto às variações ambientais. A vantagem do uso da entropia, então, está na sua reduzida sensibilidade à presença de ruído ou ausência de sinal (KHURRAM et al, 2002).

A aplicação do conceito de entropia para o problema de detecção da voz é baseada na suposição de que o espectro do sinal é mais organizado durante segmentos de voz do que segmentos de ruído, para o caso de análise no domínio da freqüência. (RENEVEY e DRYGAJLO, 2001).

2. PRODUÇÃO NATURAL DA FALA

A voz é um sinal produzido como resultado de várias transformações que ocorrem em diferentes níveis: semântico, lingüístico, articulatório, e acústico. Podem ocorrer desconformidades no sinal devido às diferenças anatômicas relacionadas ao trato vocal – características inerentes – e relacionadas ao movimento dinâmico do trato vocal, ou seja, a forma como a pessoa fala – características instruídas. As diferenças nessas transformações aparecem como diferenças nas propriedades acústicas do sinal de voz. (FECHINE, 2000).

A fala é produzida, basicamente, através da liberação de ar dos pulmões para o trato vocal, que é formado por cavidades e pelos articuladores. O trato vocal compreende a região que vai da glote até os lábios. Como não é capaz de produzir um som com intensidade suficiente para que seja perfeitamente audível, a laringe

necessita da participação de estruturas como a faringe, a cavidade oral e a cavidade nasal, que como órgãos de ressonância, melhoram a qualidade da voz (RABINER &SCHAFER, 1978).

A Figura 1 mostra um diagrama de blocos da produção de voz humana, também denominado sistema fonte-filtro, onde as cordas vocais são consideradas a fonte sonora e o trato vocal, o filtro. O sinal resultante será o sinal acústico de voz humana (DELLER, PROAKIS & HANSEN, 1993).

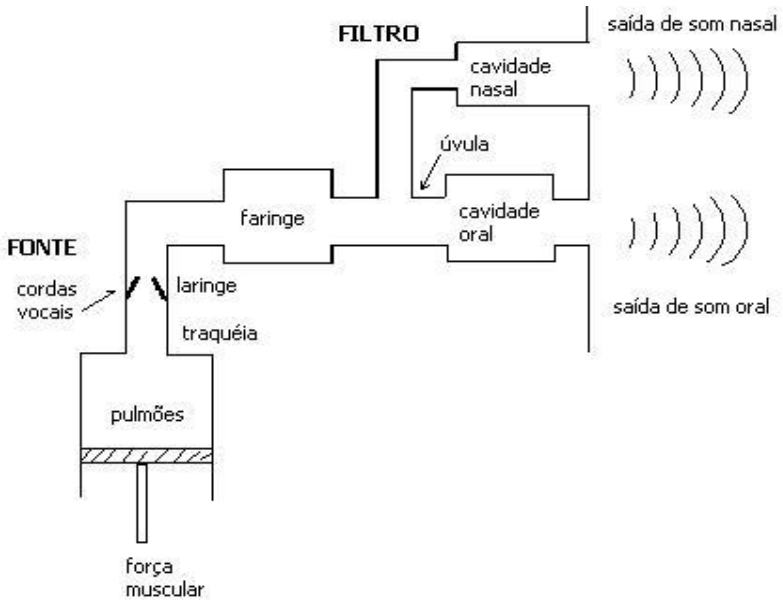

Figura 1 - Diagrama de blocos do mecanismo de produção de voz humana

A fala é produzida através da liberação de ar dos pulmões para o trato vocal, formado basicamente por cavidades e órgãos articuladores que começa na abertura entre as cordas vocais, ou glote e termina nos lábios (Figura 1). O trato vocal é uma estrutura tubular pela qual passa o fluxo de ar vindo dos pulmões, o qual é modulado nas cordas vocais. Sua principal função é modular o espectro de freqüência da onda sonora que vem das cordas vocais e promover constrições para a geração de certos tipos de som.

O trato nasal começa na úvula e termina nas narinas. Quando a úvula é abaixada, o trato nasal é acusticamente acoplado ao trato vocal para produzir os sons nasais da voz.

O ar é conduzido para fora dos pulmões pela traquéia, passando pela laringe, onde estão as cordas vocais. O espaço compreendido entre as cordas vocais é chamado de glote, e sua abertura pode ser controlada movimentando-se as cartilagens aritenóide e tireóide. É lá que o fluxo contínuo de ar dos pulmões é geralmente transformado em vibrações rápidas e audíveis durante a fala. Isso é feito pelo fechamento das cordas vocais, que causa um aumento gradativo da pressão atrás das mesmas, que acaba por fazer com que elas se abram repentinamente, liberando a pressão, para então tornarem a se fechar. Este processo produz uma seqüência de pulsos cuja freqüência é controlada pela pressão do ar e pela tensão e comprimento das cordas vocais (freqüência fundamental). Os sons assim produzidos são chamados de vozeados, ou sonoros, e normalmente incluem as vogais; caso contrário, são chamados não-vozeados, ou surdos.

No processo natural de produção da fala, alguns fatores são diretamente envolvidos com a qualidade e a inteligibilidade do som. Esses fatores são: a respiração, a articulação, a ressonância e a realimentação auditiva. Esta última recebe atenção especial no âmbito das pesquisas envolvendo pessoas com deficiência auditiva.

A realimentação auditiva é a percepção do som produzido por si mesmo. Ela que dá suporte e orienta o falante no processo de produção da fala.

3. CARACTERÍSTICAS DAS VOZES DE PESSOAS COM DEFICIÊNCIA AUDITIVA

Geralmente, pessoas com deficiência auditiva têm associadas complicações na fala devido à falta de percepção e monitoramento do som por elas produzido. A falta de auto-percepção implica no mau posicionamento da língua, dentes e lábios, o que compromete a articulação.

Os estudos voltados para análise acústica em deficientes auditivos, indicam que há alterações na velocidade, ritmo e entonação, características importantes para qualidade vocal. O padrão respiratório encontra-se alterado, apresentando uma dificuldade em administrar o fluxo aéreo a expelir ar antes de concluir a fala (COSTA, 2004).

O sujeito desprovido de realimentação auditiva tem dificuldade para aprendizagem na utilização do complexo respiração-fonação-articulação (ITON, M 1982).

Técnicas de processamento digital de sinais têm sido desenvolvidas para avaliar o comportamento das vozes de pessoas com deficiência auditiva e proporcionar ferramentas de apoio ao melhoramento da qualidade vocal (TUJAL, 1998).

A realimentação visual se revela um método bastante promissor no processo de oralização do surdo, ou de melhoria da qualidade da fala em pessoas com vozes desordenadas, tanto pela possibilidade de transmitir as variações rápidas das informações do sinal de voz, como por permitir a discriminação de uma quantidade maior de informações (TUJAL, 1998).

A velocidade e o ritmo da fala do deficiente auditivo geralmente estão alterados. A velocidade é lenta demais e o padrão de ritmo inapropriado, prejudicando a inteligibilidade. O tempo de fonação nas sílabas é um problema freqüente, sendo necessário encurtá-lo ou aumentá-lo, dependendo do caso, para que a velocidade da fala seja melhorada (WILSON, 1993).

Robb e Pang Ching (1991), em estudo com 26 deficientes auditivos adultos com perda auditiva severa e profunda e 13 sujeitos ouvintes, verificaram que a duração absoluta média na produção de frases do grupo de deficientes auditivos é 900 ms mais longa do que no grupo de ouvintes. Também observaram que no grupo de deficientes auditivos com perda profunda, a produção de frases é mais longa do que no grupo com perda severa.

Fletcher e Daly (1976), em comparação entre dois grupos: o primeiro de deficientes auditivos entre 07 e 21 anos e o segundo comparativo de ouvintes, constatam a média de 95 palavras por minuto para os deficientes auditivos e 173 palavras por minuto para os ouvintes.

Boone (1966), considera que a velocidade da fala é reduzida por duas razões: o prolongamento d as vogais e longas pausas entre as palavras. A velocidade lenta de fala pode estar associada à falta de controle respiratório. Podem acontecer sucessivas inspirações a cada poucas palavras sem conseguir terminar uma frase. Como procedimento terapêutico deve-se adequar a respiração à fala, seguida de controle muscular e acentuação. A criança é conscientizada de sua velocidade e da velocidade adequada, percebendo o contraste das situações.

Segundo Wilson (1993), uma criança com perda auditiva de leve a moderada pode ter dificuldade somente com o equilíbrio da ressonância oral-nasal. Por outro lado uma criança surda pode ter não só problemas de ressonância, como outros problemas envolvendo altura, intensidade, qualidade laríngea, velocidade e ritmo de fala. Uma criança surda normalmente tem dificuldade em manter a velocidade e o ritmo de fala em níveis adequados. Alguns falam com uma velocidade excessivamente baixa e padrões de ritmo inadequados.

4. MONITORAMENTO E CONTROLE DA VELOCIDADE E RITMO DA FALA

Como a velocidade e o ritmo da fala estão relacionados ao tempo de fonação, é possível utilizar-se de alguns recursos de processamento digital de sinais para auxiliar na sua monitoração. O objetivo deve ser sempre o de obter uma voz agradável ao ouvinte e com uma boa inteligibilidade.

Sugere-se, para monitorar a velocidade da fala, a utilização de algoritmos para detecção dos intervalos de silêncio presentes no sinal de voz. Isto pode ser feito encontrando-se os valores de tempo correspondentes às pausas entre palavras de uma mesma frase ou em pausas existentes numa mesma palavra. De posse desses valores pode-se fazer uma comparação da elocução atual com uma elocução considerada padrão/normal.

Essa comparação também pode ser feita de forma visual: uma tela mostrando as formas de onda correspondentes às duas elocuções, possibilitando ao terapeuta e ao paciente monitorar a velocidade da fala durante a elocução.

A velocidade e ritmo da fala reúnem, em sua avaliação, aspectos subjetivos. Aliando-se aspectos objetivos e subjetivos de forma eficiente, pode-se chegar a resultados satisfatórios.

Para a detecção de intervalos de silêncio numa elocução, destacam-se alguns algoritmos mais comuns que usam a energia a curto intervalo de tempo (RABINER & SCHAFER, 1978; FECHINE, 2000) e outras propostas que usam a medida de entropia (KHURRAM et al, 2002).

Khurram et al (2002) propõem um algoritmo para detecção de início e fim ou limites de uma palavra/sentença usando Entropia (medida da informação contida numa mensagem). O método consiste em usar uma função de restrição baseada na entropia entre segmentos de voz e ruído de fundo. Segundo os autores, este método apresenta melhor desempenho frente ao ruído que os métodos baseados na energia do sinal. Um limiar adaptativo é usado para a determinação dos segmentos candidatos que estão submetidos às restrições da palavra/sentença. O algoritmo é descrito a seguir.

4.1 Detecção de Início e Fim de Palavras – descrição do algoritmo implementado

A entropia é calculada pela Equação (01), em que p_k representa a probabilidade de ocorrência de uma determinada amostra, calculada através da função distribuição de probabilidade. Para determinar a distribuição de probabilidade em cada bloco de amostras, um histograma de N pacotes é construído. O histograma é normalizado para satisfazer as propriedades estatísticas da função de distribuição cumulativa (f.d.c.). A seleção do número de pacotes (*N*) é uma troca entre sensibilidade e carga computacional. Geralmente, *N* pode ser escolhido na faixa de 50-100.

$$H = -\sum_{k=1}^{M} p_k \cdot \log_2(p_k) \qquad \text{Eq. [01]}$$

O algoritmo pode ser facilmente implementado para ser utilizado em tempo real com uma unidade de atraso de um bloco de amostras. O perfil de entropia, ξ , para *m* segmentos do sinal de voz, é dado por (KHURRAM et al, 2002):

$$\xi = [H_1 \; H_2 \; \ldots \; H_m], \qquad \text{Eq.[02]}$$

Este perfil de entropia pode ser, então, usado para encontrar um limiar apropriado γ para determinação da existência de regiões de voz dentro da sentença completa. Um limiar adequado neste caso, será tomado na forma:

$$\gamma = \frac{\max(\xi) - \min(\xi)}{2} + \mu \cdot \min(\xi), \quad \mu > 0 \qquad \text{Eq. [03]}$$

O limiar é, portanto, escolhido um pouco maior que o perfil médio da entropia e sua escolha minimiza a influência excessiva do ruído de fundo. Uma vez que o limiar tenha sido determinado, qualquer valor acima deste limiar é considerado como voz e qualquer coisa abaixo do limiar ou é silêncio, ou ruído, isto é

$$\xi' = \begin{cases} \xi_i & \text{se } \xi_i \geq \gamma \\ 0 & \text{c.c.} \end{cases}; \; i = 1, 2, \; \ldots m \qquad \text{Eq. [04]}$$

A Figura 2 mostra um diagrama em blocos destacando as etapas fundamentais do algoritmo proposto por Khurram et al (2002). As Equações (01)-(04) simulam as etapas de 2 a 5.

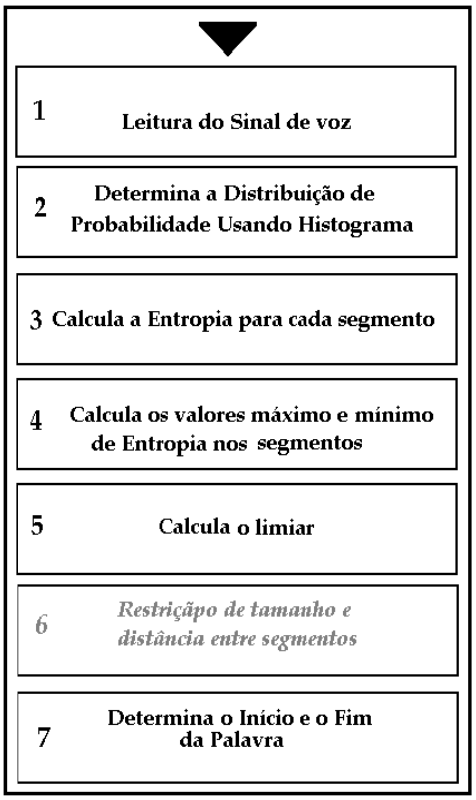

Figura 2 - Diagrama de blocos do Algoritmo utilizando entropia de Shannon.

Com base na posição na qual foi encontrado o primeiro valor de entropia acima do limiar determinado – início da palavra – o algoritmo associa o segmento a um valor correspondente de tempo. O processo de comparação com o limiar continua, mas no sentido de observar em que instante de tempo o valor da entropia cai abaixo do limiar especificado. Satisfeita esta condição, encontra-se o fim da palavra.

5. RESULTADOS OBTIDOS

Os sinais de voz a serem processados foram obtidos com utilização de um microfone conectado a um computador pessoal provido de placa de som e por meio do software *GoldWave*, versão demo.

A Figura 3 permite visualizar o comportamento no tempo do sinal de voz da palavra "bola", um dos sinais utilizados para análise. O sinal foi gravado utilizando-se uma freqüência de amostragem de 11025 Hz, resolução de 16 bits, mono. Foi utilizado um segmento de análise de 220 amostras (cerca de 20 ms) e N=50.

Na Figura 4 tem-se o gráfico correspondente a entropia segmental da palavra bola. Após a determinação e aplicação do limiar, foram encontrados os seguintes valores:

- Valor máximo de H = 3.190905; Valor mínimo de H = 0.000379;
- Valor médio de H = 1,445754; Limiar = 1,595810
- Início: 0,279 segundos; Fim: 0,758 segundos

Na Figura 5, pode-se visualizar o trecho em destaque correspondendo à nova delimitação da palavra /bola/, resultante do algoritmo implementado. Vê-se uma boa delimitação do sinal. O teste de escuta, para avaliação do desempenho do algoritmo, não detectou perda significativa no sinal.

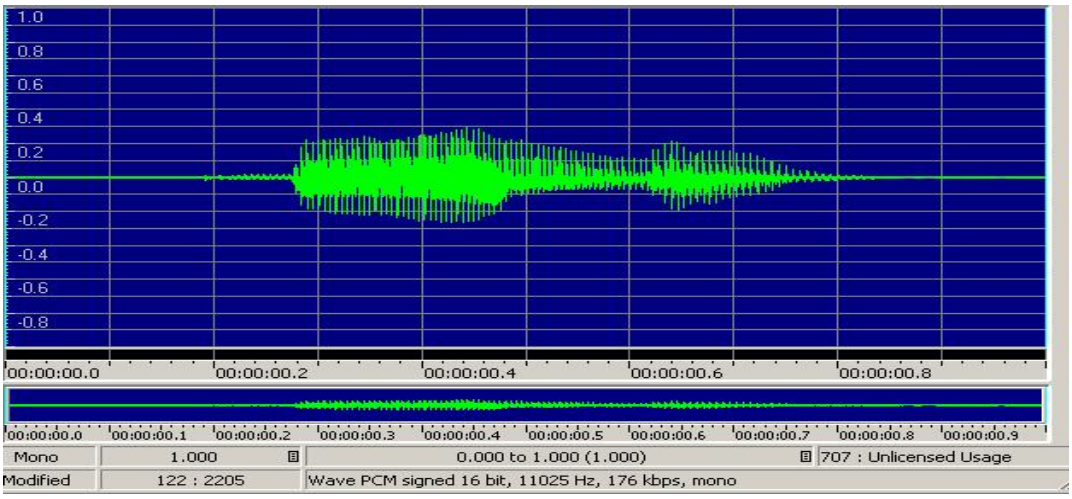

Figura 3 – Forma de onda do sinal de voz da palavra /bola/.

A Figura 4 mostra o gráfico da entropia por segmento do sinal de voz da palavra "Bola".

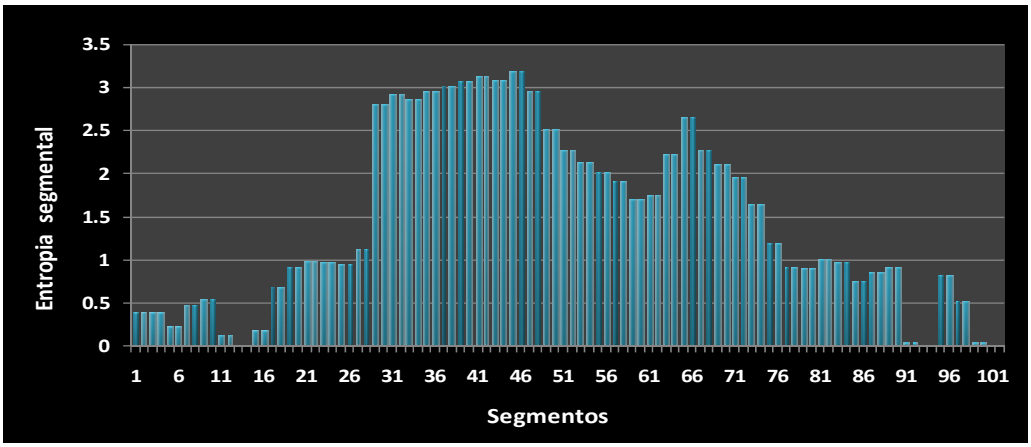

Figura 4 - Gráfico da entropia segmental correspondente ao sinal da palavra /bola/.

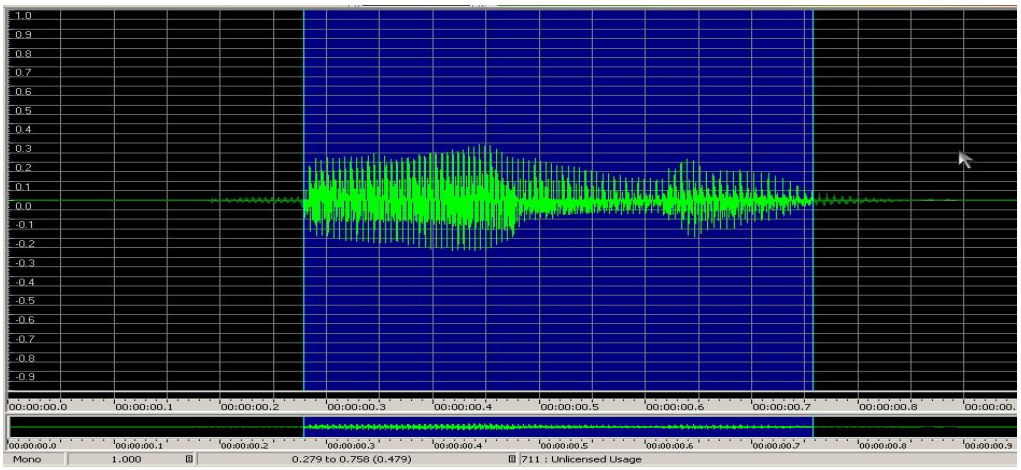

Figura 5 – Forma de onda da palavra /bola/ com as delimitações de início e fim.

O gráfico da figura 4 foi obtido a partir dos valores tomados do algoritmo, na etapa (3) - calculo da entropia para cada segmento. O mesmo simula satisfatoriamente o comportamento do sinal da palavra, atribuindo valores para os segmentos onde existe sinal e impondo valor zero às partes de silêncio (ruído).

O algoritmo deve ser capaz de determinar os limites da sentença e rejeitar períodos de silêncio (falsos limites). Por isso, é fundamental observar outros critérios como: tamanho e distância entre palavras. De acordo com a etapa 6 do diagrama de blocos do algoritmo proposto (Figura 2).

O sinal da palavra /aplausos/ na Figura 6 e o gráfico da entropia na Figura 7, mostram um segmento do sinal, que se assemelha a uma pausa. O trecho em destaque, na realidade, corresponde a uma transição entre fonemas na mesma palavra.

Figura 6 – Sinal de voz da palavra "aplausos"

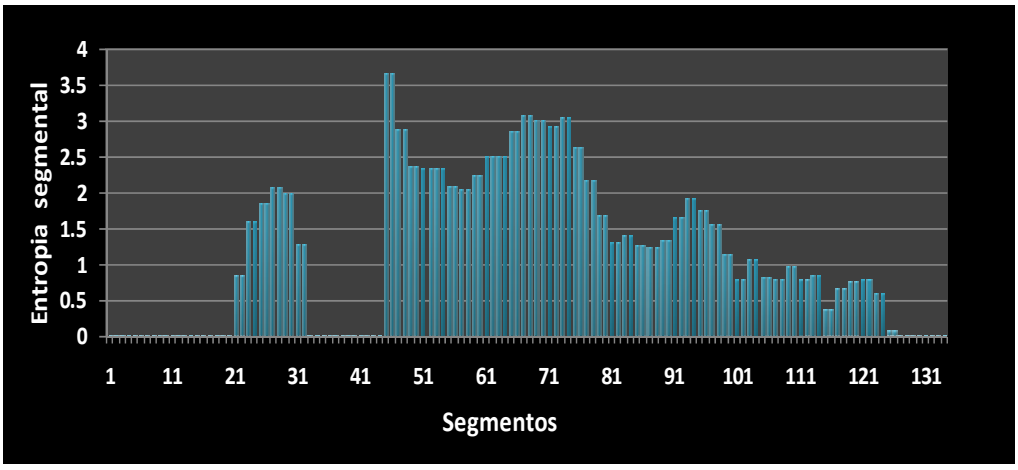

Figura 7 – Gráfico da entropia segmental da palavra /aplausos/.

No caso da palavra /aplausos/, por exemplo, o algoritmo deve ser ajustado para não detectar o final do fonema /a/ com o /p/, onde aparece uma pausa dentro da palavra, como se fosse o final da palavra.

Algumas considerações precisam ser levadas em conta para minimizar os erros decorrentes da confusão na detecção de amostras que não sejam de voz, consideradas como tal pelo algoritmo. As medidas do tamanho do segmento e de intra-segmentos são utilizadas. A distância entre o início e o fim do segmento detectado não deve ser menor que o intervalo de duração do fonema mais curto da língua em questão, ou do vocabulário que o sistema utiliza.

Projetar um sistema sem levar em conta o efeito do ruído de fundo, pode contribuir para uma queda em seu desempenho, quando operado em condições adversas àquelas onde foi implementado e testado.

Dessa forma, sistemas que sejam robustos ao ruído, adaptando-se às condições do ambiente do usuário devem ser implementados. O algoritmo escolhido deve ser aquele que proporcione melhores condições para a obtenção de uma voz clara e agradável.

Na detecção de início e fim de palavras, busca-se eliminar as pausas usuais no início e fim das elocuções. Com isso, há economia de memória em armazenamento dos sinais, além de diminuição de tempo de processamento. Assim, sistemas que necessitam de processamento em tempo real, tornam-se mais eficientes.

6. CONCLUSÃO

O algoritmo implementado mostrou-se bastante eficiente na determinação do início e fim de palavras e bastante robusto às variações ambientais.O mesmo foi utilizado em ambientes diversos, com várias fontes de ruído e não sofreu variações significativas em seu desempenho.

No estágio atual de implementação, as elocuções que apresentam pausas dentro da própria palavra, como /aplausos/, por exemplo, precisam de uma atenção maior. Como já destacado nas seções anteriores, é necessário verificar a distância intra-segmentos e tamanhos de segmentos válidos para diminuir os prováveis erros de detecção.

Testes preliminares já estão sendo feitos e os resultados encontrados são bastante promissores. Após a implementação da etapa 6 do esquema proposto, pretende-se implementar outros algoritmos e fazer uma comparação de desempenho entre os mesmos.

REFERÊNCIAS

BOONE, D. R. Modification of the Voices of Deaf Children. Volta Rev., 68: 686-92, 1966.

COSTA, S. C. **Processamento digital de sinais aplicado à oralização de surdos.** 2004. 59p. Projeto de pesquisa (Doutorado em Engenharia Elétrica) – Universidade Federal de Campina Grande, Campina Grande. Agosto, 2004.

CUKIER, Sabrina; CAMARGO, Zuleika. Abordagem da Qualidade Vocal em um Falante com Defeiciência Auditiva: Aspectos Acústicos Relevantes do Sinal de Fala.VER. CEFAC, v. 7, n. 1, 93-101, São Paulo, Jan-Mar, 2005.

DELLER, Jr. R.; PROAKIS, J. G. and HANSEN, J. H. L. **Discrete-time Processing of Speech Signals**, Macmillan Publishing Co., 1993.

FECHINE, J. M., **Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística.** Tese de Doutorado, Universidade Federal da Paraíba, 2000**.**

FLETCHER, S.G.; DALY, D.A. **Nasalance in Utterances of Hearing-impaired Speakers**. Journal of Communication Disorders, 37(1), 63-73, 1976.

ITON, M et al. **Selected Aerodynamic Characteristics of Speech Tasks**. Folia Phoniatrica, 1982.

KHURRAM, W.; WEAVER K..; SALAN, F. M.: An Entropy based Robust Speech Boundary Detection Algorithm for Realistic Noisy Environments; 2003 IEEE-INNS Joint International Conference on Neural Networks, July 20-24, 2003.

MONSEN, R. B. Voice Quality and Speech Inteligibility among Deaf Children. Am. Ann. Deaf., 128 (1): 12-19, 1983.

RENEVEY, P. DRYGAJLO, A.: **Entropy based voice activity detection in very noisy conditions**, In EUROSPEECH-

RABINER L. R. & SCHAFER R. W., **Digital processing of speech signals**. New Jersey: Prentice-Hall, 1978.

ROBB, M.P. and PANG Ching, G.K. Relative Timing Characteristics of Hearing Impaired Speakers. J. Acoustic Soc. Am. 91:2954-2960, 1992.

TUJAL, P. M. O., **Auxílio Visual à Oralização de surdos**. Tese de Mestrado, Universidade Federal do Rio de Janeiro, 1998.

WILSON, D.K. **Problemas de Voz em Deficientes Auditivos** - Problemas de Voz em Crianças. 3a ed., Manole, São Paulo, 301-327,1993.