



**Análise do emprego de técnicas de remoção de ruído como
pré-processamento de algoritmos de detecção de atividade
de voz**

PROPOSTA DE TRABALHO DE GRADUAÇÃO

Aluno: Renato Quirino de Albuquerque (rqa@cin.ufpe.br)
Orientador: Carlos Alexandre Barros de Mello (cabm@cin.ufpe.br)

29 de Agosto de 2016

Introdução

Algoritmos de detecção de atividade de voz (VAD) são responsáveis por analisar o sinal de voz e detectar em quais segmentos (*frames*) ocorre presença ou ausência de atividade de voz humana. Para realizar esta tarefa, o algoritmo de detecção deve ser capaz de ignorar trechos sem atividade de voz mesmo para ambientes com alto índice de ruído. Esta é uma tarefa complexa que requer acurada estimativa do nível de ruído no ambiente e acurada capacidade de distinção entre fala e sons ambientes diversos.

De forma geral, algoritmos de VAD são formados por três partes principais; extração de características, etapa de decisão e amortecimento da decisão. Na etapa de extração de características, são gerados conjuntos de informações de características distintas (*features*), a partir do sinal de fala. Para alcançar bons resultados, deve-se utilizar características discriminantes e robustas a ruído [1]. Na etapa de decisão, são aplicadas regras que avaliam as *features* geradas e, como resultado, decide em quais frames ocorre atividade de voz e em quais ocorre ausência de voz. Na última etapa, é realizado refinamento das decisões tomadas na etapa de decisão, um novo conjunto de regras é aplicado a fim de aprimorar os resultados eliminando potenciais falsos positivos, comumente nomeados de Falsos Alarmes (FA) [2], frames sem atividade de voz que são classificados como tendo atividade de voz. Na Fig. 1 é possível verificar os principais componentes de um algoritmo VAD e os respectivos resultados para um sinal de exemplo.

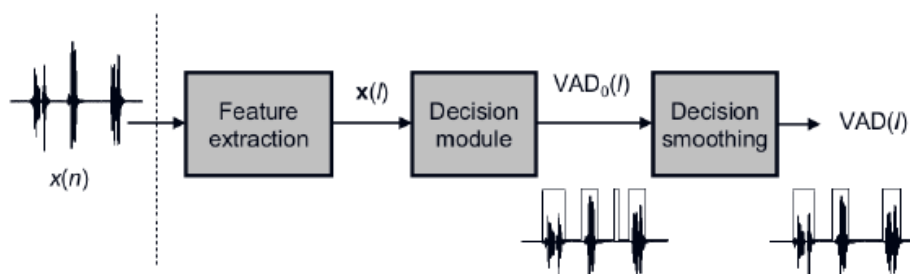


Figura 1. Principais componentes de um sistema VAD.

Algoritmos de detecção de atividade de voz desempenham papel fundamental em diversas aplicações de processamento de sinal de fala, pois em geral são aplicados como processamento inicial de separação dos segmentos de voz que serão utilizados nos passos seguintes da aplicação. Além disso, a utilização de algoritmos de VAD reduz a quantidade de informação necessária a ser processada (ou transmitida) e assim reduzir o tempo de processamento em um sistema de comunicação. Serviços de comunicação móvel (Freeman *et al.* 1989) [3], transmissão em tempo real de fala na internet (Sangwan *et al.*, 2002) [3], Redução de ruídos para aparelhos auditivos digitais (Itoh and Mizushima, 1997) [3], avaliadores de qualidade de fala, detectores de fala, reconhecedor de voz/fala, transmissão descontínua, transmissão de fala em tempo real, remoção/redução de ruído, cancelamento de eco, entre outras, são algumas das diversas aplicações que necessitam de algoritmos de detecção de fala ativa capaz de separar os

trechos ativos e que reduza a quantidade de informação processada. Tomando como exemplo, um reconhecedor de voz (ou fala) é crucial a necessidade de um VAD capaz de gerar quais os trechos que possuem voz ativa e, a partir destes, realizar o processamento necessário ao reconhecimento do falante ou do comando de voz. Codecs/Decoders de fala utilizam algoritmos de VAD para detectar trechos de silêncio e codificar ou decodificar a quantidade de informação necessária para compressão, ou descompressão.

A maior problemática ao utilizar VADs ocorre em ambientes com baixa Taxa Sinal-Ruído (SNR), nestes ambientes a discriminação entre sinal de voz e ruído é prejudicada, pois ambas as partes se confundem, dificultando as decisões do VAD. Na Fig. 2 é possível verificar o efeito da adição de ruído gaussiano a uma amostra de sinal de fala. No sinal com ruído, à direita no topo, não é possível estimar visualmente qual a separação entre as elocuições da gravação, contudo no espectrograma do mesmo sinal é visto que as baixas frequências do sinal se confundem com o ruído.

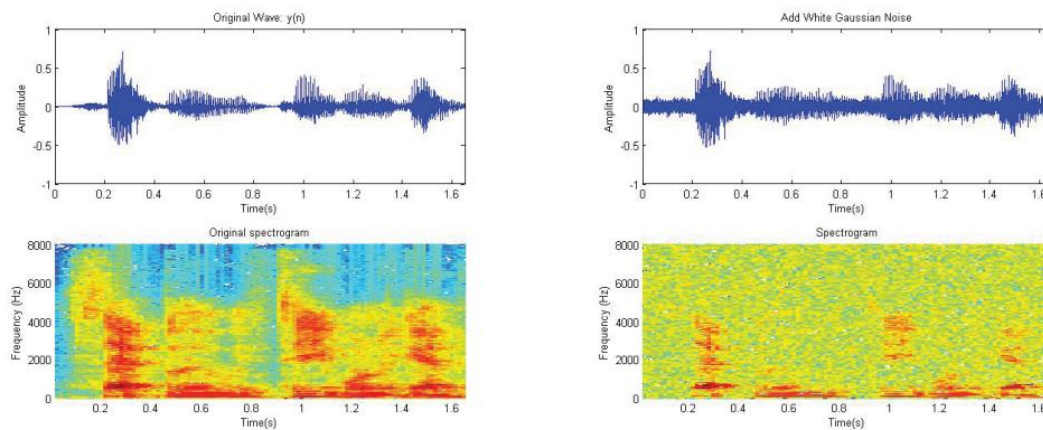


Figura 2. *Waveform* e espectrograma do sinal original à esquerda, *waveform* e espectrograma do mesmo sinal com adição de ruído gaussiano.

Para solucionar o problema de detecção de atividade de voz em baixo SNR podem ser aplicadas técnicas de remoção de ruído sobre o sinal degradado [4]. Tal solução visa reduzir o nível de ruído sem introduzir distorções na amostra de voz, e assim aumentar a distinção entre regiões de fala ativa e silêncio [5].

Objetivos

O presente trabalho tem como objetivo o estudo de algoritmos de detecção de atividade de voz com abordagens variadas, análise destes algoritmos sob diferentes níveis de SNR das amostras de entrada e a avaliação da acurácia alcançada destes algoritmos ao empregar técnicas de remoção de ruído.

Metodologia

Os algoritmos serão desenvolvidos utilizando a linguagem de programação Matlab. Caso existam soluções padronizadas dos algoritmos apresentados, estas serão utilizadas. Devido aos resultados alcançados ao aplicar diferentes técnicas de remoção de ruído, diferentes algoritmos de VAD serão utilizados para verificar quais as melhores combinações. Tabelas e gráficos expositivos serão apresentados e discutidos.

Cronograma

Atividade	Período															
	Agosto		Setembro				Outubro				Novembro				Dezembro	
Revisão bibliográfica	X	X	X	X	X	X	X	X	X	X						
Codificação			X	X	X	X	X	X	X	X						
Estudo do algoritmo (experimentos e teoria)					X	X	X	X	X	X	X	X				
Testes									X	X	X	X	X	X		
Avaliação dos resultados									X	X	X	X	X	X		
Preparação do relatório												X	X	X	X	X
Preparação da apresentação													X	X	X	X

Referências

- [1] P. C. Khoa, Noise Robust Voice Activity Detection, School of Computer Engineering, Nanyang Technological University, 2012
- [2] S. S. Meduri, R. Ananth, A Survey and Evaluation of Voice Activity Detection Algorithms, M.S. thesis, Department of Electrical Engineering, Blekinge Tekniska Hogskola, Karlskrona, Sweden, 2011.
- [3] M. Grimm, K. Kroschel, Voice Activity Detection. Fundamentals and Speech Recognition System Robustness in Robust Speech Recognition and Understanding, Vienna, Austria: I-Tech., 2007, ch. 5, pp. 460.
- [4] Jeffery J. Faneuff, D. Richard Brown, Noise Reduction and Increased VAD Accuracy Using Spectral Subtraction, Bose Corporation, The Mountain Framingham, Framingham, Worcester Polytechnic Institute, Worcester, 2003.
- [5] E. Verteletskaya, B. Simak, Enhanced spectral subtraction method for noise reduction with minimal speech distortion, Department of Telecommunication Engineering, Czech Technical University, Prague, Czech Republic, 2010.

Possíveis Avaliadores

Prof. Tsang Ing Ren

Prof. Daniel Carvalho Cunha