UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

PAULO HENRIQUE BEZERRA DE CARVALHO

CODIFICAÇÃO DE SINAIS DE VOZ HUMANA POR DECOMPOSIÇÃO EM COMPONENTES MODULANTES

São Luís
2003

PAULO HENRIQUE BEZERRA DE CARVALHO

CODIFICAÇÃO DE SINAIS DE VOZ HUMANA POR DECOMPOSIÇÃO EM COMPONENTES MODULANTES

Dissertação apresentada ao curso de Mestrado em Engenharia de Eletricidade da Universidade Federal do Maranhão para obtenção do título de Mestre em Ciências da Computação.

Orientador: Prof. Dr. Allan Kardec Duailibe Barros Filho

São Luís
2003

Carvalho, Paulo Henrique Bezerra de
Codificação de sinais de voz humana por decomposição em componentes modulantes / Paulo Henrique Bezerra de Carvalho. – São Luís, 2003.
- f.-
Dissertação (Mestrado em Engenharia de Eletricidade) – Curso de Ciência da Computação. Universidade Federal do Maranhão, 2003.
1. Software – Voz humana. 2. Sinais de voz.
I. Título
CDU 004.4:612.78

PAULO HENRIQUE BEZERRA DE CARVALHO

CODIFICAÇÃO DE SINAIS DE VOZ HUMANA POR DECOMPOSIÇÃO EM COMPONENTES MODULANTES

Dissertação apresentada ao curso de Mestrado em Ciências da Computação da Universidade Federal do Maranhão para obtenção do Título de Mestre em Ciências da Computação.

Aprovada em / /

BANCA EXAMINADORA

_____

Prof. Dr. Allan Kardec Duailibe Barros Filho
(Orientador)

_____

Prof. Dr. Juradir Nadal
(Membro da Banca Examinadora)

_____

Prof. Dr. Edson Nascimento
(Membro da Banca Examinadora)

A meus pais, irmãos, sobrinhos e em especial a Catherine, Beatriz e Eduardo.

AGRADECIMENTOS

A Deus pela minha saúde e felicidade.

Aos meus pais Carlos Henrique (in memoriam) e Marli Carvalho pela educação que me proporcionaram.

À minha esposa Catherine pela harmonia de nossas vidas, apoio e compreensão quando precisei.

Ao Professor Allan Kardec pela sua orientação segura, coerente e paciente durante todo o período de elaboração desta dissertação.

Aos colegas de trabalho que me ajudam nos momentos de ausência da empresa, em especial ao amigo Adney Teles.

Ao amigo Ozéas Lobato pelo companheirismo e colaboração.

"The whole of science is nothing more than a refinement of everyday thinking."

Albert Einstein.

RESUMO

Este trabalho propõe uma variação de codificador do sinal de voz baseada em dois conceitos: os formantes e as componentes modulantes do sinal. O método proposto de codificação extrai as componentes modulantes (amplitudes e freqüências instantâneas) para serem transmitidas. O método é baseado no fato de que a transmissão da voz pode ser substituída pelo envio de suas componentes modulantes AM-FM (amplitude modulation - frequency modulation). Desse modo, para o envio de tais componentes é utilizado o método LPC (linear predictive coding) para a determinação das freqüências correspondentes aos quatro primeiros formantes do espectro de voz na faixa de 4 kHz. Em seguida, através de uma função wavelet modificada de Gabor, são filtradas quatro faixas estreitas em torno desses formantes. Por último, utilizando-se as propriedades da transformada de Hilbert, são determinadas as componentes modulantes das faixas filtradas, ou seja, as amplitudes e freqüências instantâneas. O resultado final é a codificação de oito sinais, sendo quatro correspondentes às amplitudes instantâneas e quatro das freqüências instantâneas. Também é apresentada a recuperação da voz a partir dos oitos sinais e para a validação do método são utilizadas cinco amostras de voz humana onde são empregados testes de inteligibilidade das amostras após as suas respectivas recuperações. Os resultados obtidos mostraram que o método é factível de implementação em aplicações reais.

ABSTRACT

This work proposes an speech signal encoder variation based on two concepts: the formants and the modulating components of the speech signal. The method suggested for the codification extracts the modulating components (instantaneous amplitude and frequency) to be transmitted. The method is based on the fact that the transmission of the speech can be substituted by the transmission of its AM-FM modulating components (amplitude modulation - frequency modulation). Thus, to send such components, the LPC (linear predictive coding) method is used to determine the frequencies that correspond to the first four formants of the speech spectrum within a 4 kHz band. Then, through a modified Gabor's wavelet function, four narrow bands are filtered around the formants. Finally, the properties of the Hilbert transform are used to determine the modulating components of the filtered bands, in other words, the instantaneous amplitudes and frequencies. The final result is the codification of eight signals in which four of them correspond to the instantaneous amplitudes and the other four correspond to the instantaneous frequencies. It is also presented a recovery of human speech where tests of intelligibility of the samples are applied after their respective recoveries. The results obtained showed that the method is a promising technique to be implemented in actual applications.

LISTA DE TABELAS

LISTA DE FIGURAS

LISTA DE SIGLAS

AM	- Amplitude modulation
FM	- Frequency modulation
PCM	- Pulse code modulation
DPCM	- Differential pulse code modulation
ADPCM	- Adaptative differential pulse code modulation
LPC	- Linear predictive coding
CELP	- Code excited linear prediction
LD-CELP	- Low delay codebook excited linear prediction
AMP	- Amplitudes instantâneas
FI	- Freqüências instantâneas
ITU-T	- International Telecommunication Union
MOS	- Mean opinion score
LSF	- Line spectrum frequency pairs
BITS_AMP	- Quantidade de bits de codificação do sinal AMP;
BITS_FI	- Quantidade de bits de codificação do sinal FI;
F_AMP	- Largura do espectro do sinal de AMP;
F_FI	- Largura do espectro do sinal de FI.

SUMÁRIO

1 - INTRODUÇÃO

Atualmente, existe um grande interesse no processamento digital dos sinais voltado para aplicações em transmissão de voz em redes de computadores e de telefonia. Os trabalhos desenvolvidos nesta área têm em comum a redução da largura de banda dos sinais codificados. Assim, para um desempenho eficiente da transmissão de voz em redes, com boa qualidade e uma baixa taxa de transmissão de bits por segundo, pesquisas têm sido desenvolvidas para comprimir as taxas de transmissão dos codificadores sem que haja perda significativa da naturalidade e inteligibilidade da voz.

Existem diferentes formas de implementação de codificadores de voz [1], sendo que as comumente conhecidas podem ser divididas nas seguintes classes básicas: a dos codificadores de forma de onda, como, por exemplo, PCM (pulse code modulation) e ADPCM (adaptative differnetial pulse code modulation), e a dos codificadores paramétricos ou VOCODERs (codificação por vocalização), ou seja, os codificadores baseados em LPC (linear predictive coding). Adicionalmente, existem os codificadores híbridos que apresentam características de VOCODER e de codificação por forma de onda, como por exemplo, LD-CELP (low delay codebook excited linear prediction)

Desse modo, considerando as classes citadas, este trabalho apresenta uma variação simples e original de codificador de voz cujas características são conhecidas, mas pouco empregadas pelos codificadores padrões. O codificador proposto neste trabalho é baseado em dois conceitos: os formantes da voz e as componentes modulantes do sinal de voz.

Em linhas gerais, no método de codificação proposto, o espectro da voz limitado na faixa de 0 a 4 kHz é filtrado em quatro faixas estreitas cujo centro de cada faixa é determinada pelas freqüências ressonantes ou formantes [2], que são definidas pelos picos de

potência do espectro [3]. Em seguida, são extraídas as componentes modulantes AM-FM (amplitude modulation - frequency modulation) das faixas filtradas, as quais correspondem a quatro sinais de amplitudes instantâneas ou envelope (AMP) e a quatro sinais de freqüências instantâneas (FI). Portanto, os oito sinais obtidos são quantizados, codificados e transmitidos e, evidentemente, a partir das componentes modulantes, o sinal original é recuperado no lado da recepção.

Em resumo, este trabalho visa apresentar um método de codificação de voz para transmissão da voz humana em redes de telefonia e de computadores, onde as componentes modulantes AM e FM da voz são transmitidas no lugar da voz ou de seus parâmetros, ou seja, de modo diferente dos empregados nos codificadores de forma de onda e nos codificadores paramétricos.

Embora a idéia da decomposição do sinal de voz em componentes AM-FM seja bastante conhecida, as particularidades de cada proposta estão na forma da extração de suas componentes modulantes. Maragos, Kaiser e Quartieri [4] [5] [6] [7] desenvolveram um algoritmo baseado na separação de energia, na qual as detecções das componentes são realizadas através da aplicação de um operador de energia. Lu e Doerschuk [8] [9] [10] propuseram um modelo estatístico alternativo, onde um filtro não linear age como um banco de filtros passa- faixas, e as amplitudes e freqüências instantâneas são determinadas respectivamente a partir da freqüência central e da largura de banda de cada faixa filtrada. Kumaresan e Rao [12] [13] propuseram um algoritmo onde o sinal de voz é decomposto em um polinômio cujas raízes estão localizadas dentro e fora do circulo unitário do plano complexo, sendo que a partir das raízes obtém-se as componentes. Finalmente, neste trabalho é empregado um modelo já utilizado em extração de freqüências instantâneas aplicada em eletrocardiograma, onde se faz o uso explícito das características do sinal analítico, utilizando a transformada de Hilbert [14] [15].

No que se refere à aplicação das características modulantes AM-FM em codificadores de voz, Rao [13] identificou como fator desencorajador do uso deste método as grandes larguras de banda geradas pelas componentes modulantes. Entretanto, conforme será visto neste trabalho, os resultados obtidos aqui são satisfatórios, ou seja, as faixas das amplitudes e freqüências instantâneas são consideradas pequenas, comparadas com a largura do espectro de voz original.

Outros autores mencionaram o emprego deste modelo em codificadores. Fawe [16] abordou de forma mais otimista este assunto a partir da lei de Carson [17], onde o espectro do sinal modulante F é considerado menor que um quarto do espetro do sinal modulado B ($F < B/4$). Também, o autor sugeriu a obtenção das componentes modulantes diretamente do sinal de voz sem uma estratégia de redução da largura do sinal de voz. Entretanto, os resultados aqui apresentados divergem desta proposta por dois motivos: para aplicações em voz, a lei de Carson não apresenta um resultado confiável, pois a fase do sinal de voz altera-se de forma abrupta ao longo do tempo. Com isso, uma vez que a freqüência instantânea é a derivada da fase, o resultado não é o esperado quando da aplicação direta da fórmula $F < B/4$. Segundo, a obtenção das componentes direta do sinal de voz sem uma filtragem prévia, torna os espectros das componentes instantâneas muito largos. Desse modo foi utilizada no presente trabalho a técnica da redução da faixa de voz em quatro faixas estreitas em torno dos formantes, antes da extração das componentes modulantes.

Por último, Hanson [11] citou um modelo que se aproxima bastante do implementado neste trabalho, entretanto a proposta do codificador é colocada no seu artigo apenas como sugestão para futuros trabalhos, ou seja, não é apresentada uma implementação de fato. Além disso, na sua abordagem, ele sugeriu a obtenção das componentes AM-FM utilizando o operador de energia [4] [5] [6] [7]. Esta proposta diverge do presente trabalho, pois aqui será utilizada a associação das propriedades da técnica LPC [18] [19] para a

obtenção dos formantes e da transformada de Hilbert para a obtenção das componentes AM-FM [20] [21] [22].

Este trabalho de dissertação está divido da seguinte forma:

No capitulo 2 será feita uma abordagem dos motivos pelos quais o método está sendo proposto. Será feita também uma apresentação das fundamentações teóricas relativas às modulações AM-FM e aos formantes aplicados na codificação de voz.

No capítulo 3 será apresentado o método de fato, incluindo todas as suas etapas: estimação das freqüências dos formantes, filtragem, extração das componentes modulantes e recuperação do sinal de voz.

No capítulo 4 será feita uma ilustração de todas as etapas abordadas no capítulo anterior através de resultados práticos. Também, serão mostrados resultados de testes simplificados de inteligibilidade das amostras codificadas.

Finalmente, no capítulo 5 serão apresentadas explicações para os resultados obtidos, bem como serão feitas observações relacionadas a melhorias e propostas para novos trabalhos.

2 – FUNDAMENTAÇÃO TEÓRICA

2.1 Introdução

Atualmente existem diferentes tipos de codificadores de voz. Assim, o método proposto neste trabalho apresenta uma variação e originalidade em relação aos métodos padronizados de codificação.

Nas próximas seções serão apresentadas as classificações de codificadores de voz, bem como serão abordados os conceitos que sustentam o método proposto neste trabalho, ou seja, os formantes da voz e os sinais modulantes .

2.2 Padrões de Codificadores de Voz

Na telefonia digital convencional, o sinal de voz analógico gerado na fonte (telefone) é digitalizado para permitir o seu tráfego pelos sistemas de telecomunicações. A técnica de codificação padrão utilizada em telefonia para a digitalização da voz é denominada de modulação por codificação de pulso ou PCM. Nesta técnica, o sinal analógico da voz, limitado na faixa de 0 a 4 KHz, é amostrado a uma taxa de 8 KHz obedecendo ao teorema de Nyquist. Em seguida cada pulso é quantizado através do ajuste dos níveis de cada pulso para valores pré-estabelecidos. Por último, os pulsos quantizados são codificados através de códigos de 8 bits. O sinal resultante produz uma taxa de 64 kbps. A técnica PCM é muito antiga, sendo considerada a referência dos codificadores de forma de onda. Outros métodos de codificadores de forma de onda são o DPCM e o ADPCM [1]. As principais características desses codificadores são taxas entre 64 kbps e 32 kbps, podendo chegar 16 kbps, e boa

qualidade da voz codificada. Os codificadores de forma de onda são empregados principalmente redes de telefonia.

Em outro extremo, existem os codificadores paramétricos que, ao invés de tratarem a voz a partir do seu sinal produzido, extraem os parâmetros que a produzem, e os transmitem no lugar do sinal da voz propriamente dita. Na recepção a voz é modelada, e assim, recuperada. O codificador paramétrico padrão desta classe é o codificador por predição linear ou LPC [18]. As principais características desses codificadores são as baixas taxas alcançadas, em torno de 2,4 kbps, e a baixa qualidade da voz produzida. Por isso, os codificadores LPC não são utilizados em sistemas comerciais de comunicação. Atualmente, existem evoluções de codificadores que empregam modelos parecidos com o LPC, como por exemplo, o codificador CELP (Code excited linear prediction) [23].

Embora os exemplos de codificadores citados sejam antigos, tendo uma certa consolidação em aplicações de telecomunicação, técnicas mais recentes de codificação de voz têm sido propostas a fim de que sejam atendidos os requisitos para a transmissão de voz em tecnologias de redes emergentes, tais como redes de computadores de alta velocidade, Internet, e principalmente, redes de telefonia móveis [1] [24] [25]. Estes requisitos básicos são: baixas taxas de transmissão da voz, normalmente abaixo de 8 kbps, baixo atraso e bom nível de inteligibilidade e naturalidade da voz.

Na linha dos codificadores mais modernos estão aqueles que associam as características das baixas taxas dos paramétricos e a boa qualidade dos codificadores de forma de onda. Estes codificadores são denominados de híbridos [26].

A tabela 2.1 apresenta um resumo dos principais codificadores de voz. São apresentadas também as referências das recomendações de padronização do ITU-T (International Telecommunication Union) [1] [27].

Tabela 2.1 – Padrões de codificadores de voz

ITU-T	SIGLA	*DENOMINAÇÃO*	TAXA	CLASSIFICAÇÃO
G.711	PCM	Pulse code modulation	64 kbps	Forma de Onda
G.721	ADPCM	Adaptative differretial pulse code modulation	32 kbps	Forma de Onda
G.722	SB-ADPCM	Sub band adaptative differretial pulse code modulation	64 kbps	Forma de Onda
G.723.1	MP-MLQ	Algebraic code excited linear prediction	5,3 kbps	Paramétrico
G.723.1	ACELP	Algebraic code excited linear prediction	6,3 kbps	Paramétrico
G.728	LD-CELP	Low delay code excited linear prediction	16 kbps	Híbrido
G.729	CS-ACELP	Conjugate structure algebraic code excited linear prediction	8 kbps	Híbrido

2.3 Codificação a Partir dos Sinais Modulantes da Voz

Todas as técnicas de codificação empregadas e padronizadas pelo ITU-T, conforme tabela 2.1 estão baseadas nos princípios básicos da codificação da forma de onda do sinal da voz, da codificação dos parâmetros da voz, ou da associação de ambos. Em nenhum dos codificadores utilizam-se técnicas similares à proposta neste trabalho, ou seja, codificação a partir da extração das componentes modulantes da voz.

A técnica de extração das componentes modulantes da voz baseia-se no princípio de que o sinal de voz é constituído de respectivas amplitudes e freqüências instantâneas. Desse modo, a voz pode ser considerada um sinal modulado, e a extração ou decomposição deste em amplitudes e freqüências instantâneas, é considerada uma demodulação.

A obtenção das amplitudes (AMPs) e freqüências (FIs) instantâneas remete à utilização das propriedades AM e FM do espectro da voz, onde o sinal de voz pode perfeitamente ser decomposto em suas componentes modulantes AM e FM [20] [21] [22]. A demodulação funciona como um redutor das taxas de codificação do sinal de voz, uma vez

que as larguras das faixas ocupadas pelos sinais modulantes podem ser consideradas bem menores que a faixa da voz original.

A figura 2.1 ilustra o comportamento das componentes modulantes AMPs e FIs do método proposto. Conforme já citado, para a extração das componentes modulantes, inicialmente a faixa de voz de largura de 4 kHz é reduzida em faixas estreitas em torno dos formantes da voz. A quantidade de faixas escolhida foi quatro, uma vez que dentro da faixa de voz, ocorrem normalmente apenas quatro formantes. Portanto, as quatro faixas filtradas são moduladas em amplitude pelos sinais AMPs e em freqüência pelos FIs, e assim, ao se decompor o sinal original da voz de cada faixa, obtém-se sinais instantâneos que correspondem às variações de amplitudes e freqüências no tempo. Quanto menor a taxa de variação dos sinais AMPs e FIs, menor será a taxa de codificação obtida para a transmissão.

Utilizando também a ilustração da figura 2.1, pode-se entender melhor o comportamento das componentes AM e FM. No domínio do tempo a componente modulante AM, de cada faixa do sinal de voz filtrada, corresponde à envoltória do sinal, produzindo assim, sinais AMPs. Este mesmo sinal de amplitudes instatâneas corresponde às metades das larguras das faixas filtradas no domínio da freqüência.

De outro modo, a componente modulante FM é definida no domínio do tempo como as variações da fase do sinal, produzindo assim, as freqüências instatâneas. No domínio da freqüência este mesmo sinal de FIs corresponde aos centros das faixas filtradas.

Vale identificar que o codificador proposto pode ser classificado como híbrido, uma vez que os formantes estão relacionados aos parâmetros da voz, enquanto os sinais modulantes extraídos são codificados de forma similar aos codificadores de forma de onda.

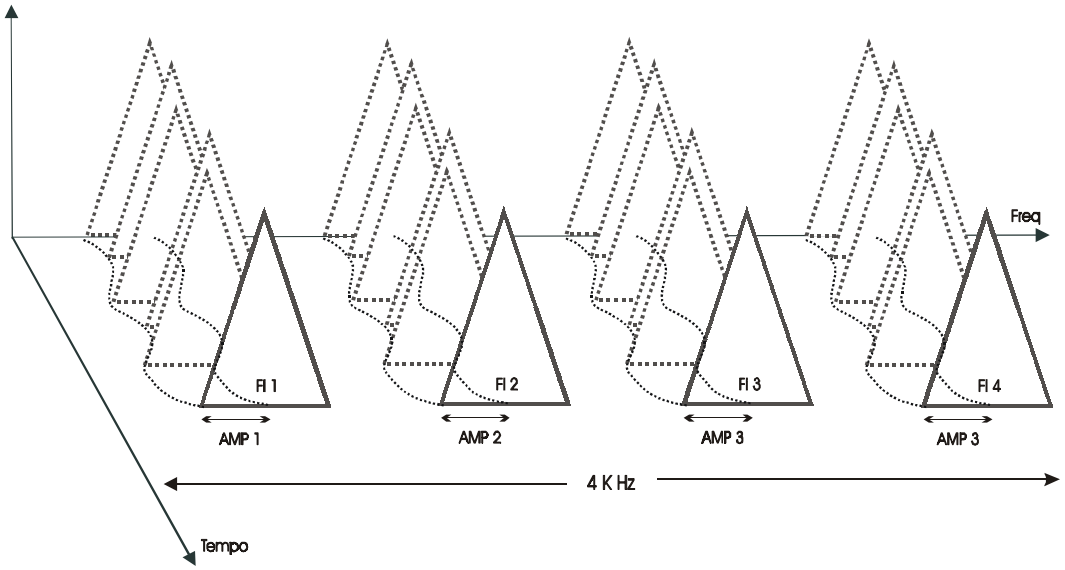


Figura 2.1 – Modulação AM e FM

2.4 Formantes da Voz

Embora a técnica apresentada neste trabalho considere o sinal de voz composto a partir de suas componentes modulantes, ou seja, amplitudes e frequências instantâneas, de certa forma na prática não existem uma modulação AM e FM, conforme o método tradicional aplicado em sistemas de comunicação. Na realidade, a voz é resultante da modulação, pelo trato vocal, de excitações provenientes da glote e das cordas vocais [2].

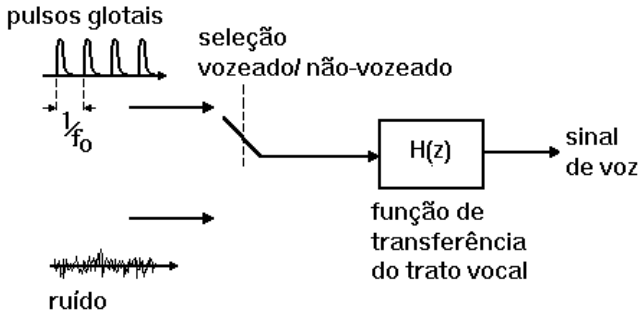


Figura 2.2 – Modelo do Trato Vocal

das consoantes, tais como, “s” e “f”. Estes sons são denominados de surdos ou não vozeados.

Basicamente, o ar produzido nos pulmões passa pela glote. Em seguida, atravessa a laringe, faringe, e as cavidades bucal e nasal, os quais representam o trato vocal. Quando o ar atravessa a glote livremente sem obstruções, a excitação corresponde a ruídos de largo espectro e os sinais modulados pelo trato vocal correspondem aos sons presentes na maioria

De outro modo, quando o ar atravessa a glote sofrendo obstruções provenientes das vibrações das cordas vocais, a excitação corresponde a pulsos globais, cuja freqüência fundamental é denominada de *pitch*. Os sinais modulados pelo trato vocal correspondem aos sons presentes nas vogais. Estes sons são denominados de sonoros ou vozeados. A figura 2.2 [2] apresenta o modelo do trato vocal.

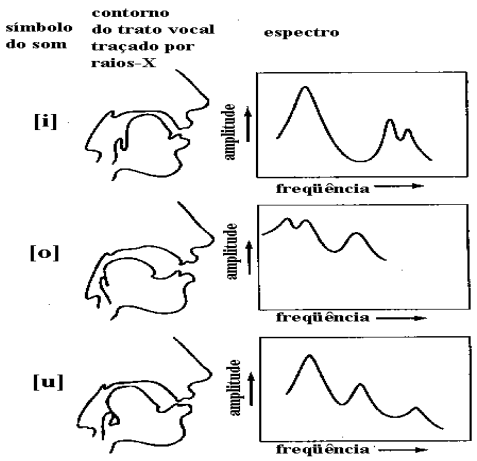


Figura 2.3 – Formantes da Voz

Para os sons vozeados, o trato vocal funciona como cavidades ressonantes produzindo os harmônicos. As freqüências de ressonância do trato vocal são denominadas de formantes [2] [3] [28]. Dependendo do comportamento das cavidades, o espetro de freqüência do sinal da voz varia continuamente. Por exemplo: ao emitir o som vozeado de uma vogal “a”

os formantes gerados são particulares e diferentes de outras vogais. Cada som vozeado possui um espectro característico. Os formantes são denominados como 1º, 2º, 3º, 4º,5º ..., e assim por diante.

Embora haja um número infinito de formantes ao longo do espectro da voz , o método apresentado neste trabalho tem um interesse particular nos primeiros quatro formantes, uma vez que estes estão dentro da faixa de voz limitada entre 0 a 4 kHz.. A partir do quinto formante, as ocorrências dos mesmos, estão além da faixa de 4 kHz.

A figura 2.3 [2] apresenta três exemplos de sons emitidos pelas vogais "i", "o" e "u". Na ilustração, os formantes são as freqüências onde ocorrem os pic os destes espectros, sendo também, os centróides de maior de energia. Logo, pode-se considerar que a maior concentração de energia do sinal do sinal de voz está em torno dos formantes. Isto explica o fato de o codificador proposto reduzir o espetro de largura de 4 kHz em quatro faixas estreitas em torno d os quatro formantes, sem que haja perda significativa da inteligibilidade da voz.

3 - O MÉTODO

3.1 Introdução

O método de codificação proposto neste trabalho pode ser dividido em três partes básicas: determinação das freqüências correspondentes aos quatro formantes do espectro, obtenção de quatro faixas estreitas em torno dos quatro primeiros formantes, e por último, determinação das componentes modulantes das faixas filtradas. Além das três partes básicas, existe uma etapa complementar, que corresponde à recuperação da voz a partir das componentes modulantes.

Nas seções seguintes serão apresentadas de forma mais detalhada as partes que compõem o método bem como a base teórica empregada para a obtenção dos parâmetros (formantes) da voz, as filtragens e as extrações das componentes modulantes. Por último, será apresentada a forma utilizada para a recuperação da voz a partir das componentes modulantes.

3.2 Detalhamento do Método

A figura 3.1 apresenta em diagrama de blocos o método proposto. Neste diagrama são apresentadas as três partes básicas do codificador. A seguir será explicado mais detalhadamente o método:

1º Inicialmente, a partir do sinal original amostrado a 8 kHz são estimadas as freqüências correspondentes aos quatro primeiros formantes do espectro de uma janela de 240 amostras (30 ms). Esta estimação é feita utilizando o método de predição linear (LPC).

2º Em seguida, as freqüências dos formantes são utilizadas como referência para as filtragens passa faixa utilizando uma função wavelet modificada de Gabor [14] [15]. Assim, são obtidas quatro faixas de espectro em torno dos quatro formantes;

3º A partir das quatro faixas filtradas são determinadas as amplitudes e freqüências instantâneas correspondentes de cada faixa utilizando-se as propriedades da transformada de Hilbert;

4º Os sinais correspondentes às amplitudes e freqüências instantâneas de cada faixa correspondem à informação que será codificada. Portanto, serão transmitidos os quatro sinais de amplitude e quatro de freqüência instantâneas codificados;

5º Em seguida, retorna-se a 1º etapa para a codificação da próxima janela, obedecendo a um deslocamento de 180 amostras, conforme ilustrada na figura 3.2.

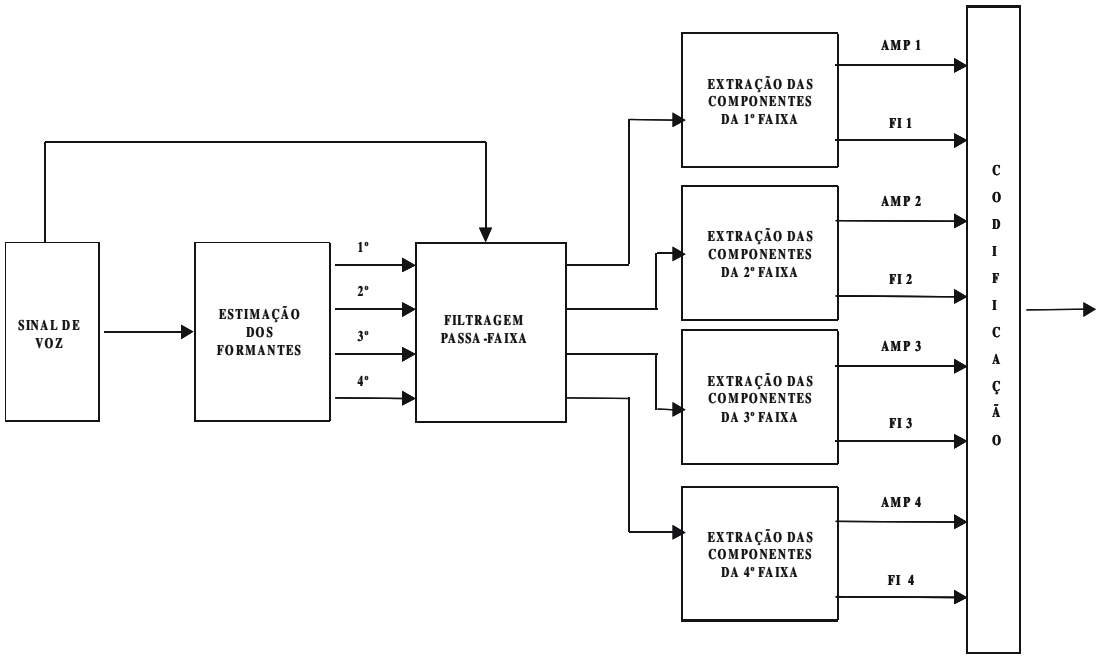


Figura 3.1 – Diagrama de blocos do codificador

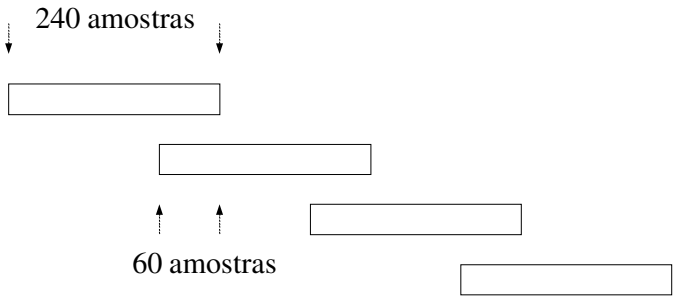


Figura 3.2 – Janela deslizante para obtenção dos formantes

3.3 Estimação dos Formantes

Para a estimação das freqüências dos formantes utiliza-se LPC com o emprego do método da auto-correlação ou abordagem auto-regressiva de Yule-Walker [3] [28]. Estas freqüências são os picos do espectro de potência que, do ponto de vista matemático, são as raízes do polinômio do denominador da função de transferência, isto é, os pólos [3].

O processo de estimação das freqüências pode ser resumido através do diagrama de blocos a figura 3.3.

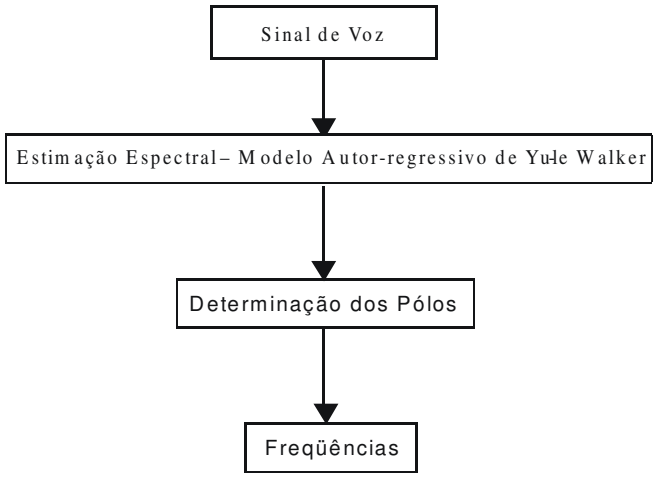


Figura 3.3 – Estimação das freqüências dos formantes

Os coeficientes da função de auto-correlação são determinados considerando o sinal estacionário num intervalo de tempo. No codificador em questão foi escolhida uma janela de 30 ms.

Embora a voz produza um sinal não estacionário, na prática o trato vocal varia lentamente, por isso o sinal de voz pode ser suposto estacionário entre 10 a 40 ms [3]. Isto explica a escolha de uma janela de 30 ms.

Após a obtenção dos coeficientes da função de autocorrelação, são determinadas as raízes do polinômio da função de transferência de *(1)* que correspondem aos pólos da função. Deve ser observado que c_k em $H(z)$ são os coeficientes de autocorrelação .

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} c_k z^{-1}} \qquad (1)$$

Após a obtenção das raízes do polinômio, basta encontrar as freqüências angulares correspondentes aos pólos da função. Entretanto, para o problema em questão, as freqüências, cujos termos reais são negativos, são descartadas, e as freqüências angulares restantes correspondem aos valores desejados. Nota-se que o número de freqüências encontradas é correspondente à quantidade de coeficientes, mas neste método são utilizadas apenas as quatro primeiras freqüências.

3.4 Filtragem Passa Faixa

As filtragens em torno das freqüências correspondentes aos picos do espectro de potência da janela são realizadas utilizando-se a wavelet básica de Gabor [14] [15] definida por *(2)*.

$$\psi(t) = \frac{1}{2\pi}\frac{d}{dt}\left[\exp\left(-\pi\left\{\frac{\overline{\delta(t)}t}{2}\right\}^2\right)\cos\left(2\pi t\int_{\Omega}\delta(\tau)d\tau\right)\right], \qquad \overline{\delta(t)} = \frac{1}{\Omega}\sum_{\Omega}\delta(t) \qquad (2)$$

onde Ω é um pequeno intervalo de tempo. O sinal filtrado nesse intervalo é dado por

$$x_{\Omega}(t) = \int_{\Omega} \psi(\tau)\, y_i(t-\tau) d\tau \qquad (3)$$

que corresponde a uma convolução entre o $\psi(t)$ e o sinal de voz [29].

Vale frisar que para a filtragem passa-faixa requerida, qualquer método de filtragem adaptativa pode ser empregado, entretanto a escolha da wavelet modificada de Gabor deve-se a sua boa localização tempo-freqüência, comprovada em recentes trabalhos de Barros e Ohnishi [14];

3.5 Determinação das Freqüências e Amplitudes Instantâneas

Para a obtenção dos sinais AMPs e FIs, as freqüências positivas e negativas do sinal real filtrado devem ser isoladas, a fim de tornar o sinal filtrado analítico. Para se conseguir o sinal analítico, basta obter a componente imaginária do sinal original através da transformada de Hilbert [20] [21] [22] dada por

$$\bar{x}_i(t) = \frac{1}{\pi} \int \frac{x(\tau)}{t-\tau} d\tau \qquad (4)$$

onde $x(t)$ é o sinal real filtrado e $\bar{x}_i(t)$ o termo imaginário obtido.

O sinal analítico a partir de $x(t)$ é definido por

$$s_i(t) = x_i(t) + j\bar{x}_i(t) \qquad (5)$$

A partir de $s(t)$ obtém-se AMP conforme *(6)*

$$a_i(t) = \left| H[s_i(t)] \right| \qquad (6)$$

Também, a FI é obtida a partir de $s(t)$ conforme *(7)*

$$\omega_i(t) = \frac{d\phi_i(t)}{dt}, \quad \phi_i(t) = \arctan\left(\frac{\left| H[s_i(t)] \right|}{s_i(t)} \right) \qquad (7)$$

A utilização da transformada de Hilbert para a determinação das componentes instatâneas do sinal está baseada no fato de que em um sinal analítico existem duas componentes únicas: a fase $\phi_i(t)$ e o seu módulo $\left|H[s_i(t)]\right|$. Estas componentes são obtidas a partir dos respectivos termos: um real e outro imaginário. Por outro lado, diferente do sinal analítico, a fase $\phi_i(t)$ de um sinal real é indeterminada haja vista que não existe o termo imaginário.

3.6 Recuperação do sinal de voz

Uma vez codificadas e transmitidas as componentes, será necessária a recuperação na recepção das quatro faixas reais para compor o sinal de voz. A figura 3.4 apresenta em diagrama de blocos o decodificador para a recuperação do sinal de voz.

O sinal recuperado é o termo real do sinal analítico [21] [22], ou seja:

$$y_i(t) = a_i(t)\cos 2\pi \int \omega(t)dt \qquad (8)$$

O sinal de voz recuperado corresponde à soma dos quatro sinais determinados a partir de *(8).* Deve ser ressaltado também que o sinal recuperado perde parte da informação, pois o espectro de voz resultante é formado apenas pelas quatro faixas.

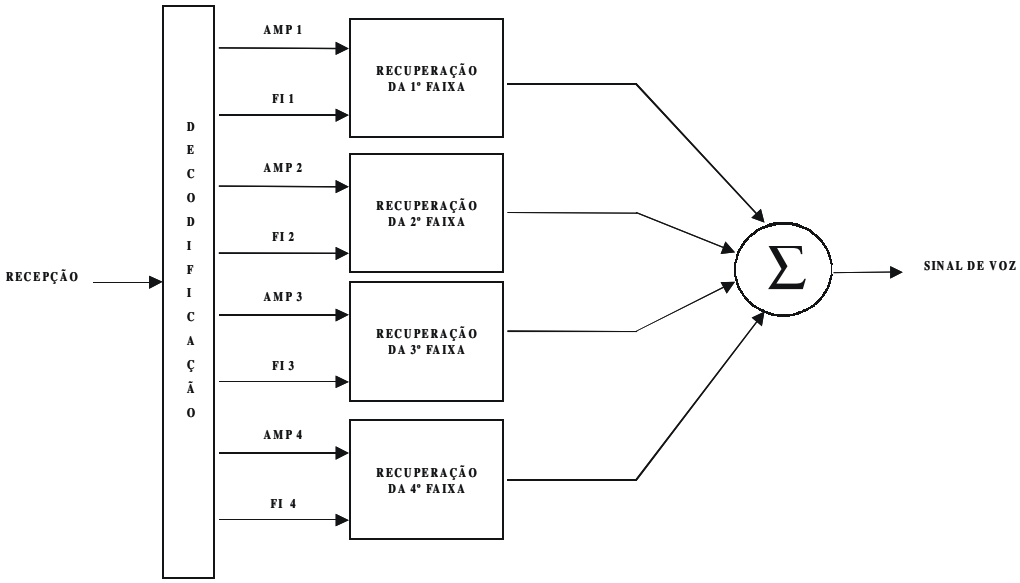


Figura 3.4 – Diagrama de blocos do decodificador

4 - RESULTADOS

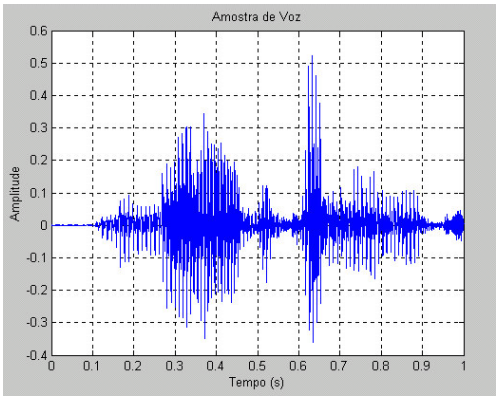
4.1 Introdução

Partindo-se do embasamento teórico do capítulo anterior, este capítulo apresenta os resultados práticos do método proposto. Nas seções seguintes serão apresentados os resultados de todas as etapas do método, tanto no domínio do tempo quanto no domínio da freqüência. Após as ilustrações, serão abordadas as questões relacionadas com a estimativa da taxa do codificador e a avaliação do codificador quanta à qualidade da voz.

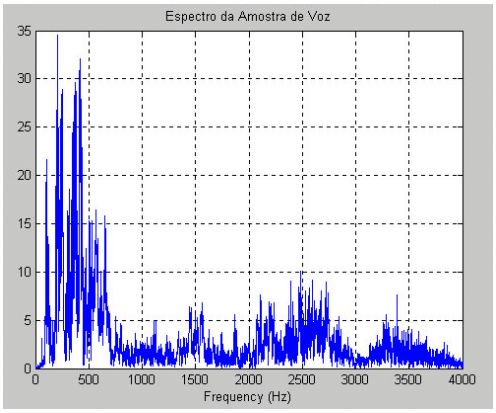
4.2 Ilustração do Método

Para ilustrar o método, as figuras 4.1, 4.2, 4.3 e 4.4 apresentam os resultados práticos obtidos de uma das amostras analisadas. Foi escolhida a amostra 1 do apêndice A para a ilustração. A figura 4.1.a representa as 8000 primeiras amostras ou os 1000 ms iniciais, sendo que a figura 4.1.b o seu espectro de freqüência correspondente.

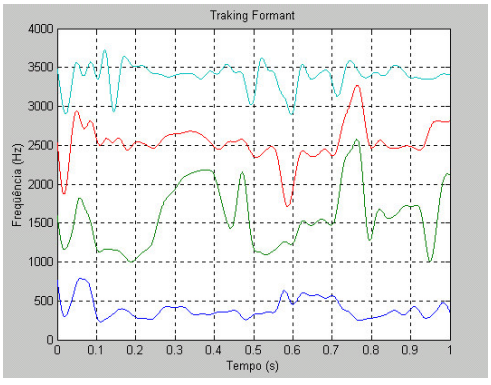
A partir da amostra de voz, as freqüências dos formantes foram calculadas dentro de janelas de 30 ms, conforme *(1).* A figura 4.1.c mostra as freqüências correspondentes encontradas no tempo. Em seguida, a partir das freqüências dos formantes estimadas, foram construídas as wavelets aplicando-se *(2)* e filtradas as faixas, aplicando-se *(3)* dentro de intervalos de 30 ms. A figura 4.2 mostra as wavelets construídas e a figura 4.3 mostra os espectros produzidos pelas quatro faixas filtradas.



a) Domínio do Tempo

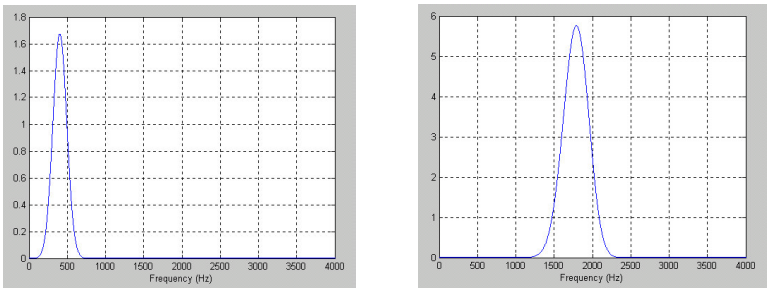


b) Domínio da Freqüência



c) Estimação dos Formantes

Figura 4.1 – Primeiras 8000 amostras do sinal



a) Wavelet do 1º formante b) Wavelet do 2º formante

c) Wavelet do 3º formante d) Wavelet do 4º formante

Figura 4.2 – Construção das wavelets

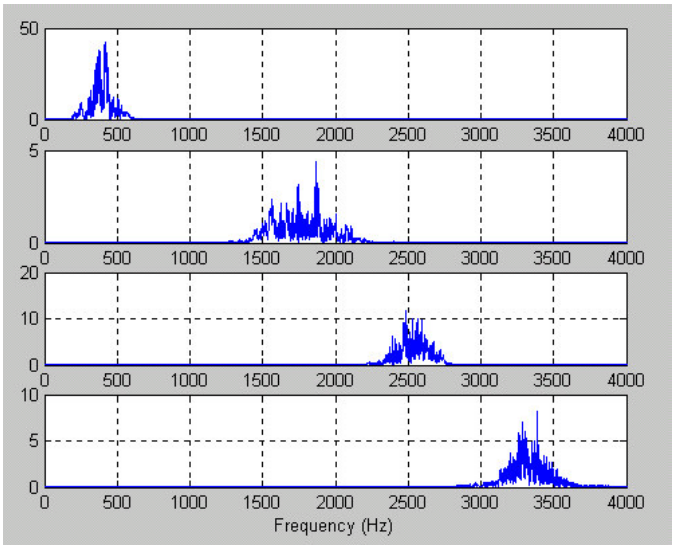
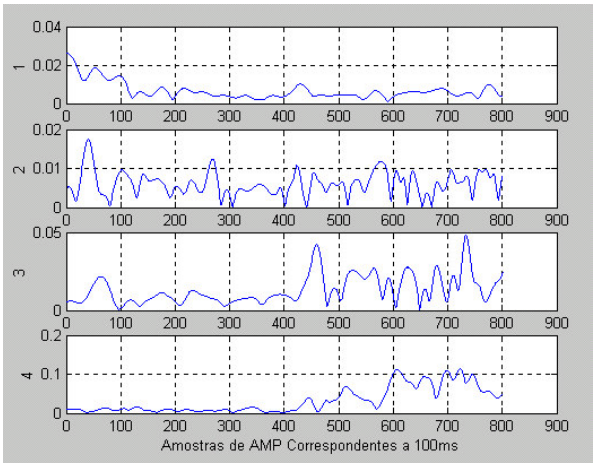


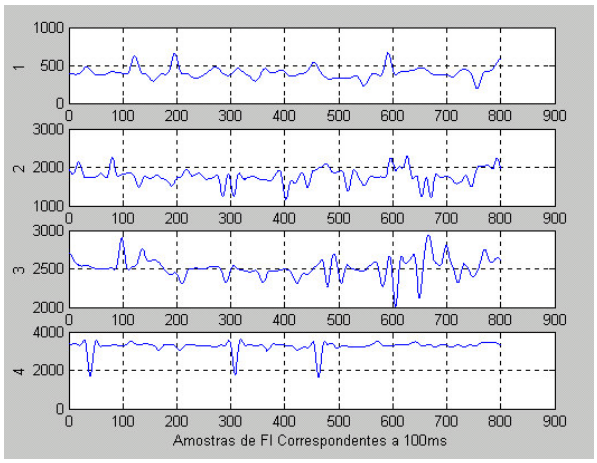
Figura 4.3 – Faixas Filtradas

A partir das quatro faixas filtradas foram determinadas as amplitudes e freqüências instantâneas correspondentes de cada faixa, utilizando a transformada de Hilbert conforme

(4), (5), (6) e *(7).* As figuras 4.4.a e 4.4.b mostram 800 amostras ou 100 ms dos sinais AMPs e FIs encontrados, entre 900 ms e 1000 ms.



a) Amplitudes instantâneas



b) Freqüências instantâneas

Figura 4.4 – Domínio do tempo dos sinais AMP e FI das quatro faixas

Após a obtenção dos oito sinais partiu-se para a última etapa, ou seja, a estimativa da codificação através da quantização dos respectivos sinais.

4.3 Estimativa da Taxa do Codificador

Embora a proposta deste trabalho seja apresentar um método para futuras aplicações em codificação de voz sem entrar no mérito da taxa de transmissão alcançado para o codificador de voz, vale citar a estimativa da taxa obtida. Para isto, foi necessário inicialmente estimar a largura do espectro dos sinais AMPs e FIs. Neste caso, observou-se através de testes que para a obtenção de uma voz inteligível, os sinais de AMPs teriam que ter no mínimo 100 Hz de largura de banda e FI com 400 Hz. Assim, para fixar os sinais nestas larguras, foram utilizados filtros passa baixa com freqüências de corte em 100 Hz e 400 Hz..

Outro ponto importante na estimativa da taxa de transmissão é a quantidade de bits de codificação. Neste caso, observou-se também na prática que para a obtenção de uma voz inteligível, os sinais de AMPs teriam que ter no mínimo 16 níveis de quantização (4 bits) e FI 32 níveis de quantização (5 bits). A taxa final estimada foi de 19200 bps aplicando-se:

Taxa de Transmissão do Codificado = [(FORMANTS x BITS_AMP x 2 x F_AMP)+ (FORMANTS x BITS_FI x 2 x F_FI)] bps (9)

onde,

FORMANTS: número de formantes;

BITS_AMP: quantidade de bits de codificação do sinal AMP;

BITS_FI: quantidade de bits de codificação do sinal FI;

F_AMP: largura do espectro do sinal de AMP;

F_FI: largura do espectro do sinal de FI.

Taxa de Transmissão do Codificado = [(4 x 4 x 2 x 100)+ (4 x 5 x 2 x 400)] = 19200 bps

Para as quantizações dos sinais AMPs e FIs aplicam-se *(10)* e *(11)* no sinal. Observando-se os resultados de (10) e (11), verifica-se que sinais AMPs e FIs são os mesmos, porém quantizados.

$$a_i(t) = round\left(\frac{2^{NQ} a_i(t)}{\max(abs(a_i(t)))}\right) x \left(\frac{\max(abs(a_i(t)))}{2^{NQ}}\right) \qquad (10)$$

onde, *NQ* corresponde aos níveis de quantização de FI. Neste caso *NQ* é igual a 4.

$$\omega_i(t) = round\left(\frac{2^{NQ} \omega_i(t)}{\max(abs(\omega_i(t)))}\right) x \left(\frac{\max(abs(\omega_i(t)))}{2^{NQ}}\right) \qquad (11)$$

onde, *NQ* corresponde aos níveis de quantização de AMP. Neste caso *NQ* é igual a 5.

4.4 Avaliação da Qualidade *Versus* a Taxa do Codificador

Existem diversos tipos de testes para medir o desempenho do codificador com relação à inteligibilidade e naturalidade da voz codificada. Estes testes podem ser objetivos ou subjetivos [30] [31].

Considerando que este trabalho visa apresentar apenas um método, sem a preocupação na implementação do codificador, a validação da codificação da voz não obedeceu aos rigores exigidos pelos padrões existentes. Sendo assim, foi escolhido um teste conhecido como método de opinião subjetiva, mais conhecido como teste de MOS (mean opinion score) [31]. Neste caso, um conjunto de sentenças, mais precisamente 24, é submetido aos ouvintes, e estes dão notas que variam conforme escala da tabela 4.1.

Mais uma vez vale ressaltar que a forma de avaliação da voz codificada teve o simples objetivo de provar que o método é factível e a voz é inteligível, e desse modo, coube ao autor do trabalho implantar uma variante do método MOS, de forma totalmente simplificada.

Para a avaliação do método, foram utilizadas cinco amostras de frases as quais estão representadas no apêndice A. Para os testes, as frases foram originalmente geradas com taxas de amostragem de 48 kHz e codificação a 16 bits a fim de se melhorar a qualidade das amostras. Uma vez que o teste foi feito considerando apenas os requisitos necessários para a inteligibilidade, houve a preocupação de que os ruídos não interferissem nos resultados, portanto, as amostras originais foram gravadas obedecendo a relação sinal/ruído de 40 dB. Em seguida, as mesmas foram sub amostradas a 8 kHz uma vez que o objetivo da codificação proposta neste trabalho é a telefonia.

Nas amostras originais, foram aplicadas todas as etapas do método para a obtenção das componentes modulantes, conforme descrito no capitulo 3. Em seguida, para testar o grau de inteligibilidade a partir dos sinais modulantes codificados a 19200 bps, aplicou-se *(8)* para a recuperação do sinal de voz.

Tabela 4.1 – Teste de MOS

Qualidade de Voz	**Pontos**
Excelente	5
Boa	4
Razoável	3
Pobre	2
Ruim	1

As frases recuperadas foram entregues a dez ouvintes, os quais deram uma pontuação para cada frase ouvida, obedecendo ao critério do MOS, conforme tabela 4.1. As 50 pontuações variaram entre 2 e 3,5, sendo que o valor médio final ficou próximo de 3. A figura 4.5 mostra o resultado final do teste de MOS do codificador, bem como um gráfico

comparativo do resultado deste em relação aos codificadores tradicionais, levando-se em consideração, além da pontuação do MOS, a taxa de transmissão em Kbps conseguida.

Também, a fim de se observar o desempenho do codificador, o teste de MOS foi utilizando para diferentes taxas, tanto acima como abaixo do ponto estimado anteriormente. Desse modo, para conseguir diferentes estimativas de bps do codificador, as larguras dos espectros dos sinais AMPs e FIs foram modificadas, conforme tabela 4.2, sendo que os níveis de quantizações foram mantidos constantes. Logo, foram obtidas 10 combinações e aplicando-se as 10 combinações em *(9),* foram obtidas 10 taxas que variaram de 7200 bps a 38400.

Aplicando-se o mesmo procedimento do teste MOS foram obtidas pontuações que variaram entre 3 e 1,5 da escala MOS. A figura 4.5 [32] mostra os resultados das dez pontuações obtidas utilizando as taxa da tabela 4.2. As dez pontuações estão plotadas através da linha mais grossa do gráfico.

De acordo com os resultados, foi observado que quando a taxa do codificador aumenta a partir do ponto ótimo (19200 bps), o ganho em relação ao MOS é pequeno. Por outro lado, quando a taxa é reduzida a partir do mesmo ponto o MOS cai rapidamente.

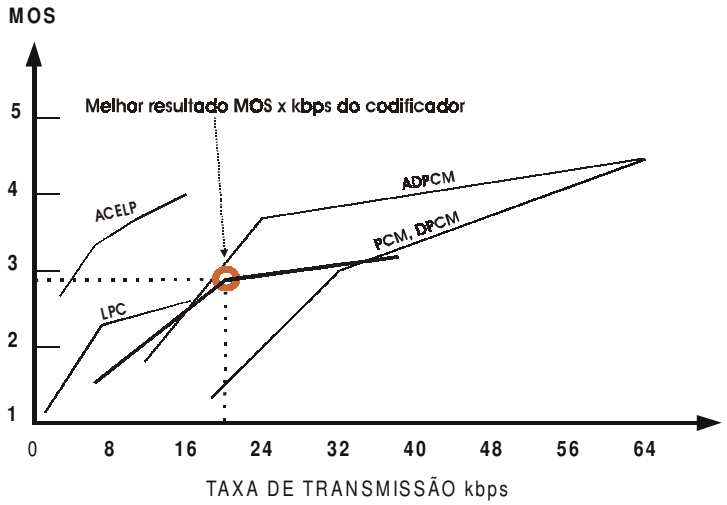


Figura 4.5 – MOS x taxa de transmissão em kbps

Tabela. 4.2 – Taxa de transmissão em Kbps

F_AMP (Hz)	F_FI (Hz)	BITS_AMP	BITS_FI	Tx (bps)
100	100	4	5	7200
200	100	4	5	10400
100	200	4	5	11200
200	200	4	5	14400
100	**400**	**4**	**5**	**19200**
200	400	4	5	22400
100	600	4	5	27200
200	600	4	5	30400
100	800	4	5	35200
200	800	4	5	38400

As explicações para o comportamento do MOS para diferentes taxas, as quais estão mostradas na figura 4.5 e na tabela 4.2 são as seguintes:

- Para taxas acima de 19200bps, o aumento da largura de banda dos sinais FIs não contribuem para o aumento do MOS. A perda da inteligibilidade do codificador neste intervalo (19200 a 38400 bps) está ligada diretamente a dois fatores intrínsecos e preponderantes do codificador: a técnica de estimação de formante LPC pouco eficiente e a perda de informação da voz após a filtragem em torno dos formantes;
- Por outro lado, para taxas abaixo de 19200bps, a diminuição das bandas dos sinais FIs contribuem significativamente para a redução da inteligibilidade, somando-se aos fatores intrínsecos já citados (filtragens em torno dos formantes e estimação dos formantes LPC). Por isso, o MOS caia rapidamente no intervalo de 19200 a 7200 bps;

- Para as diferentes taxas, a variação da largura de banda dos sinais AMPs não contribui de forma perceptível para a variação do MOS.

5 - DISCUSSÃO

5.1 Introdução

A proposta deste trabalho é apresentar um método novo para codificação e transmissão de voz diferente dos métodos tradicionais já conhecidos. Desse modo, embora o método aqui apresentado não priorize a questão da taxa de transmissão, algumas considerações devem ser feitas a fim de que futuros trabalhos possam se aprofundar nas questões referentes à eficiência da inteligibilidade do codificador em baixas taxas de transmissão.

Nas próximas seções deste capítulo serão apresentadas considerações sobre os resultados obtidos no capítulo anterior, bem como algumas observações sobre os parâmetros e as suas influências na performance do codificador. Também, serão apresentadas propostas de melhorias que possam servir de base para futuros trabalhos.

5.2 Avaliação da Taxa de Transmissão do Codificador

Conforme resultados obtidos, verificou-se que as variações dos sinais instantâneos AMPs e FIs foram altas, conforme mostrados na figura 4.4. Desse modo, para se implementar um codificador eficiente que permita a transmissão em baixas taxas, por exemplo abaixo de 8Kbps, sem o comprometimento da qualidade, ficou clara a necessidade de se reduzir ainda mais as variações dos sinais AMPs e FIs. Conforme análises, observou-se que as variações dos sinais instantâneos de freqüência estão relacionadas às inversões ou mudanças bruscas das fases dos sinais de voz. As variações de fase, por sua vez, estão relacionadas às variações da amplitude do sinal de voz. Desse modo, caberá a futuros trabalhos o entendimento da relação

existente entre as amplitudes e fases instantâneas do sinal de voz, e os reflexos desses nas freqüências instantâneas.

Esta questão merece um certo aprofundamento, pois uma vez conhecido previamente o comportamento da amplitude e da fase sinal de voz gerado pelo trato vocal, as variações dessas variáveis poderão ser mascaradas ou parametrizadas na transmissão, e recuperadas na recepção, diminuindo possivelmente a largura de banda das informações de voz do codificador.

5.2.1 Influência da Fase do Sinal de Voz na Freqüência Instantânea

As figuras 5.1, 5.2 e 5.3 ilustram as observações citadas do comportamento da fase e a sua influência no resultado das amplitudes e freqüências instantâneas. Neste exemplo foram extraídas 100 amostras de um sinal de voz que correspondem a 125 ms. Conforme mostrado nas figuras, a fase do sinal de voz é sempre crescente e quase linear, entretanto em dois pontos ocorrem alterações bruscas no seu percurso, os quais produzem variações correspondentes nas amplitudes e freqüências instantâneas. Este comportamento é explicado através da equação $\omega_i(t) = \frac{d\phi_i(t)}{dt}$ da seção 3.5, ou seja, a freqüência instantânea é definida como a derivada da fase.

Pelo exposto acima, considerando que a variação da fase apresenta um comportamento mais suave no tempo em relação a freqüência, sugere-se portanto a possibilidade da transmissão da fase instantânea ao invés da freqüência. Isto é plenamente factível uma vez que na recuperação do sinal da voz as duas variáveis necessárias são a fase e a amplitude instantâneas, conforme equação $y_i(t) = a_i(t)\cos 2\pi \int \omega(t)dt$ ou $y_i(t) = a_i(t)\cos 2\pi\phi_i(t)$ da seção 3.5.

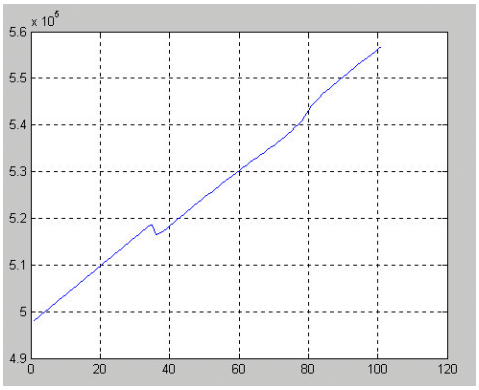


Figura 5.1 – Amostra da fase do sinal de voz

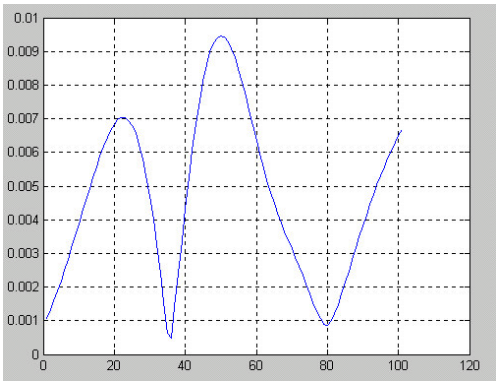


Figura 5.2 – Amostra da amplitude correspondente

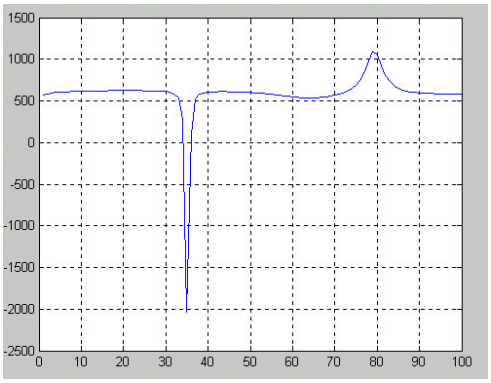


Figura 5.3 – Amostra da freqüência correspondente

Considerando a largura de banda dos sinais FIs, esta opção da transmissão da fase no lugar da freqüência torna o codificador mais atraente do ponto de vista da redução da taxa do codificador, se caso for aplicado um algoritmo que tire proveito do comportamento suave e linear da fase do sinal.

5.3 Avaliação do Traking Formant do Codificador

A técnica de traking formant [13] pode ser definida como a forma de estimação das freqüências dos formantes ao logo do tempo. Quanto mais eficiente esta técnica, mais próximas as freqüências obtidas pela estimação dos formantes ficarão das freqüências reais, Também, quanto mais coincidentes forem estas freqüências, mais eficientes será o codificador, pois conforme já explicada, a filtragem ocorre em torno dos formantes. Se a estimação não for precisa, a filtragem não ocorrerá nos pontos do espectro de maior potência do sinal.

A figura 5.4 ilustra o espectograma referente aos três primeiros segundos da amostra 1 do anexo A. A figura 5.5 mostra as estimações dos formantes correspondentes. Portanto, o traking formant pode ser avaliado através da comparação das duas figuras, onde a localização real dos formantes é feita utilizando o espectograma da figura 5.4 e as estimações dos mesmos através do método LPC são plotadas na figura 5.5.

Sobre os resultados obtidos faz-se necessário as seguintes considerações:

- Deve-se ter um cuidado especial com as filtragens em torno dos formantes vizinhos, pois caso as freqüências estimadas estejam muito próximas, pode ocorrer sobreposição de espectros na recuperação do sinal. Logo, antes da extração das componentes modulantes, as quatro faixas devem estar suficientemente espaçadas, a fim de que não ocorram distorções na voz recuperada;
- O processo de estimação dos formantes utilizando LPC não é robusto, pois conforme abordado no capitulo 2, os formantes estão ligados aos sons vozeados, não sendo totalmente confiáveis para sons não sonoros. Uma alternativa interessante para um codificador de voz é utilizar parâmetros LSF (line spectrum frequency pairs) [18], os

quais são intimamente ligados aos valores dos formantes no caso de vogais e são estimados de forma robusta no caso de qualquer fonema;

- A proposta deste trabalho limita-se apenas em apresenta um método, sem a preocupação com a robustez e a eficiência do codificador. O método LPC foi escolhido pela sua simplicidade, e desse modo, caberá a novos estudos, a utilização de extratores de formantes mais eficientes, como por exemplo o método já citado LSF .

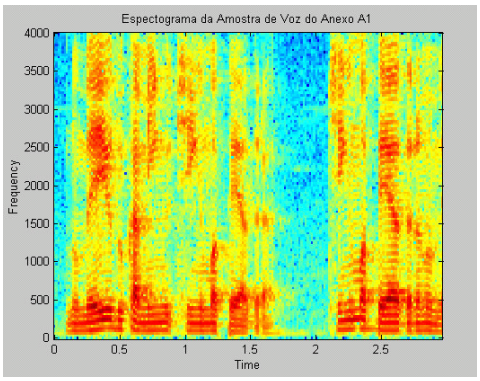


Figura 5.4 - Espectograma da amostra de voz

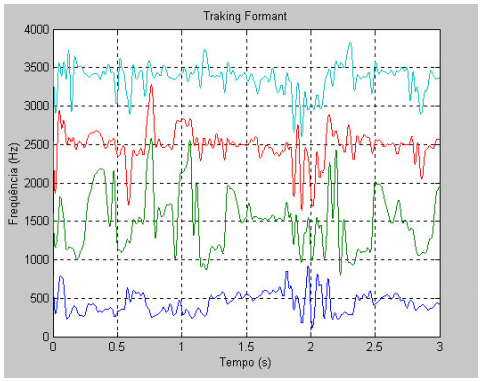


Figura 5.5 – Estimação dos formantes

5.4 Comportamento das Freqüências dos Formantes em Relação as Freqüências Instantâneas

Finalmente, para ilustrar o comportamento das freqüências dos formantes em relação às freqüências instantâneas das quatro faixas filtradas, as figuras 5.6 e 5.7 comparam as mesmas no intervalo de 1 a 1,5 s da amostra 1 do anexo A. Pode-se notar que os sinais FIs da voz apresentam maior variabilidade que as freqüências dos formantes do trato vocal.

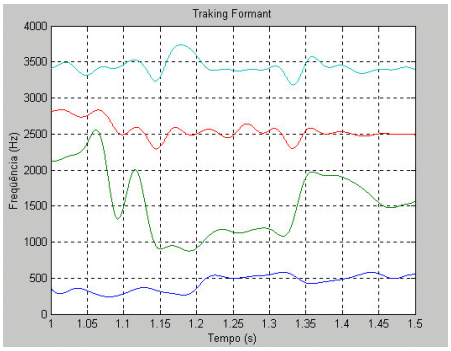


Figura 5.6 - Estimação dos formantes

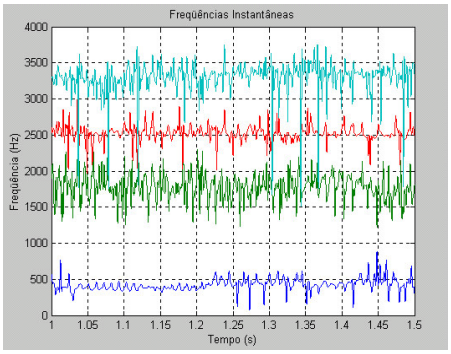


Figura 5.7 – Freqüências instantâneas

6 - CONCLUSÃO

Neste trabalho foi apresentada uma proposta de decomposição dos sinais de voz humana em suas componentes modulantes AM e FM para aplicação em codificadores de voz. O método baseou-se no fato de que a transmissão da voz modulada pode ser substituída pelo envio de seus sinais modulantes, os quais correspondem às freqüências e amplitudes instatâneas, sendo que para extrair estas componentes, o método utilizou o conceito de formantes e aplicou o método LPC para a obtenção das freqüências dos picos do espectro de potência voz. Também foram utilizadas as propriedades de filtragens passa-faixas baseadas em wavelets para a obtenção de faixas estreitas em torno dos picos, e por último, foi aplicada a transformada de Hilbert para se chegar as componentes modulantes.

Os resultados obtidos mostraram que o método é factível de implementação, embora para aplicações de fato seja necessário um aprofundamento maior nas pesquisas para a melhoria da performance do codificador, ou seja, boa inteligibilidade a baixas taxas de transmissão. Com relação à melhoria da inteligibilidade, foi sugerido o emprego de outro método para a estimação de formantes. No que tange a taxa de transmissão estimada, verificou-se que para reduzi-la deve-se diminuir a banda das componentes extraídas, em especial as freqüências instantâneas. Para isso foi sugerida a substituição da transmissão destas pelas fases do sinal.

REFERÊNCIAS

[1] HASEGAWA-JOHNSON, M., Alwan, A. **Speech coding: Fudamentals and Applications.** In: Wiley Encyclopedia of Telecomunications and Signal Processing, NY: J.Proks, NY, dezembro 2002.

[2] TAFNER, MALCON A. **Reconhecimento de Palavras Isoladas Usando Redes Neurais Artificiais,** 1996. Trabalho de Conclusão de Curso Mestrado – Universidade Federal de Santa Catarina, Florianópolis, 1996. Disponível em: www.eps.ufsc.br/disserta96/tafner/index/index.htm

[3] SILVA, A.O.; JOAQUIM, M.B. **Estimação de Freqüências Formantes de Sinais de Voz**.[s.n.].São Paulo.Escola de Engenharia de São Carlos - Universidade de São Paulo, pp.7. Disponível em: http://www.eesc.usp.br/cetepe/cicte/ric/ric1/art-2%28p07%29.pdf

[4] KAISER, J. F. **On a simple algorithm to calculate the 'energy' of a signal.** In: IEEE International Conference on, Acoustics, Speech, and Signal Processing,1990, Albuquerque, New Mexico. Procedings..., Albuquerque, New Mexico, abril 1990, pp. 381-384, abril 1990.

[5] MARAGOS, P.; QUARTIERI, T. F.; KAISER, J. F. **Speech nonlinearities, modulations, and energy operator.** In: Proc. IEEE ICASSP-91,1991, Toronto, Canada, pp.421-424, maio 1991.

[6] MARAGOS, P.; QUARTIERI, T. F.; KAISER, J. F. **On separating amplitude from frequency modulations using energy operators**. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992, Estados Unidos. Proceedings..., Estados Unidos, 1992, pp. II-1-II-4.

[7] MARAGOS, P., KAISER, J. F**. Energy separation in signal modulations with aplication to speech analysis**. In: IEEE Trans. Signals Processing, 1993, 41 v., No.10, Proceedings...,. October 1993, pp.3024-3051.

[8] LU, S.; DOERSCHUK, P. C. **Modeling and processing speech with sums of AM-FM formants models.** In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1995, West Lafayette, Estados Unidos. Proceedings..., West Lafayette, Estados Unidos, maio 1995, pp. 764-767.

[9] LU, S.; DOERSCHUK, P. C. **Nonlinear modeling and processing of speech based on sums of AM-FM formant models.** In: IEEE Trans. Signals Processing, 44 v., No.4, 1996, West Lafayette, Estados Unidos, Proceedings..., West Lafayette, Estados Unidos, abril 1996, pp.773-782.

[10] LU, S.; DOERSCHUK, P. C. **Demodulators for AM-FM models of speech signals: a comparison**. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1996, Estados Unidos, 1 v. Proceeding..., Estados Unidos, maio 1996, pp. 263-266.

[11] HANSON, H. M.; MARAGOS, P.; POTAMIANOS, A.**Finding speech formants and modulations via energy separation: with aplication to a vocoder**. In: IEEE International Conference on Acoustics, Speech, and Signal Processing,1993, vol. II. Proceeding..., 1993, pp. 716-719.

[12] KUMARESAN, R.; RAO, A. **Algorithm for decomposing an analytic signal into AM and positive FM components**. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998, 3 v, maio 1998. Proceeding..., 1998, pp.1561 –1564.

[13] KUMARESAN, R.; RAO, A. **On decomposing speech into modulated components**, In: Proc. IEEE Trans. On Speech and Audio Processing, 8 v, n.3,p. 240-254, maio 2000.

[14] BARROS, A.K.; WISBECK, J. OHNISHI, N., **Heart Instantaneous frequency (HIF): An Alternative Approach to Extract Heart Rate Variability**, In: IEEE Trans. On Biomedical Engineering, 48 v, n 8, Proceedings..., agosto 2001, p. 850-855.

[15] BARROS, A.K.; OHNISHI, N. **Amplitude Estimation of Quasi-Periodic Physiological Signals by Wavelets**. In: IEICE Trans. Fundamentals, E82 v, Janeiro 1999.

[16] FAWE, A.L. **Analytical speech encoding**. In: *Electrotechnical Conference - MELECON '96, 8th Mediterranean,1996,* 3 v, pp. 1690-1693, maio 1996.

[17] CARLSON, A. B.; CRILLY, P. B.; RUTLEDGE, J. C. **Communication Systems**, 4º ed, McGraw-Hill, p. 199-202, 2002.

[18] HUANG, X.; ACERO, A.; HON, HSIAO-WUEN. **Spoken Language Processing: a guide to theory, algorithm, and system development,** ed.Prentice-Hall, p.303-305, 2001.

[19] RIBEIRO, C.E.M. **Predição Linear – Aplicações na Codificação de Fala.** In: Jornadas Matemáticas – Instituto de Engenharia de Lisboa, Lisboa, 1998. Disponível em: www.deetc.isel.ipl.pt/comunicacoesep/disciplinas/pdf/cmr/publica

[20] S. WARDLE. **A Hilbert-Transformer Frequency Shifter for Audio.** Available online at http://www.iua.upf.es/dafx98/papers.

[21] SMITH, J.O. **Generalized Complex Sinusoids**.[2001?]. Disponível em: *http://ccrma-www.stanford.edu/~jos/mdft/Generalized_Complex_Sinusoids.html*

[22] LATHI,B.P. **Modern Digital and Analog Communication Systems,** 3º , Oxforrd University Press, p. 172-174, 1998.

[23] GERSHO, A. **Advances in Speech and Audio Compression. In:** Proceedings of the IEEE, 1994, Santa Bárbara, CA, 1994, v. 82, n.6, p. 900-918.

[24] HERSENT, O.; GUIDE, D.; PETIT, J. **Telefonia IP comunicação multimídia baseada em pacotes**, Addison Wesley. São Paulo, 2002.

[25] AHONEN, J.; LAINE, A. **Realtime Speech and Voice Transmission on the Internet**. [s.n], Helsinki, Helsinki University of Technology. Telecomunications Software and Multimedia Laboratory. Disponível em: www.tcm.hut.fi/opinnot/tik-110.551/1997/seminar_paper.html

[26] MARTINO, E.; ROCHA, F. M. F.; et all; **Codificação de Voz Para Sistema de Atendimento de Pequenas Localidades**. In: Revista Telebrás Tecnologia, ed.out, Brasília, DF, 1992, p.119-128.

[27] OLIVEIRA, S. **Voz**. [2001?], Belo Horizonte, MG, Disponível em: www.lecom.dcc.ufmg.br/~sergiool/telefonia/voz.html

[28] RABINER L. R.; SCHAFER, R. W. **Digital Processing of Speech Signals. Englewood Cliffs,** Nj: Prentice Hall signal processing series, p. 398-404, 1978.

[29] OPPENHIMEIM, A.V.; SHAFER, R. W. **Discrete-Time Sinal Processing**. Ed. Pretice-Hall, 1999.

[30] BABEDO, J. **Avaliação Objetiva de Qualidade de CODECS de Voz na Faixa de Telefonia**. 2001. p.128, Trabalho de Conclusão de Curso Mestrado – Universidade Estadual de Campinas, Campinas, 2001.

[31] CAMPOS NETO, S. F. **Metodologias de Avaliação de Algoritmos de Codificação de Voz.** 1993. p.160,Trabalho de Conclusão de Curso Mestrado – Universidade Estadual de Campinas, Campinas, 1993.

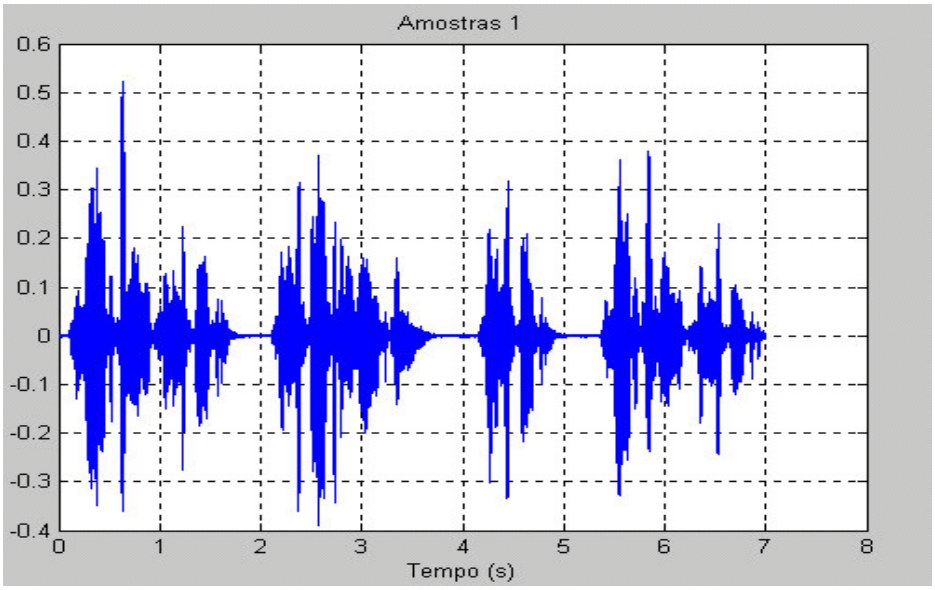
[32] AGUIAR NETO, BENEDITO G. Complexidade dos Codificadores. In: **Curso de Especialização em Comunicação de Dados - Modulo: Transmissão Digital em Banda Básica.** São Luís: UFMA, maio1995. p. 29.

APÊNDICE A – RELAÇÃO DE AMOSTRAS UTILIZADAS PARA VALIDAR O MÉTODO

AMOSTRA 1

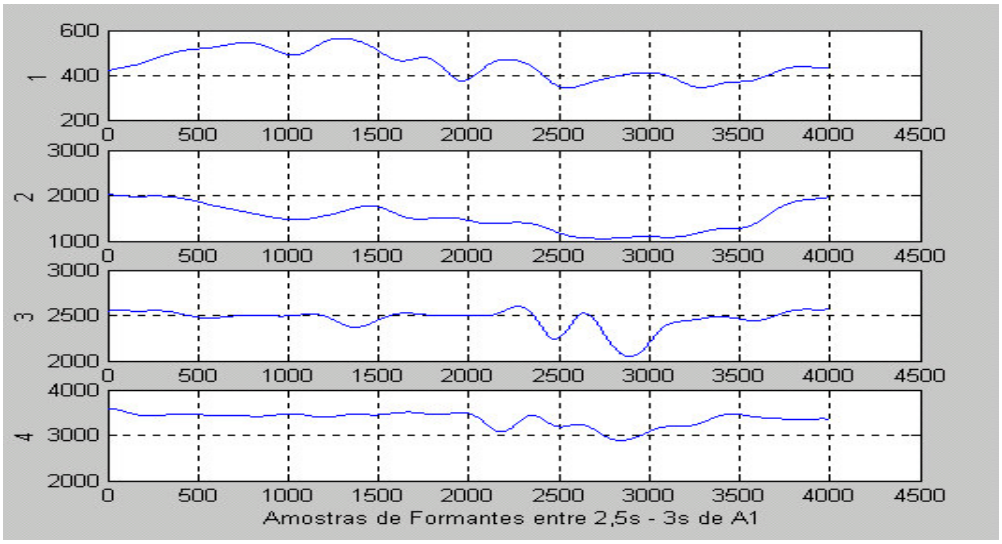
"No meio do caminho tinha uma pedra, tinha uma pedra no meio do caminho, tinha uma pedra, no meio do caminho tinha uma pedra"

1º Sinal original a 48Kbps e16 bits:

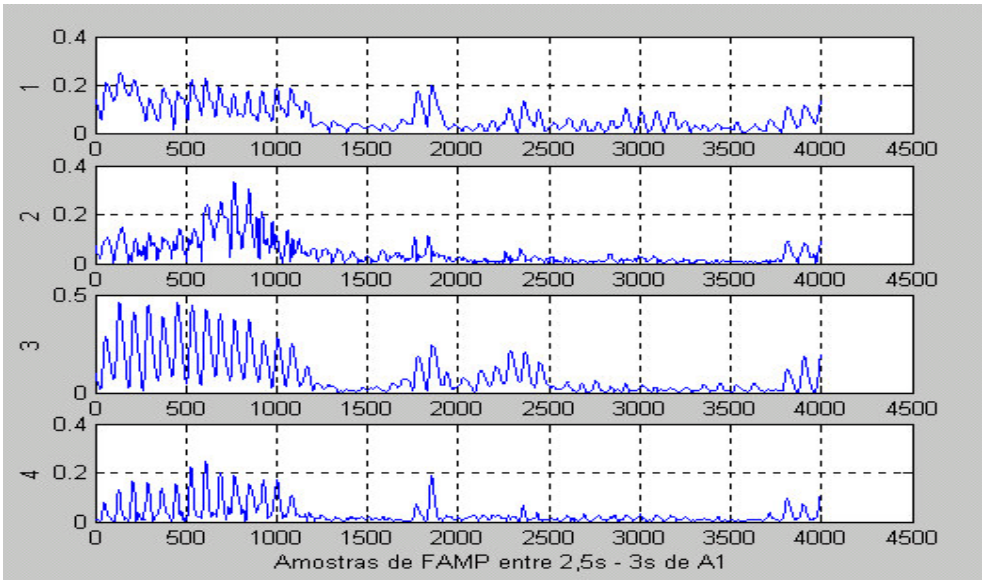


pedra.wav

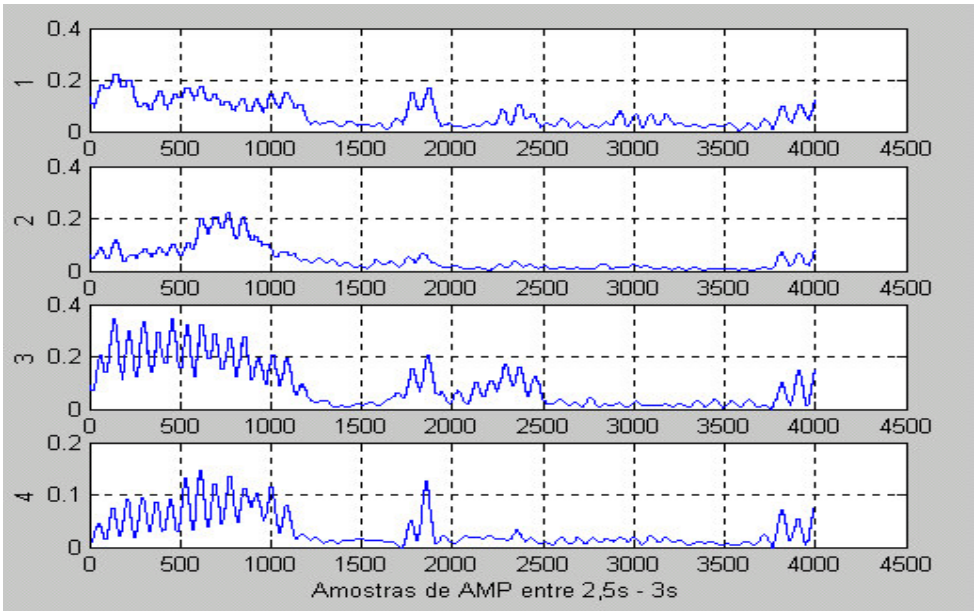
2º Traking Formant entre 2,5 e 3 s da amostra 1



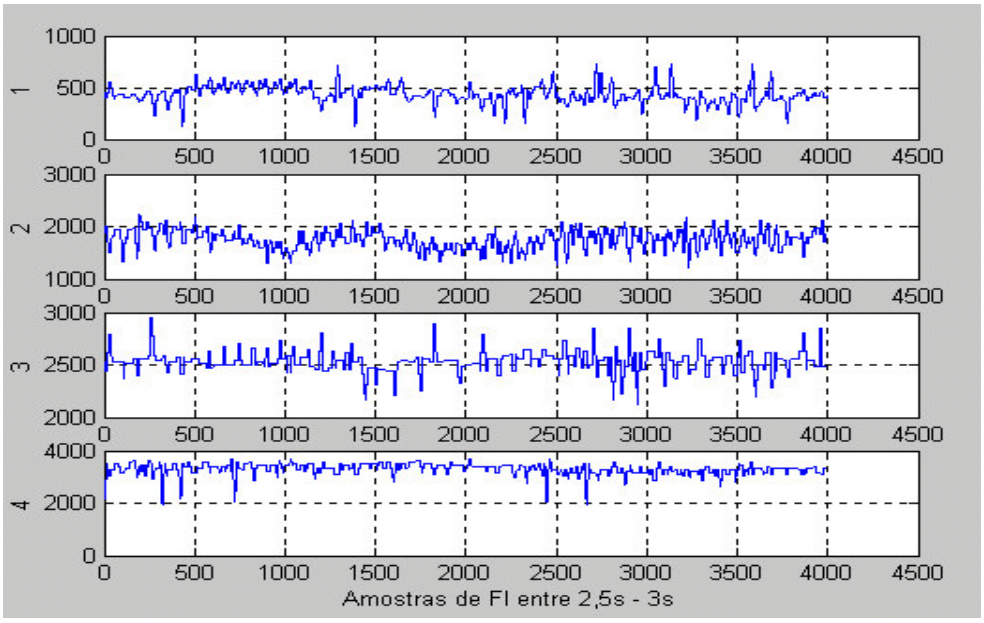
3º Amplitudes instatâneas entre 2,5 e 3 s da amostra 1 sem restrições de largura de banda e quantização:



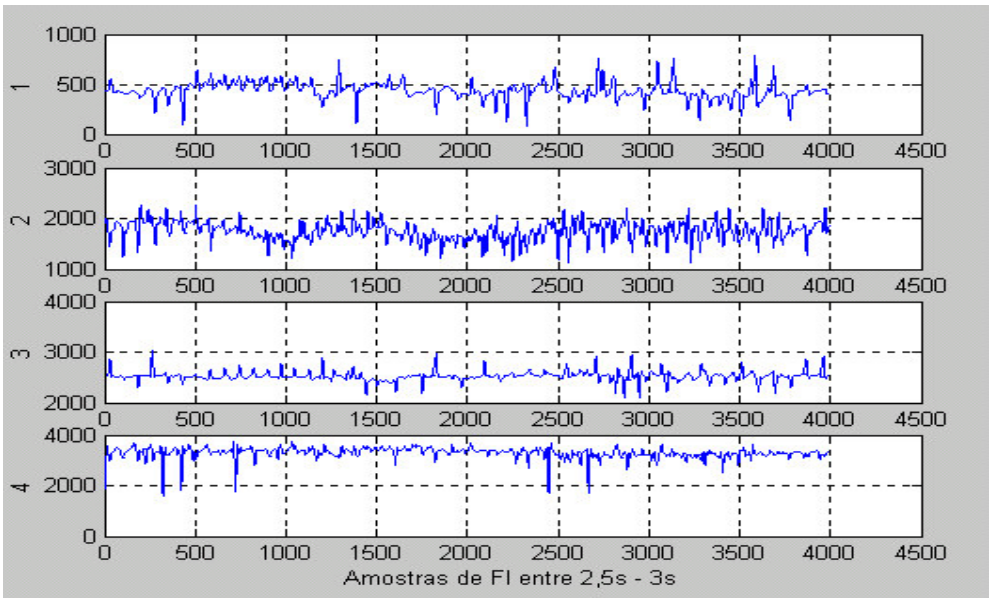
4º Amplitudes instatâneas entre 2,5 e 3 s da amostra 1 com filtragem a 100 Hz com quantização a 4 bits:



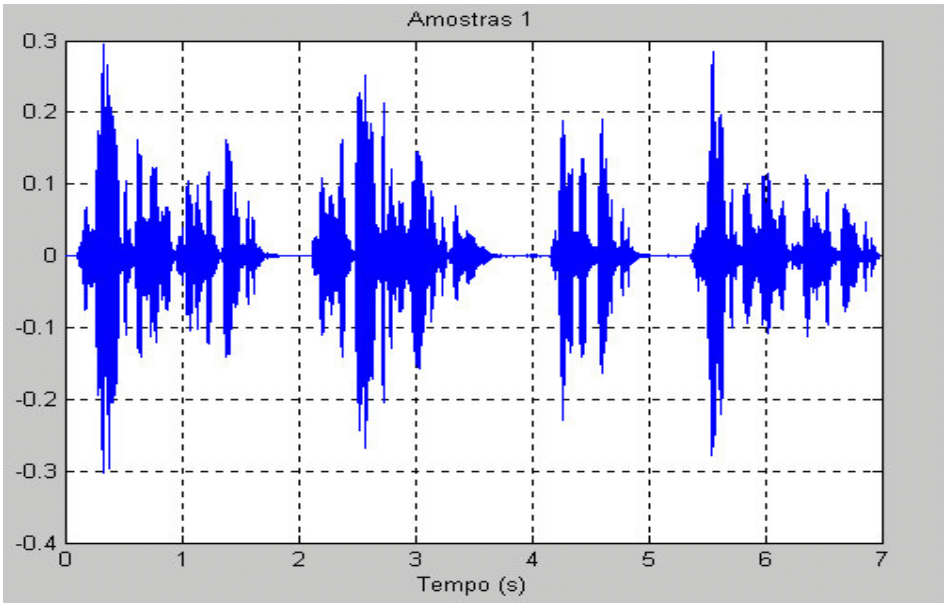
5º Freqüências instatâneas entre 2,5 e 3 s da amostra 1 sem restrições de largura de banda e quantização:



6º Freqüências instatâneas entre 2,5 e 3 s da amostra 1 com filtragem a 100 Hz com quantização a 4 bits:

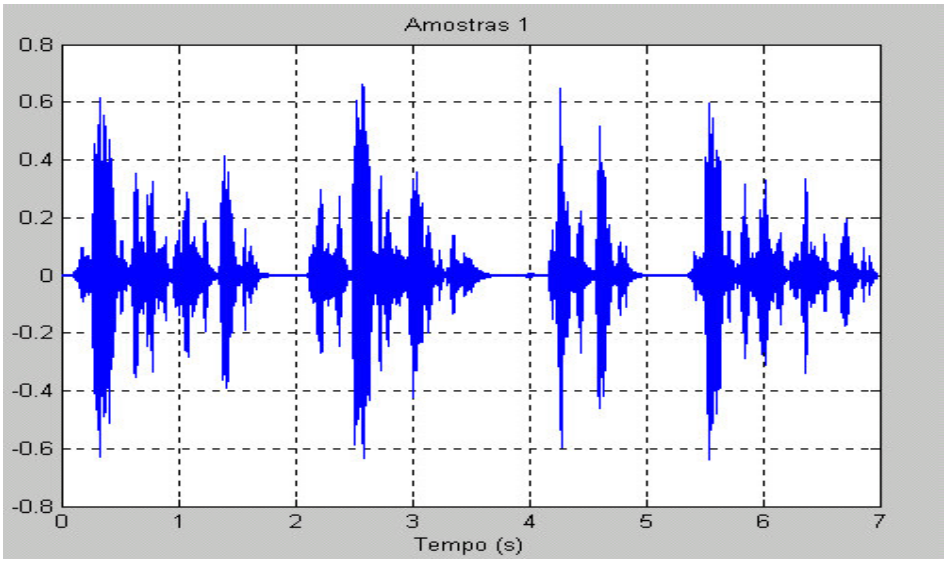


7º Sinal recuperado a 8Kbps e 8bits a partir dos sinais modulantes sem restrições de largura de banda e quantização dos sinais AMPs e FIs:



pedra_h.wav.

8º Sinal recuperado a partir dos sinais modulantes com sinais AMPs filtrados a 100 Hz com quantização a 4 bits e FIs a 400 Hz com 5 bits.

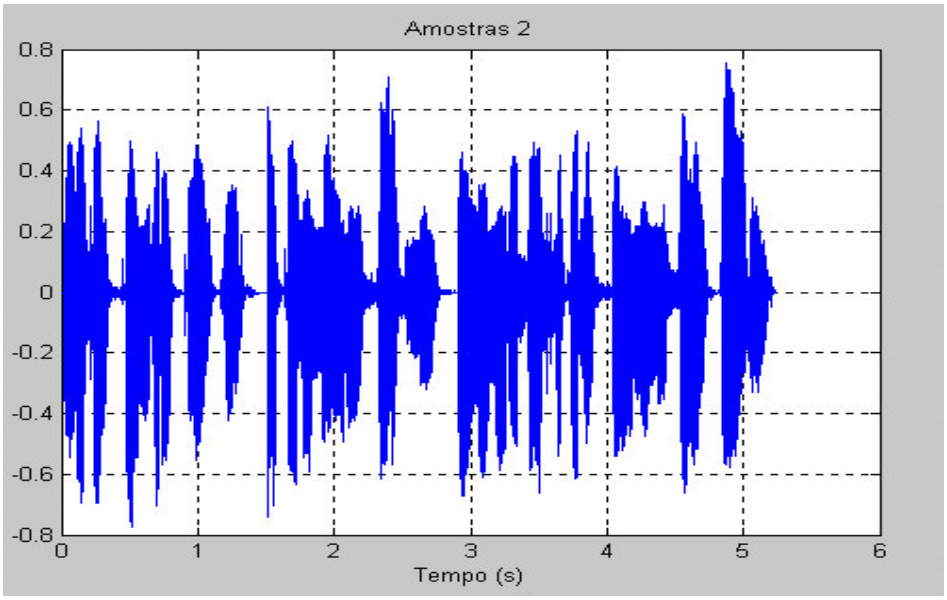


pedra_g.wav.

AMOSTRA 2

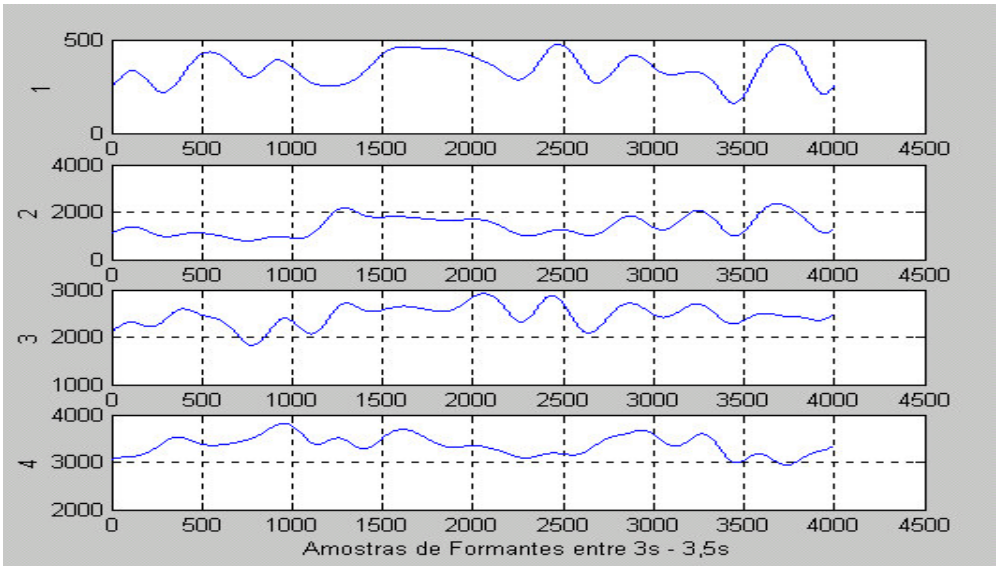
"Ligação gratuita, a Telemar informa: o número deste telefone mudou para"

1º Sinal original a 48Kbps e16 bits sub amostrado em 8Kbps e 8bits:

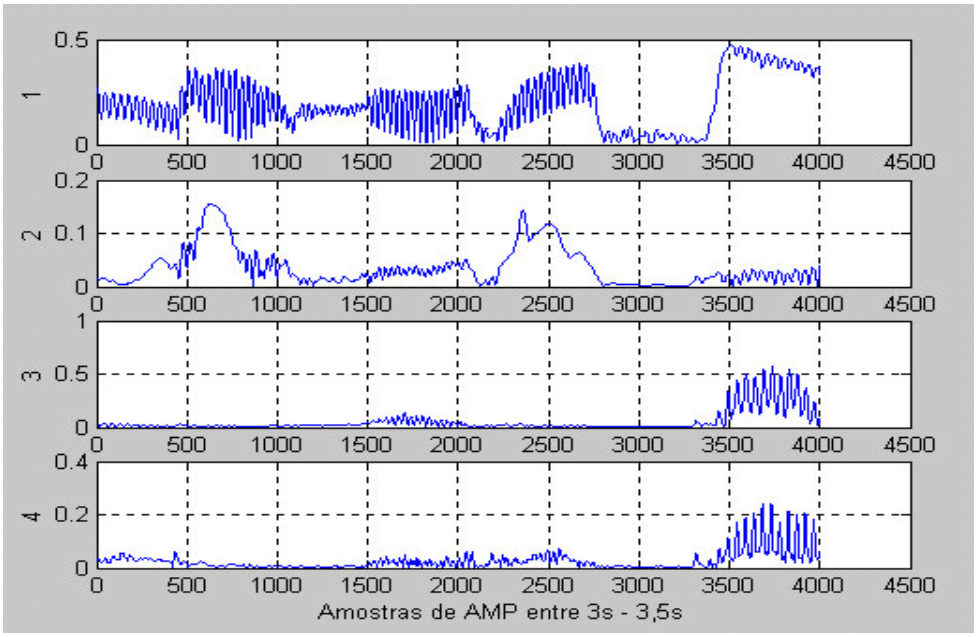


ligacao.wav

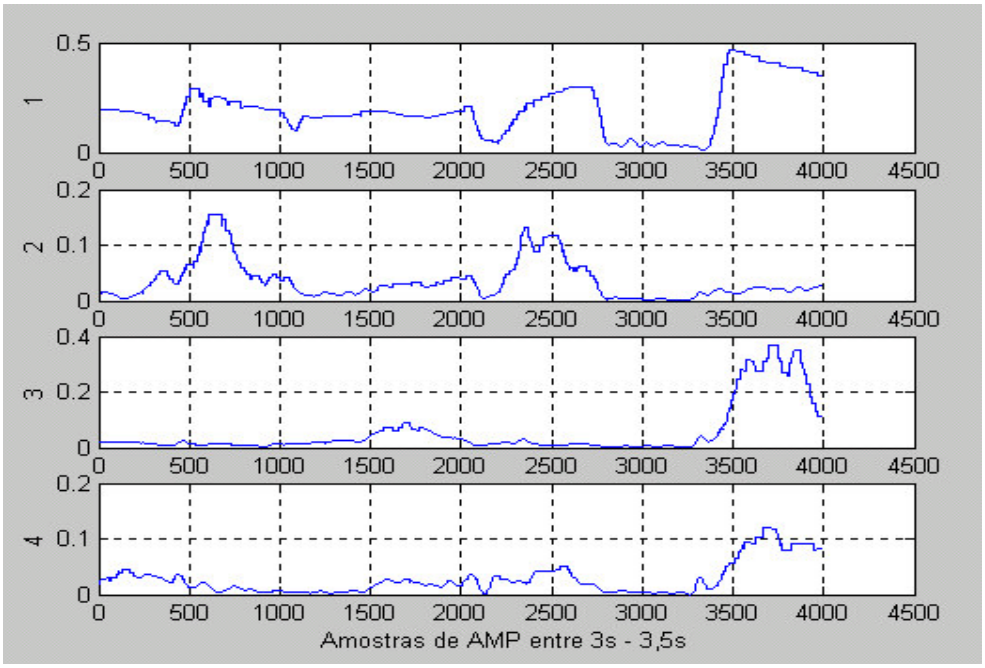
2º Traking Formant entre 3 e 3,5 s da amostra 2



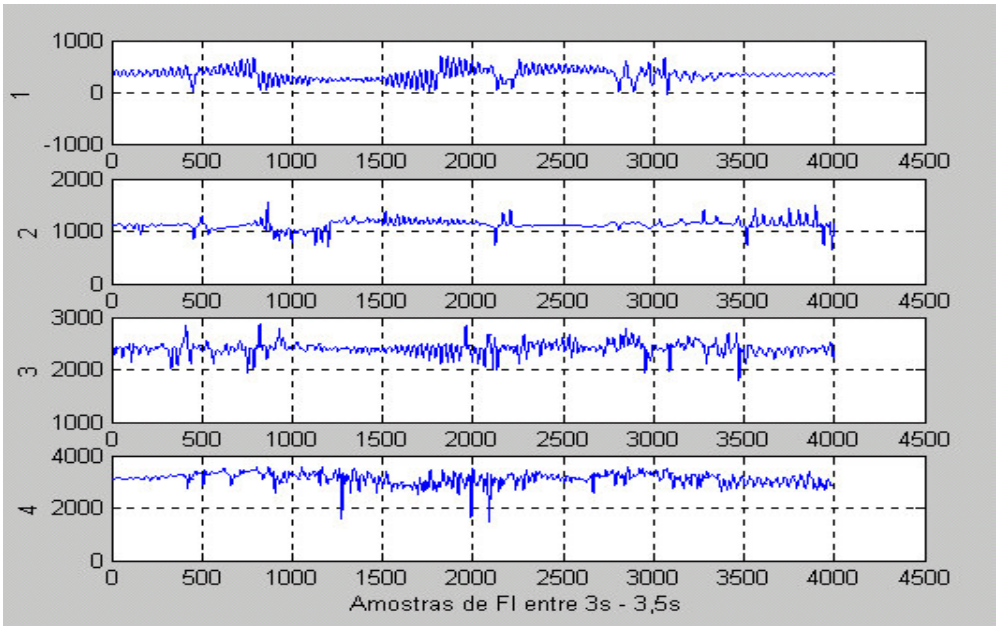
3º Amplitudes instatâneas entre 3 e 3,5 s da amostra 2 sem restrições de largura de banda e quantização:



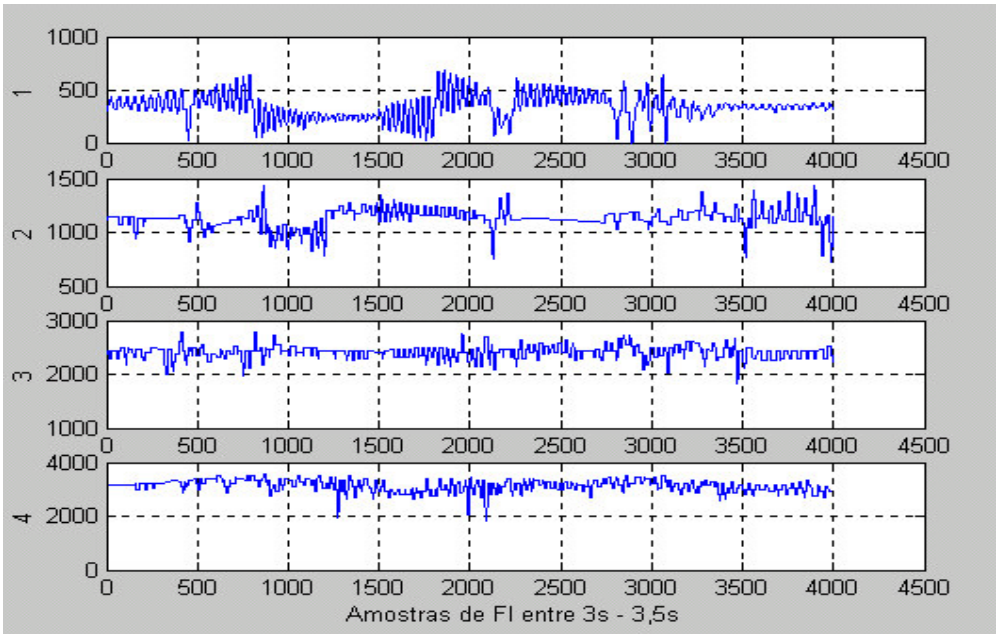
4º Amplitudes instatâneas entre 3 e 3,5 s da amostra 2 com filtragem a 100 Hz com quantização a 4 bits:



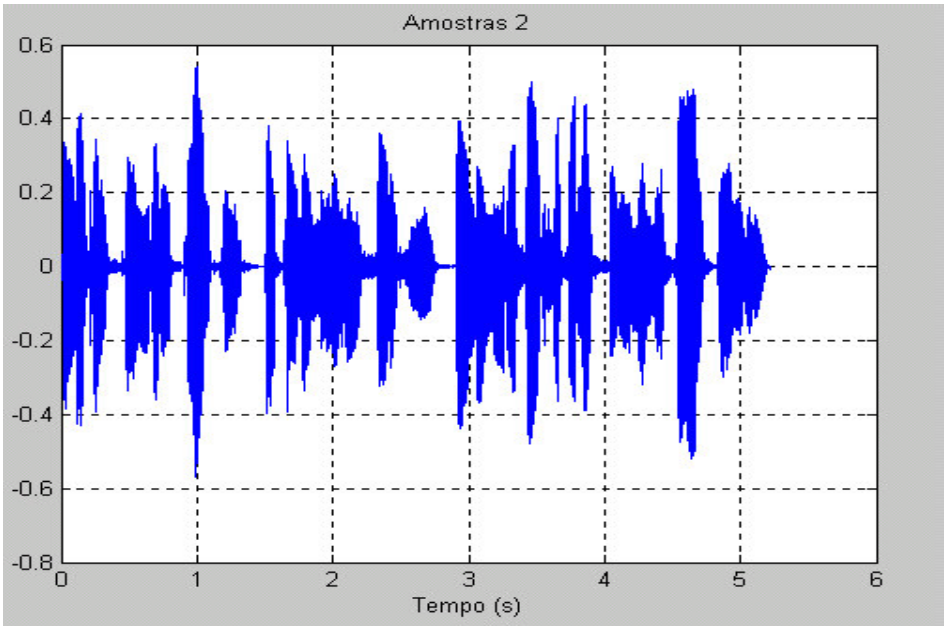
5º Freqüências instatâneas entre 3 e 3,5 s da amostra 2 sem restrições de largura de banda e quantização:



6º Freqüências instatâneas entre 3 e 3,5 s da amostra 2 com filtragem a 100 Hz com quantização a 4 bits:

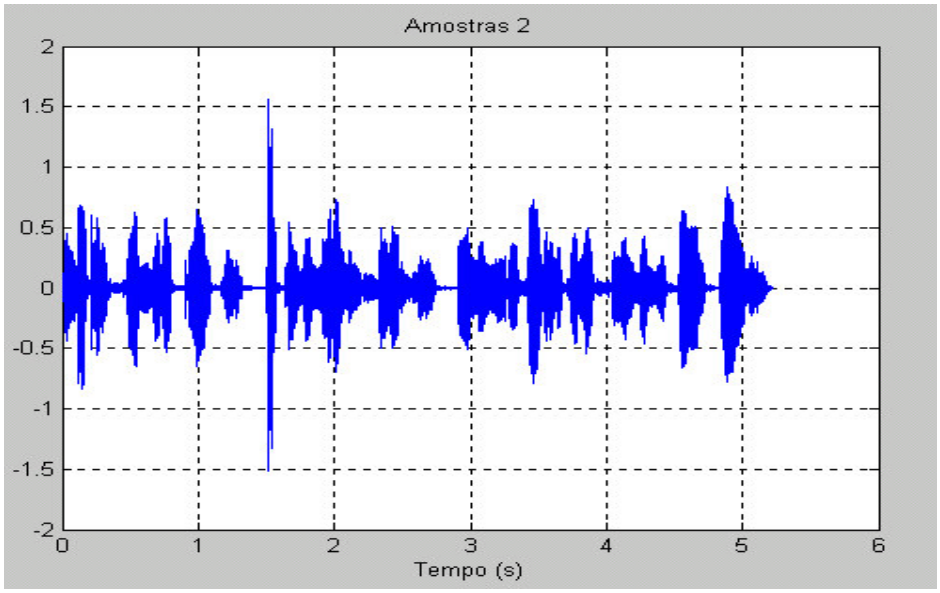


7º Sinal recuperado a 8Kbps e 8bits a partir dos sinais modulantes sem restrições de largura de banda e quantização dos sinais AMPs e FIs:



ligacao_h.wav

8º Sinal recuperado a partir dos sinais modulantes com sinais AMPs filtrados a 100 Hz com quantização a 4 bits e FIs a 400 Hz com 5 bits.

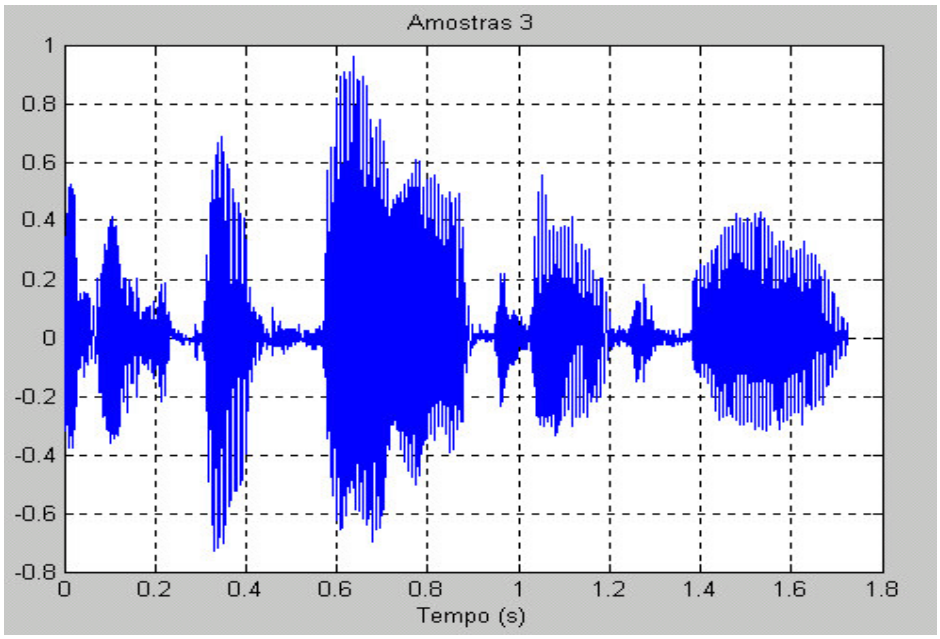


ligação_g.wav

AMOSTRA 3

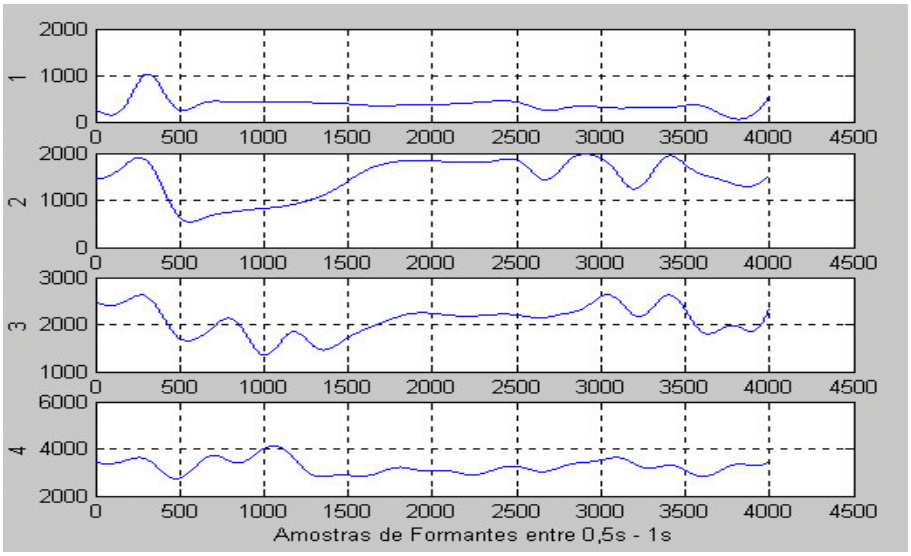
"The discrete Fourier Transform."

1º Sinal original a 48Kbps e16 bits sub amostrado em 8Kbps e 8bits:

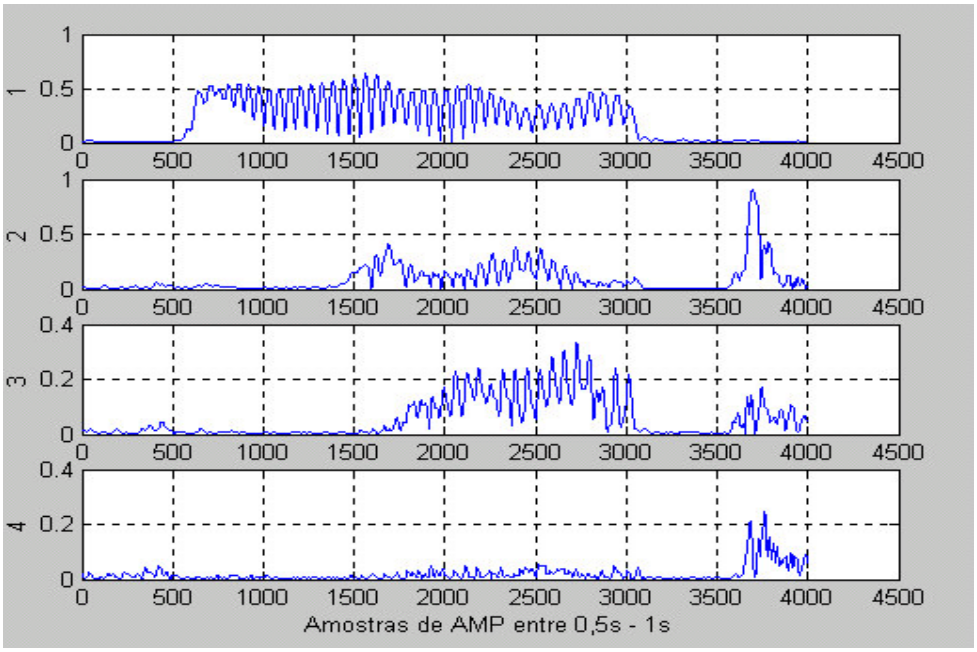


wakosound.wav

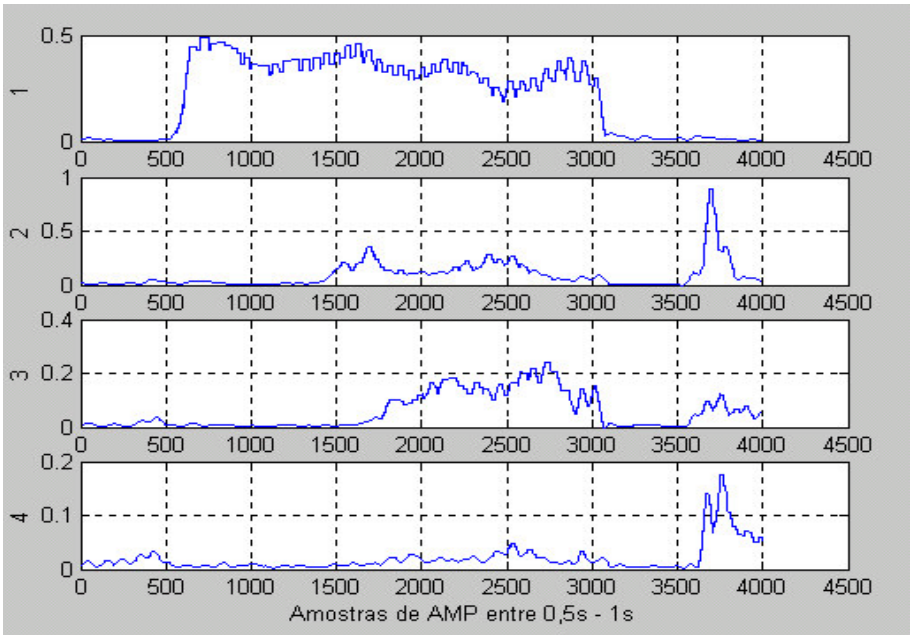
2º Traking Formant entre 0,5 e 1 s da amostra 3



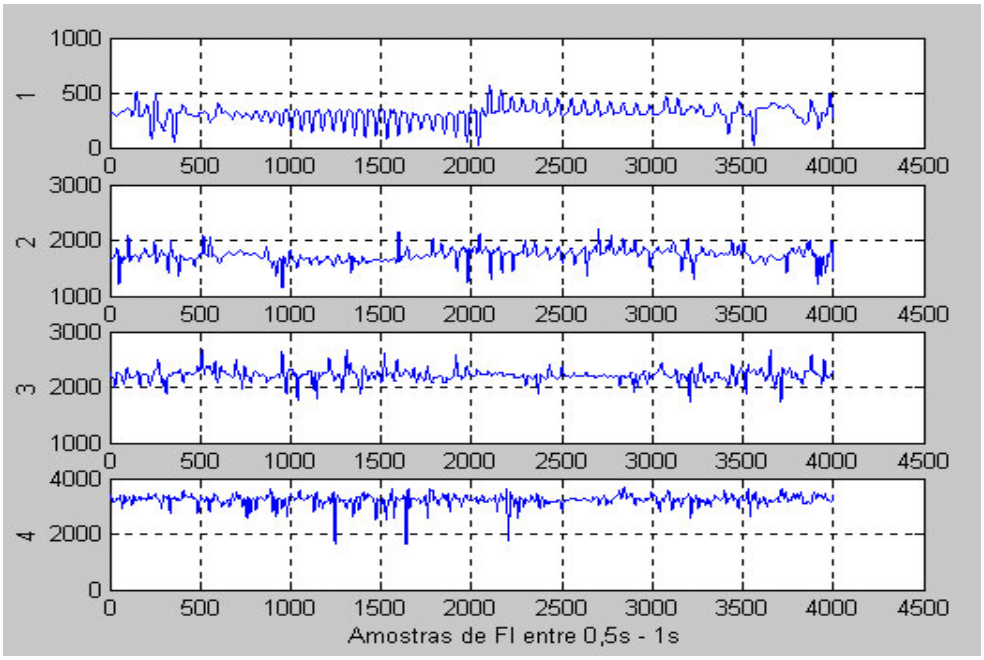
3º Amplitudes instatâneas entre 0,5 e 1 s da amostra 3 sem restrições de largura de banda e quantização:



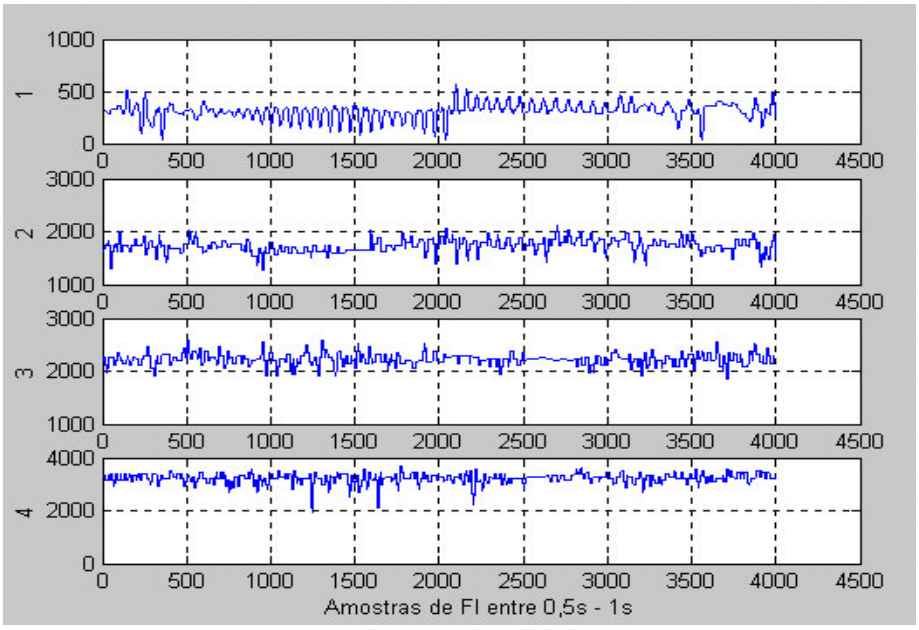
4º Amplitudes instatâneas entre 0,5 e 1 s da amostra 3 com filtragem a 100 Hz com quantização a 4 bits:



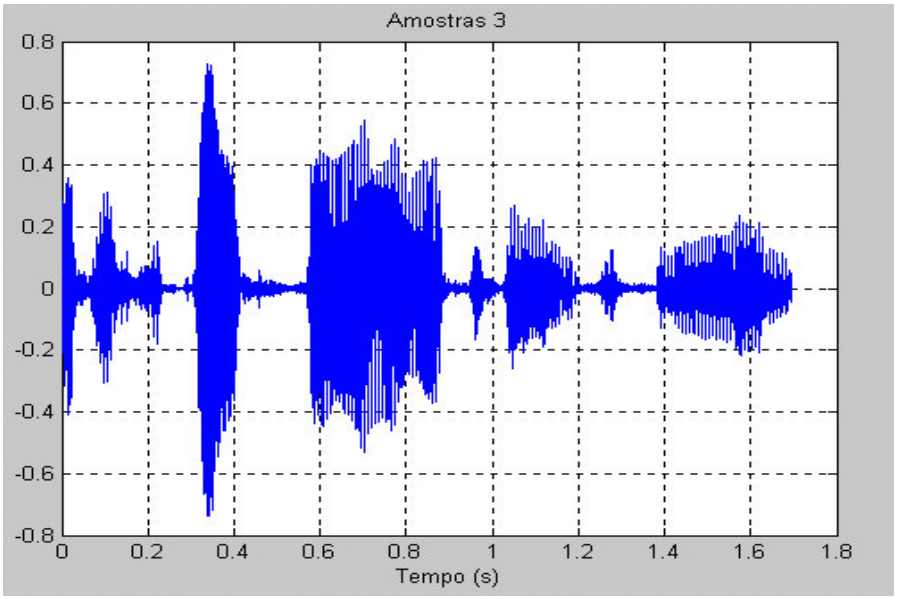
5º Freqüências instatâneas entre 0,5 e 1 s da amostra 3 sem restrições de largura de banda e quantização:



6º Freqüências instatâneas entre 0,5 e 1 s da amostra 3 com filtragem a 100 Hz com quantização a 4 bits:

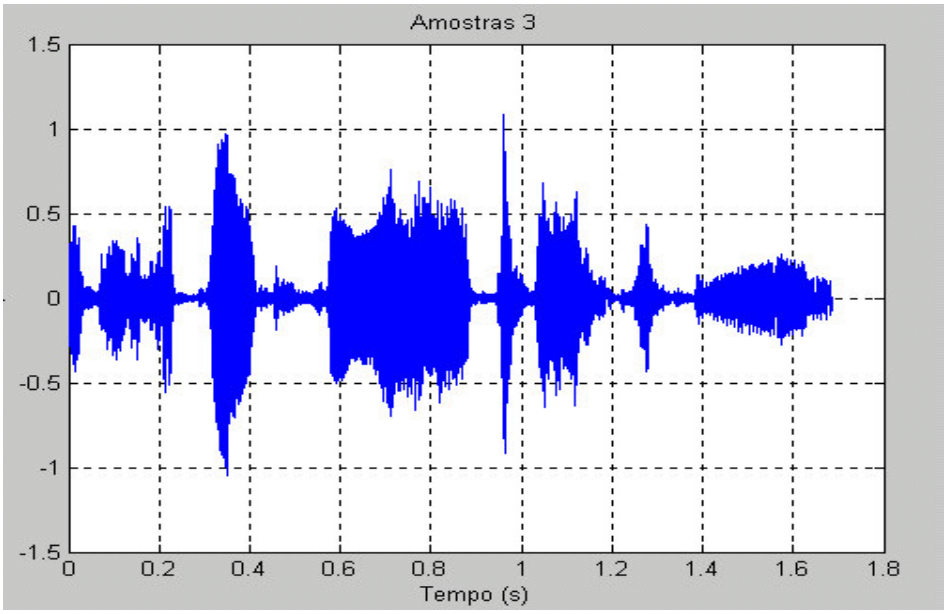


7º Sinal recuperado a 8Kbps e 8bits a partir dos sinais modulantes sem restrições de largura de banda e quantização dos sinais AMPs e FIs:



wakosound_h.wav

8º Sinal recuperado a partir dos sinais modulantes com sinais AMPs filtrados a 100 Hz com quantização a 4 bits e FIs a 400 Hz com 5 bits.

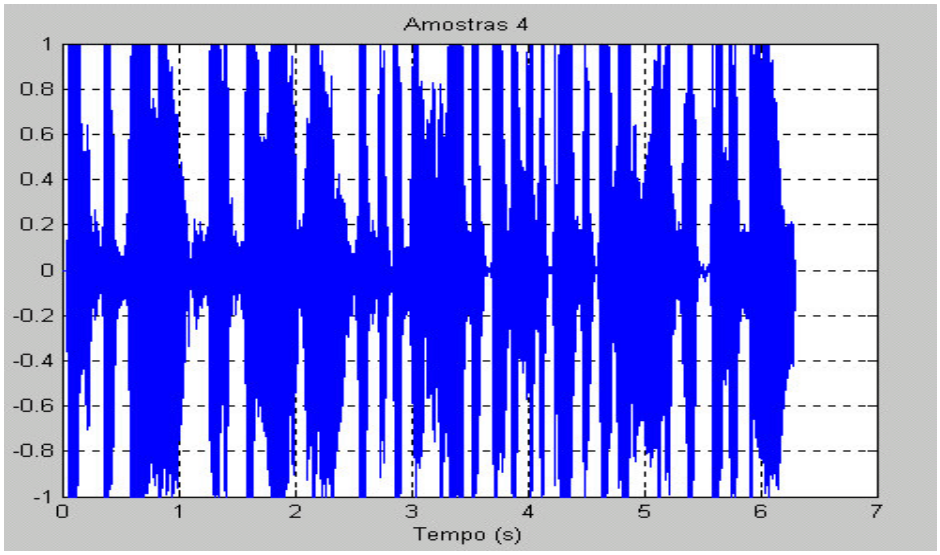


wakosound_g.wav

AMOSTRA 4

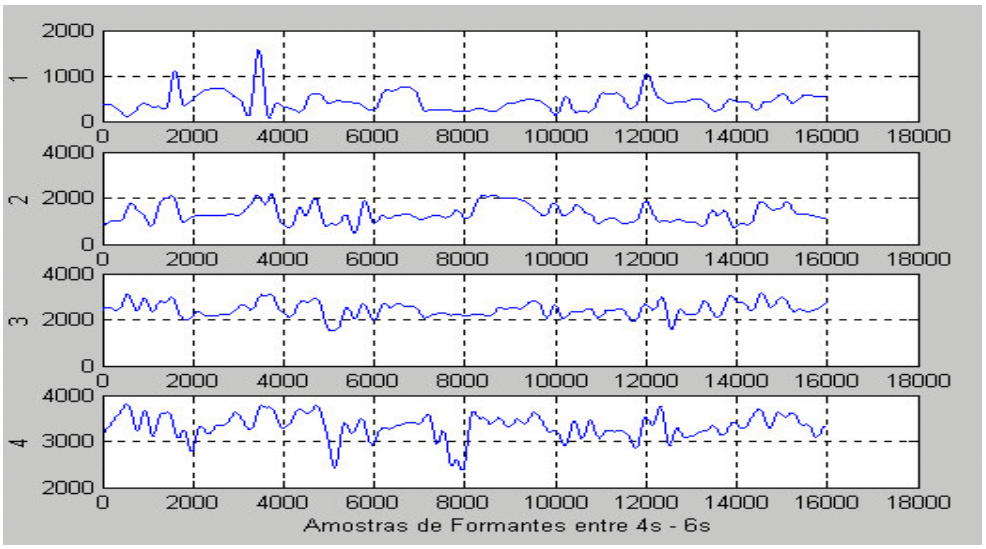
"Conseguimos, isso mesmo, você pediu e nós estendemos o prazo com a mesma promoção"

1º Sinal original a 48Kbps e16 bits sub amostrado em 8Kbps e 8bits:

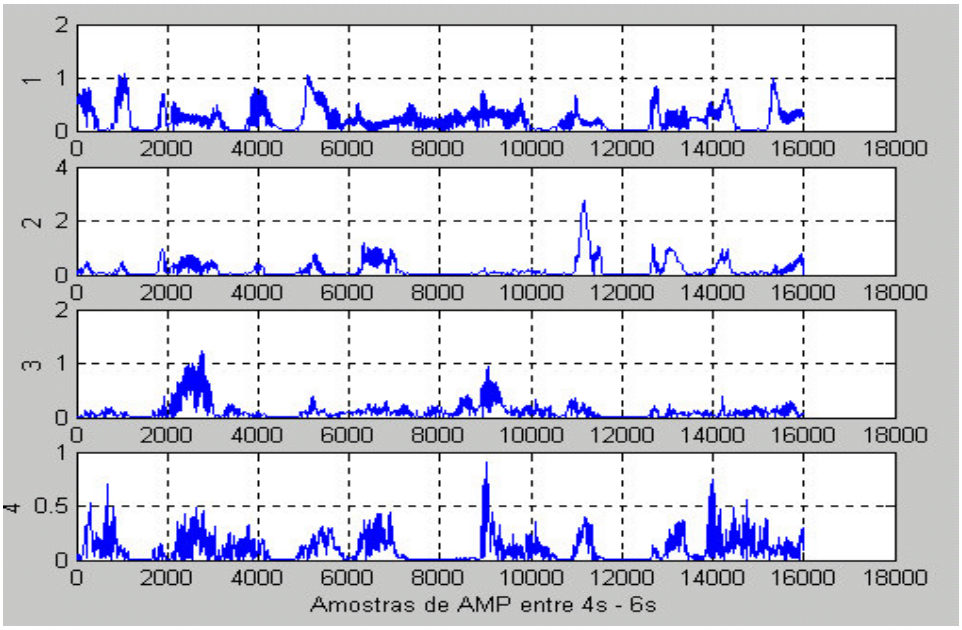


promoção.wav

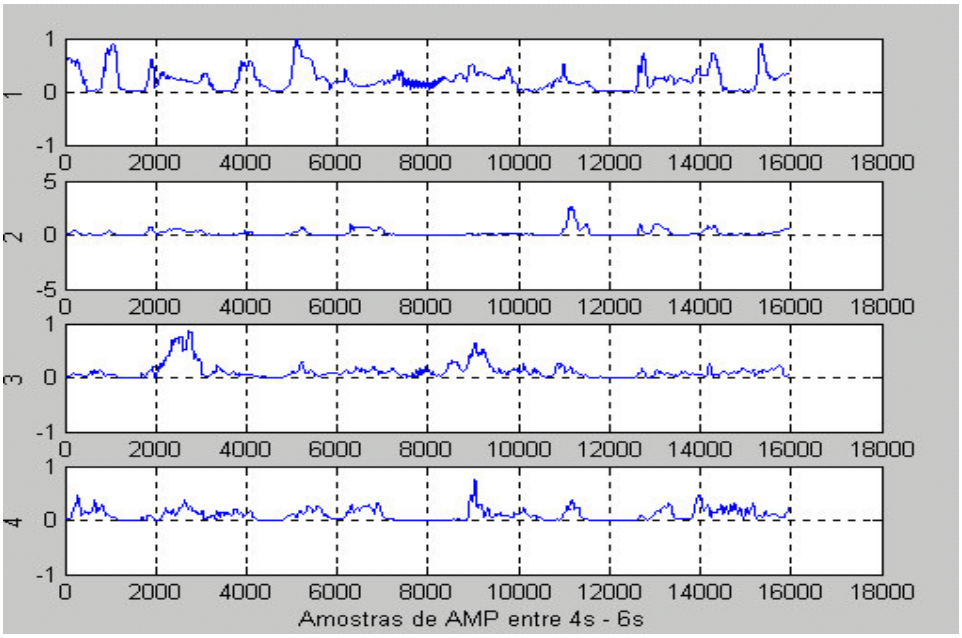
2º Traking Formant entre 4 e 6 s da amostra 4



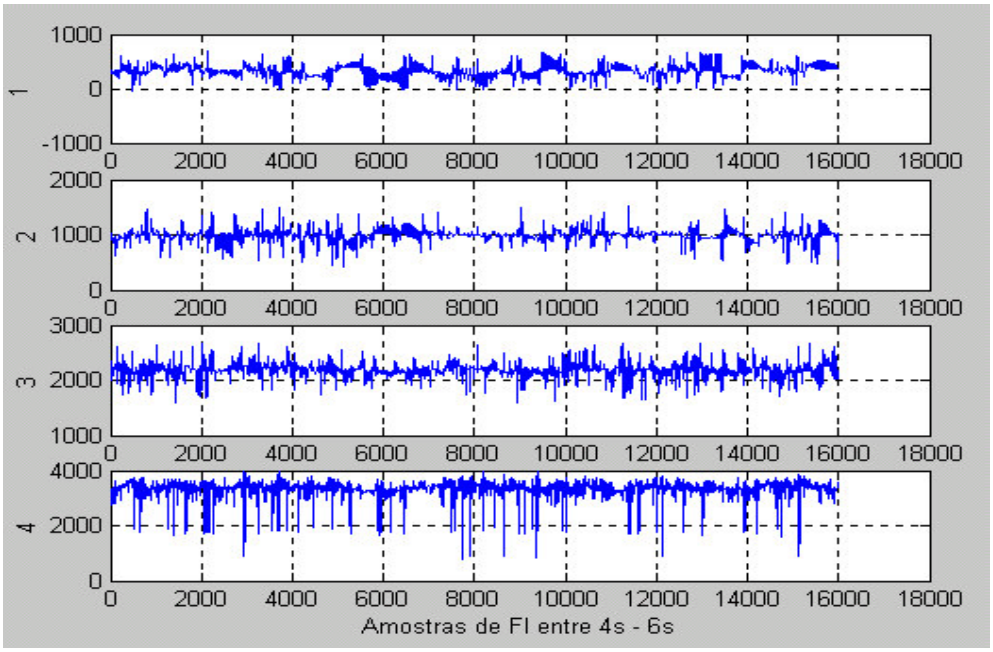
3º Amplitudes instatâneas entre 4 e 6 s da amostra 4 sem restrições de largura de banda e quantização:



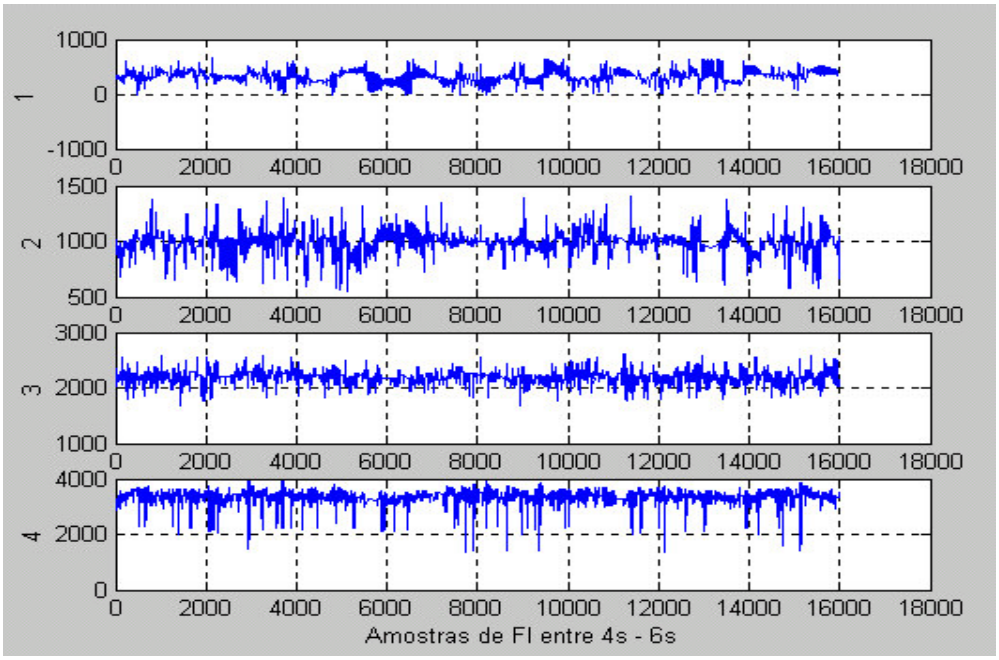
4º Amplitudes instatâneas entre 4 e 6 s da amostra 4 com filtragem a 100 Hz com quantização a 4 bits:



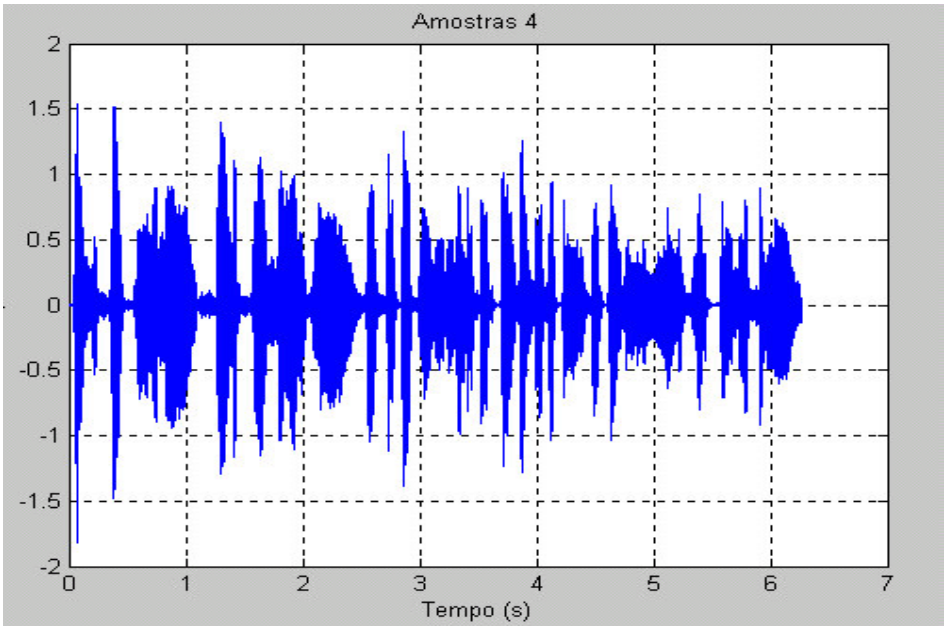
5º Freqüências instatâneas 4 e 6 s da amostra 4 sem restrições de largura de banda e quantização:



6º Freqüências instatâneas entre 4 e 6 s da amostra 4 com filtragem a 100 Hz com quantização a 4 bits:

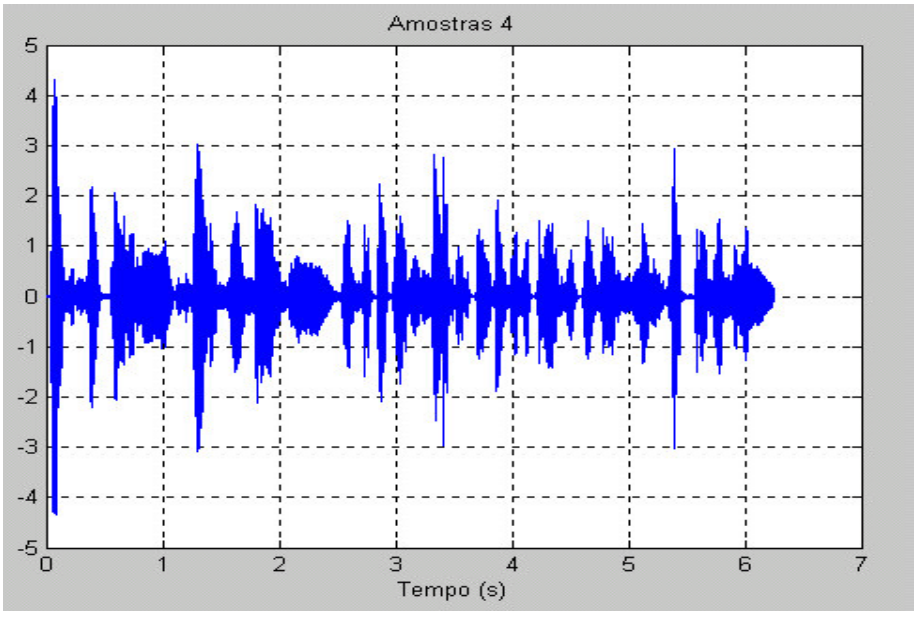


7º Sinal recuperado a 8Kbps e 8bits a partir dos sinais modulantes sem restrições de largura de banda e quantização dos sinais AMPs e FIs:



promocao_h.wav

8º Sinal recuperado a partir dos sinais modulantes com sinais AMPs filtrados a 100 Hz com quantização a 4 bits e FIs a 400 Hz com 5 bits.

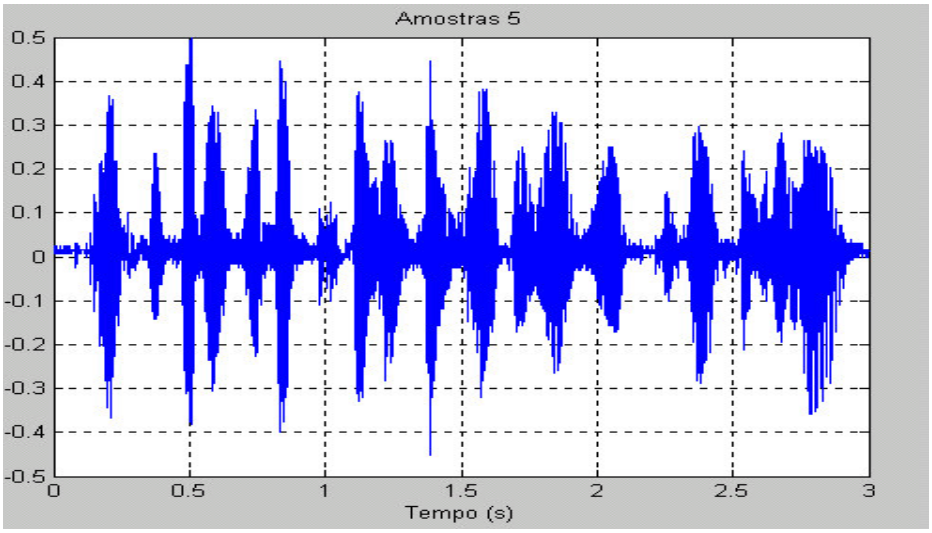


promocao_g.wav

AMOSTRA 5

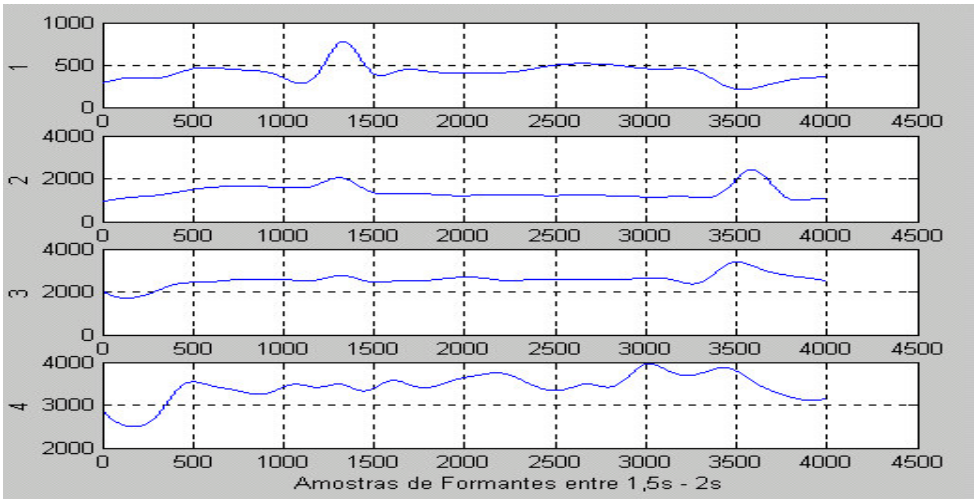
"A essência da vida está nas borboletas azuis que voam sobre o bar"

1º Sinal original a 48Kbps e16 bits sub amostrado em 8Kbps e 8bits:

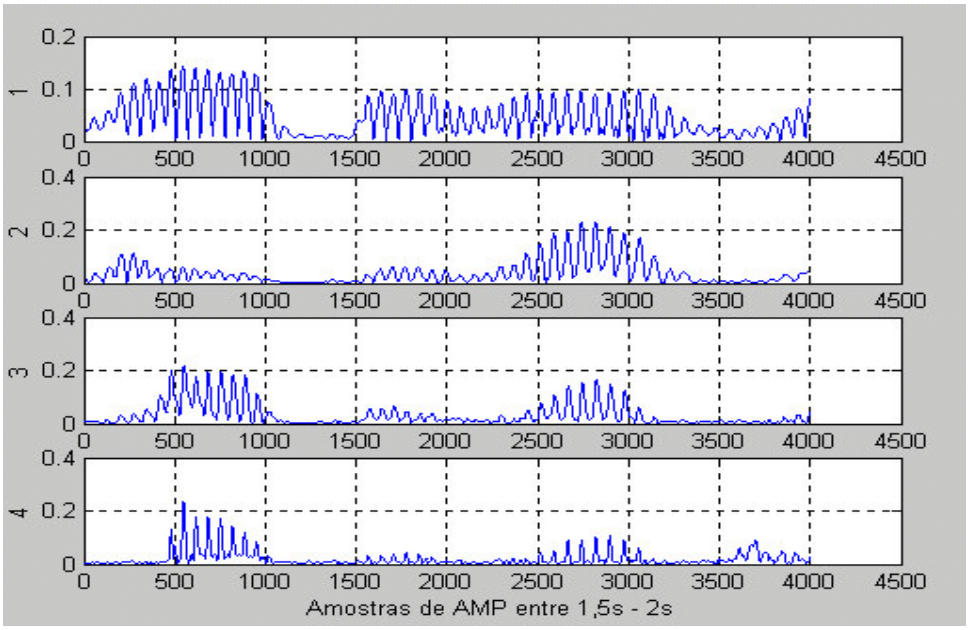


borboleta.wav

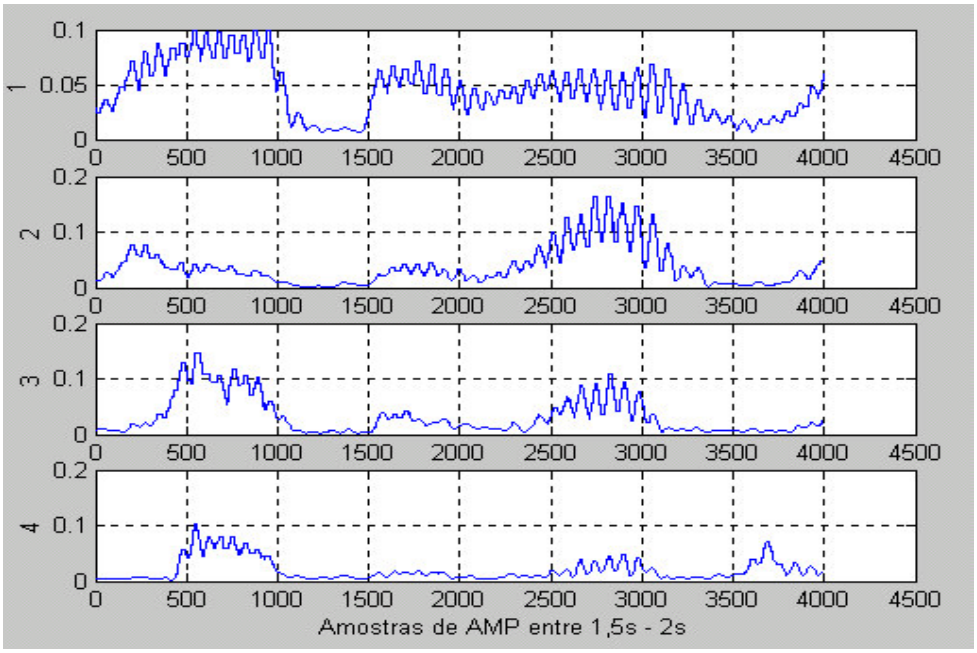
2º Traking Formant entre 1,5 e 2 s da amostra 5



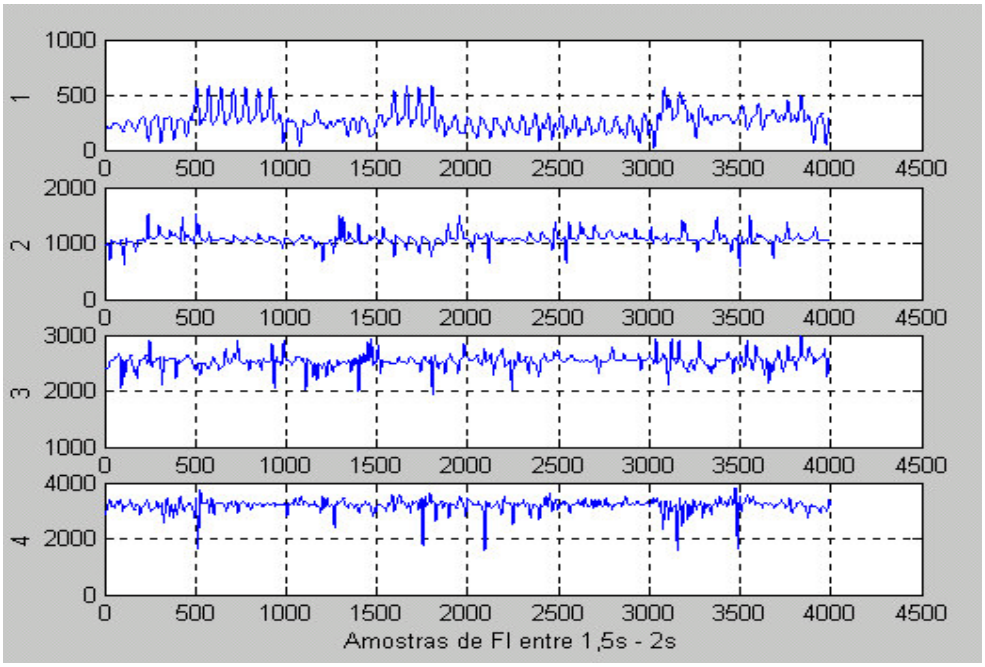
3º Amplitudes instatâneas entre 1,5 e 2 s da amostra 5 sem restrições de largura de banda e quantização:



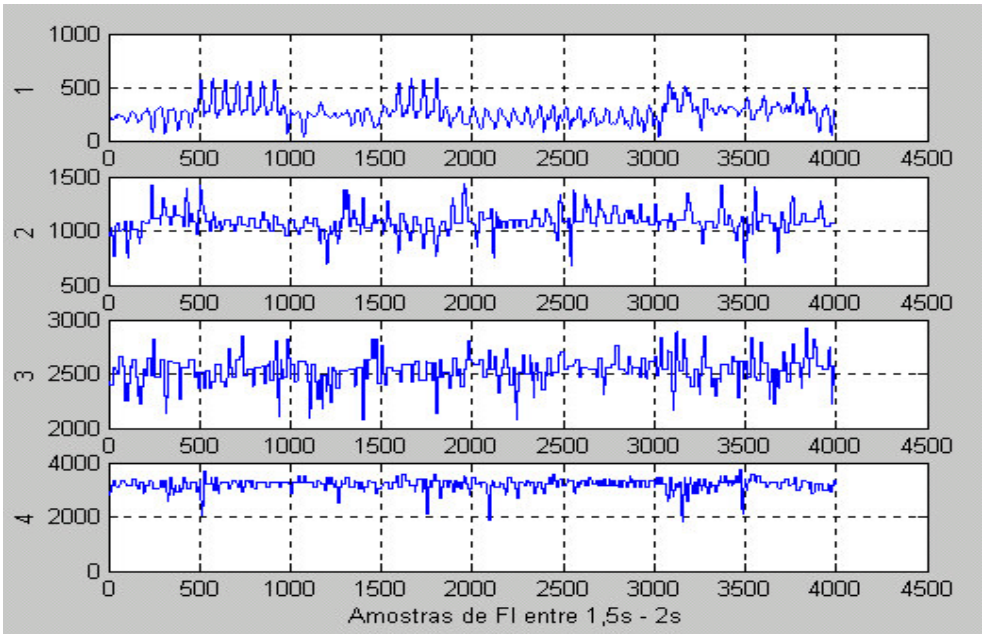
4º Amplitudes instatâneas entre 1,5 e 2 s da amostra 5 com filtragem a 100 Hz com quantização a 4 bits:



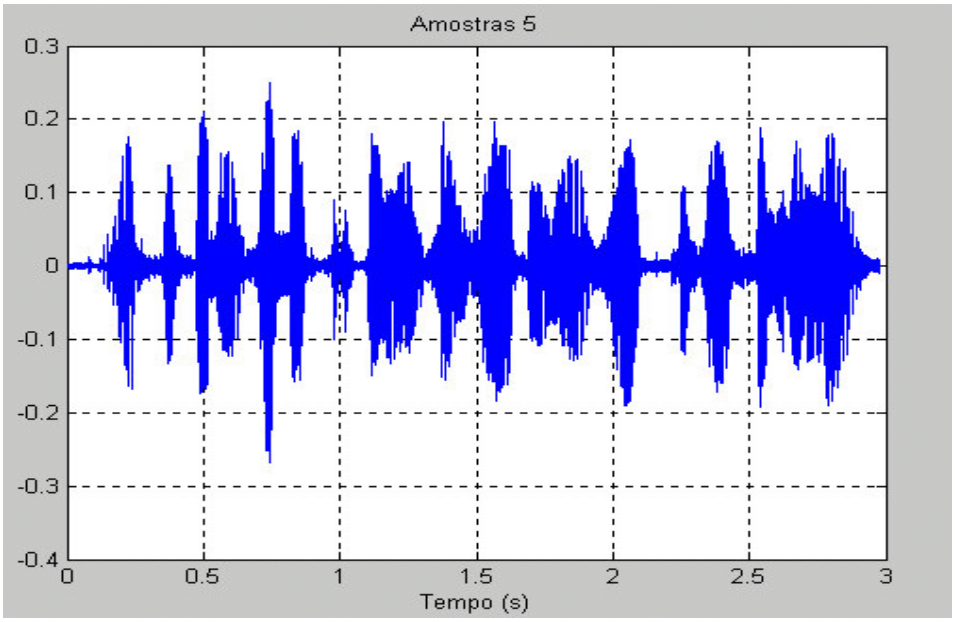
5º Freqüências instatâneas 1,5 e 2 s da amostra 5 sem restrições de largura de banda e quantização:



6º Freqüências instatâneas entre 1,5 e 2 s da amostra 5 com filtragem a 100 Hz com quantização a 4 bits:

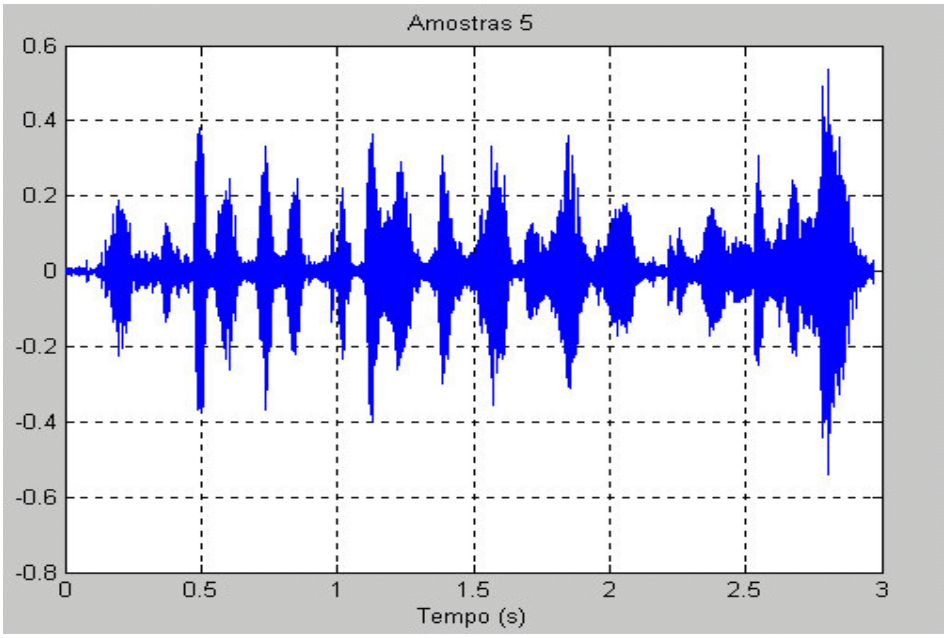


7º Sinal recuperado a 8Kbps e 8bits a partir dos sinais modulantes sem restrições de largura de banda e quantização dos sinais AMPs e FIs:



borboleta_h.wav

8º Sinal recuperado a partir dos sinais modulantes com sinais AMPs filtrados a 100 Hz com quantização a 4 bits e FIs a 400 Hz com 5 bits.



borboleta_g.wav