

INSTITUTO MILITAR DE ENGENHARIA

LEVANTAMENTO DE CARACTERÍSTICAS DA  
LÍNGUA PORTUGUESA UTILIZÁVEIS NO  
PROCESSAMENTO DIGITAL DOS SINAIS  
DE VOZ

POR

ROBERTO MISCOW FILHO  
TENENTE CORONEL

TESE SUBMETIDA COMO  
REQUISITO PARCIAL PARA  
A OBTENÇÃO DO GRAU DE  
MESTRE EM CIÊNCIAS

ASSINATURA DO ORIENTADOR DA TESE

RIO, RIO DE JANEIRO  
DEZEMBRO, 1977

À memória de meu pai - que  
despertou em mim o humano  
e natural desejo de  
conhecer.

## AGRADECIMENTOS

À Seção de Eletricidade do Instituto Militar de Engenharia e aos professores que me fizeram engenheiro e me ensinaram durante o Curso de Pós-Graduação.

Ao Professor Dr. José Paulo de Almeida e Albuquerque, aos Majores Luiz Wilson Marques Daudt e Luiz Costa da Silva - pela orientação e pelo estímulo.

Ao Instituto Nacional de Tecnologia, na pessoa do engenheiro Aimone Camardella, pelo total apóio. Ao Capitão Álvaro Simões - pelo inestimável auxílio, prestado com prejuízo de suas horas de estudo e lazer. Aos Capitães Carlos Alberto Reinert de Lima e Eduardo Augusto Orosco Galvão - pelas utilíssimas sugestões.

Ao Instituto Nacional de Pesquisas Espaciais, nas pessoas dos engenheiros Francisco Viola, José Roberto de Oliveira e Arry Burs - pelo carinho e dedicação com que me ajudaram.

Ao Serviço Cine Fotográfico do Exército, na pessoa do Capitão Antonio Carlos de Assis Brasil - pela valiosa cooperação.

Ao Núcleo de Processamento de Dados do Centro

Brasileiro de Pesquisas Físicas nas pessoas do professor Antonio Julio Lossio Botelho e sua equipe - pela paciência e boa vontade durante longas horas de processamento digital.

À Seção de Engenharia Mecânica do Instituto Militar de Engenharia - na pessoa do Major Edival Ponciano, pelo apôio amigo.

À minha filha mais velha, pela ajuda no trabalho da gravação dos sinais de voz.

Finalmente, mas não menos importante, à minha esposa pelo permanente incentivo e por ter achado tempo, entre os afazeres domésticos, para ajudar no preparo das figuras incluídas neste compêndio.



## SUMÁRIO

O presente trabalho visa testar a viabilidade do levantamento de características da voz em língua portuguesa através do emprego de processamento de sinais reais em computador digital.

Inicialmente, são gravados (em voz feminina e voz masculina) fonemas isolados, combinações simples de fonemas e frases triviais possíveis de ocorrer em conversas telefônicas comuns.

Em seguida, opera-se a conversão análoga/digital dos sinais gravados.

Finalmente, empregando-se algoritmos simples, programados em FORTRAN acessível, são analisadas diversas características dos sinais.

Os resultados obtidos podem servir de subsídios às pesquisas e projetos de sistemas de processamento do sinal de voz.

## ABSTRACT

The objective of this work is to verify the viability of the determination of voice characteristics, in portuguese, processing real signals in a digital computer.

Initially, individual phonems, simple combinations and usual sentences, that normally appear in telephonic conversation, are recorded (using male and female voices).

An analog-to-digital conversion is followed. Finally, several characteristics of the signals are analised employing some simple algorithms, in FORTRAN language.

The results obtained can be applied to Research and development of voice signal processing systems.

## ÍNDICE

SUMÁRIO	iv
LISTA DE ILUSTRAÇÕES	vii
I - INTRODUÇÃO	1
II - CARACTERÍSTICAS E PARÂMETROS DA VOZ	4
III - MÉTODO DE TRABALHO	12
IV - RESULTADOS OBTIDOS	25
V - CONCLUSÕES	87
REFERÊNCIAS	92
PROGRAMA "FFT"	95
PROGRAMA "SPWCH"	96
PROGRAMA "ALISA"	98
PROGRAMA "FRMNT"	100
PROGRAMA "ZEROS"	104
PROGRAMA "VARVA"	105
PROGRAMA "VARCO"	106
PROGRAMA "FQREL"	107

## LISTA DE ILUSTRAÇÕES

Figura 1 - Diagrama esquemático do aparelho fonador (segundo Flanagan <sup>18</sup> )	6
Figura 2 - Circuito da estrutura de acoplamento	15
Figura 3 - Forma de onda de um /a/, v.fem.	19
Figura 4 - Forma de onda de um /i/, v.fem.	20
Figura 5 - Forma de onda de um /ê/, v.fem.	21
Figura 6 - Forma de onda de um /a/, v.masc.	22
Figura 7 - Forma de onda de um /i/, v.masc.	23
Figura 8 - Forma de onda de um /ô/, v.masc.	24
Figura 9 - Esp. (SPWCH) - /a/, v.fem.	29
Figura 10 - Esp. (SPWCH) - /i/, v.fem.	30
Figura 11 - Esp. (SPWCH) - /ô/, v.fem.	31
Figura 12 - Esp. (SPWCH) - /an/, v.fem.	32
Figura 13 - Esp. (SPWCH) - /a/, v.masc.	33
Figura 14 - Esp. (SPWCH) - /ê/, v.masc.	34
Figura 15 - Esp. (SPWCH) - /i/, v.masc.	35
Figura 16 - Esp. (SPWCH) - /u/, v.masc.	36
Figura 17 - Esp. (SPWCH) - /an/, v.masc.	37
Figura 18 - Esp. (SPWCH) - /pa/, v.fem.	38
Figura 19 - Esp. (SPWCH) - /ba/, v.fem.	39
Figura 20 - Esp. (SPWCH) - /ta/, v.masc.	40
Figura 21 - Esp. (SPWCH) - /zã/, v.masc.	41
Figura 22 - Esp. (ALISA) - /a/, v.fem.	44
Figura 23 - Esp. (ALISA) - /i/, v.fem.	45

Figura 24 - Esp. (ALISA) - /ê/, v.fem.	46
Figura 25 - Esp. (ALISA) - /an/, v.fem.	47
Figura 26 - Esp. (ALISA) - /i/, v.masc.	48
Figura 27 - Esp. (ALISA) - /u/, v.masc.	49
Figura 28 - Env. espectral - /a/, v.fem.	56
Figura 29 - Env. espectral - /ê/, v.fem.	57
Figura 30 - Env. espectral - /i/, v.fem.	58
Figura 31 - Env. espectral - /ô/, v.fem.	59
Figura 32 - Env. espectral - /u/, v.masc.	60
Figura 33 - Env. espectral - /an/, v.fem.	61
Figura 34 - Env. espectral - /a/, v.masc.	62
Figura 35 - Env. espectral - /i/, v.masc.	63
Figura 36 - Esp. var. formantes (baixos-v.fem.)	66
Figura 37 - Esp. var. formantes (médios-v.fem.)	67
Figura 38 - Esp. var. formantes (altos-v.fem.)	68
Figura 39 - Esp. var. formantes (baixos-v.masc.)	69
Figura 40 - Esp. var. formantes (médios-v.masc.)	70
Figura 41 - Esp. var. formantes (altos-v.masc.)	71
Figura 42 - Var. variância ("O João quebrou o dedo" - v.fem.)	73
Figura 43 - Var. variância ("O João quebrou o dedo" - v.masc.)	74
Figura 44 - Var. variância ("A Regina <del>ca</del> a com o Raul" - v.fem.)	75
Figura 45 - Var. variância ("A Regina casa com o Raul" - v.masc.)	76

Figura 46 - Var. variância ("Das Herz München")	77
Figura 47 - Freq. relativa ("A Regina casa com o Raul" - v.fem. - três segmentos iniciais)	80
Figura 48 - Freq. relativa ("O João quebrou o dedo" - v.masc. - três segmentos iniciais)	81
Figura 49 - Var. correlação, ret. = 1 ("O João quebrou o dedo" - v.fem.)	85
Figura 50 - Var. correlação, ret. = 1 ("O João quebrou o dedo" - v.masc.)	86

## I - INTRODUÇÃO

1.1 - Em 1779, Christian Gottlieb Kratzenstein ganhou prêmio oferecido pela Academia Imperial de São Petersburgo construindo ressoadores capazes de imitar "com tolerável precisão" os sons das vogais a, e, i, o, u<sup>5</sup>. Aceitando aquela data como referência, pode-se conceder à pesquisa da voz sintética uma existência de quase duzentos anos.

De lá para nossos dias, muito progresso foi feito no processamento da voz humana, incluindo as "máquinas falantes" de von Kempelen (1791) e de Wheatstone (1875), passando pela genial invenção de Graham Bell (1875) e chegando, em 1939, ao Vocoder, construído por Homer Dudley dos laboratórios Bell; o Vocoder marca, sem dúvida, a definitiva presença da eletrônica na síntese da voz.

1.2 - Entretanto, não apenas a pesquisa da voz sintética vem atraindo os engenheiros e cientistas empenhados no estudo da voz humana; nas últimas décadas, o aumento explosivo da demanda telefônica, acompanhado pelo crescente encarecimento do cobre, (matéria prima essencial na indústria de cabos), tem levado os pesquisadores à procura de novos processos de modulação capazes de aliar as vantagens da transmissão digital a uma utilização efi-

ciente da faixa disponível. É exatamente o que se observa com a evolução dos sistemas PCM para sistemas DPCM, delta e as técnicas de quantização adaptiva<sup>8,9</sup>.

1.3 - Qualquer que seja o alvo das pesquisas, elas se realizam com larga aplicação do processamento digital, fazendo uso intenso dos circuitos integrados, dos microprocessadores e dos computadores digitais.

1.4 - Eis algumas das possíveis aplicações do processamento digital dos sinais de voz:

- (1) transmissão telefônica;
- (2) síntese da voz;
- (3) armazenagem em computador e "fala" do computador em diálogo direto homem ↔ máquina;
- (4) identificação e reconhecimento de pessoas;
- (5) auxílio às pessoas portadoras de defeitos relacionados com a elocução (p. ex.: defeitos da fala ou da audição causados por doenças ou acidentes).

1.5 - Compulsando a abundante literatura técnica alienígena disponível, verifica-se existirem, em outros países, estudos exaustivos sobre a fonética das línguas respectivas e sobre o papel daquela disciplina no processamento dos sinais de voz. Constata-se, também, através da leitura dos artigos a íntima colaboração de engenheiros, físicos e foneticistas no ramo da acústica aplicada<sup>7</sup>.



1.6 - O presente trabalho não pretendeu introduzir novas técnicas e diferentes processos em área sobejamente explorada e conhecida por outros países. Desejou-se, aplicar sabidos processos e técnicas no levantamento de parâmetros da voz em língua portuguesa, visando obter subsídios para pesquisadores e engenheiros brasileiros interessados diretamente nas aplicações retro mencionadas.

## II - CARACTERÍSTICAS E PARÂMETROS DA VOZ

2.1 - Um estudo judicioso e preciso do processamento dos sinais de voz depende, naturalmente, do estudo preliminar do objeto dessa operação, isso é, da voz humana, suas características físicas e seus parâmetros, particularmente os parâmetros utilizáveis no processamento.

2.2 - Pode-se definir a voz humana como o " processo de converter mensagem intencional formulada com unidades linguísticas discretas em sinal acústico contínuo capaz de levar a mensagem ao ouvinte"<sup>3</sup>. Chama-se fonema a unidade sonora que não se pode analisar em unidades menores e que permite diferenciar uma palavra da outra, por exemplo: mas de mês , faço de maço, correr de correu etc.

2.3 - O aparelho fonador é o conjunto de órgãos capazes de criar as condições para a produção da voz humana, compreendendo:

- (1) pulmões, brônquios e traquéia - (fornecem a corrente de ar, essencial para a fonação);
- (2) laringe - (onde se localizam as cordas vocais, responsáveis pelo caráter quase periódico dos sons vocálicos e da maior parte dos sons consonantais);
- (3) cavidades supralaríngeas (faringe, boca e fossas na-

sais - atuam como ressoadores<sup>2,6</sup> (ver Figura 1).

2.4 - O fluxo de ar proveniente dos pulmões pode ser utilizado de três modos na produção dos sons da voz:

- (1) "modulação" pelas cordas vocais - as cordas (na realidade, ligamentos carnosos esticados entre as cartilagens aritnóideas e a cartilagem tireóidea) são postas em vibração pelo impulso nervoso oriundo do córtex cerebral. O ar, forçado através da glote, acelera-se na estreita passagem e, em seguida, retarda-se na cavidade maior da faringe. O efeito Bernouilli reduz a pressão na glote fazendo as cordas se aproximarem mais ainda para interromper o fluxo; por outro lado, a elasticidade das cordas provoca seu afastamento, aumentando a abertura da glote. Esse processo se repete a uma taxa variável entre 50 e 400 vezes por segundo, independentemente do movimento dos órgãos da articulação (mandíbula, língua, lábios e véu palatino). Os sons originados das cordas vocais são chamados VOZEADOS (designação sugerida pelo professor Gladstone Chaves de Melo para substituir o termo "sonoros"), incluindo as vogais e muitas consoantes tais como: g, j, m, n, r etc;
- (2) obstrução parcial no trato vocal - é processo tipicamente aleatório e aparece quando alguma constrição se interpõe no percurso da corrente expiratória oriunda da glote. Os sons desse grupo, geralmente, são

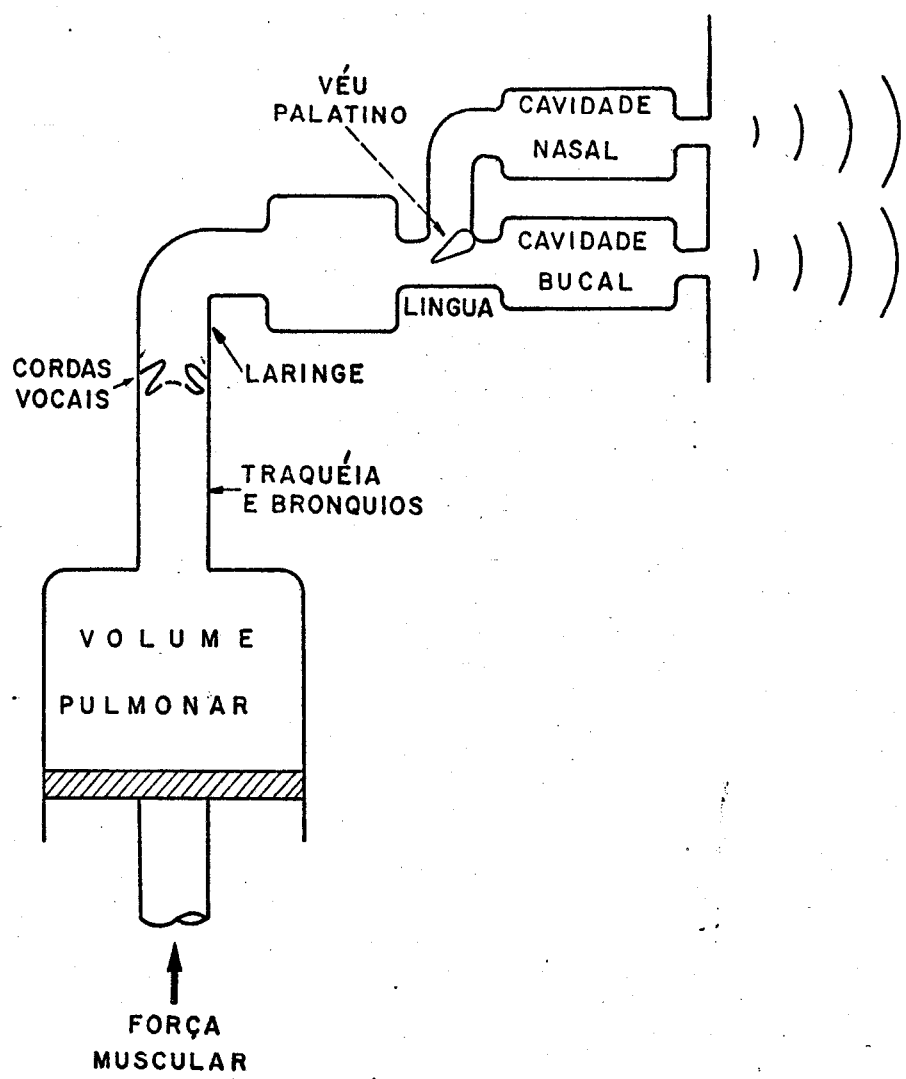


Figura 1 - Diagrama esquemático do aparelho fonador (segundo Flanagan<sup>18</sup>).

caracterizados por uma sibilância, por exemplo: *f* ou *s*. Se houver paralelamente, a atividade das cordas vocais, teremos os sons FRICATIVOS VOZEADOS: *z* e *v*, por exemplo;

- (3) liberação repentina após fechamento do trato vocal - uma queda súbita da pressão gera excitação transitória do trato produzindo um som "explosivo". Se as cordas vocais estiverem vibrando durante o fechamento, isso é, durante o aumento da pressão, surgem os sons EXPLOSIVOS VOZEADOS, tipo *b*, *d*. Se as cordas estiverem em repouso, aparecem as consoantes EXPLOSIVAS NÃO VOZEADAS: *p*, *t*, *k*.

2.5 - Espectro da voz - O espectro da voz estende-se, com níveis de energia significativos até 5 kHz. No presente trabalho utilizou-se a faixa passante de 4 kHz, largura essa compatível com o canal de voz recomendado pela EMBRATEL<sup>22</sup>; em aplicações onde se faça necessária a análise de frequências no limite da faixa natural (5 kHz), o método de trabalho poderá ser o mesmo alternando-se apenas os valores de alguns parâmetros referidos no Capítulo III.

2.5.1- Sons oriundos das cordas vocais - sendo normalmente periódicos (ou quase periódicos), sua energia se divide entre a frequência fundamental de vibração e um grande número de harmônicos até 5 kHz. A análise es-

pectral revela estar na fundamental a maior amplitude, seguindo-se uma queda progressiva à taxa de 12 dB/oitava<sup>3</sup>

2.5.2- Os sons fricativos, de natureza turbulenta, apresentam espectro substancialmente plano, tipo "ruído branco", isso é, distribuição contínua de energia.

2.5.3- Os sons explosivos têm espectro semelhante aos dos fricativos.

2.5.4- O efeito do trato vocal - A posição da língua, (órgão cuja importância na fonação transformou o termo em sinônimo de idioma<sup>2</sup>), dos lábios, da mandíbula e do véu palatino altera as ressonâncias e as antiressonâncias do trato vocal fazendo a energia concentrar-se em determinadas frequências conhecidas como FORMANTES da voz. Geralmente, ocorrem cerca de cinco formantes na faixa até 5 kHz.

2.5.5- Na produção de sons nasais, o véu palatino é abaixado permitindo fluxo de ar pelo nariz; em consequência, são produzidos novos formantes (polos da transmissão) e novas anti-ressonâncias (zeros da transmissão)<sup>1</sup>.

2.5.6- Qualquer vogal pode ser nasalizada, dependendo da menor elevação do véu palatino; "o inglês falado nos EE.UU. é caracterizado pela ressonância nasal do seu

vocalismo"<sup>1</sup> (fato esse aproveitado por humoristas do rádio e da televisão ao fazerem imitação de americano). Normalmente, as vogais do português têm uma relativa nasalização cuja intensidade cresce segundo a sequência: *i* → *u* → *e* → *o* → *a*. O a é a vogal cuja emissão é acompanhada de um fluxo de ar pelo nariz porquanto, sendo a vogal mais aberta, o véu palatino se encontra mais descido durante sua articulação<sup>1</sup>. Esse fato é comprovado pelo maior número de formantes no espectro do a<sup>1</sup>.

2.5.7- Timbre e altura das vogais - O timbre da vogal é caracterizado pelos FORMANTES; a altura ("pitch") está relacionada com a inflexão, p. ex. nas duas frases:

"Ele está aqui."

"Ele está aqui?", a palavra aqui é a mesma, entretanto a mudança de altura no i caracteriza a inflexão interrogativa.<sup>4</sup>

2.6 - Intensidade da voz - Geralmente, a intensidade da corrente de ar expirado é mais forte para u e i que para a; as consoantes NÃO vozeadas ("surdas") exigem muito mais ar que as VOZEADAS ("sonoras").<sup>1</sup>

2.6.1- Os sinais de voz não escapam à lei geral da natureza: economizar energia, procurar o menor esforço. Assim, verificações experimentais revelam estarem as amplitudes do sinal de voz abaixo de um quarto do valor

RMS, em 50% do tempo<sup>16</sup>.

2.6.2- Essa característica do sinal de voz explica o motivo do uso da quantização não uniforme (isso é, precedida pelo compressor), nos sistemas PCM.

2.7 - Alguns elementos de fonética - Basicamente, os sons da voz podem ser divididos em dois grupos: vogais e consoantes. As consoantes do português falado no Brasil compreendem 19 sons característicos grupados no quadro a baixo<sup>2</sup>:

	CONSOANTES EXPLOSIVAS	ORAIS FRICATIVAS	CONSOAN- TES NASAIS
VOZE-	/b/: <u>b</u> ule	/v/: <u>v</u> aca; /z/: <u>z</u> ero;	/m/: <u>m</u> ola
	/d/: <u>d</u> ama	(ʒ): <u>j</u> ato; /r/: <u>m</u> oro;	/n/: <u>n</u> ada
ADAS	/g/: <u>g</u> ato	/ʁ/: <u>m</u> orro	/ɲ/: <u>u</u> na
		/l/: <u>l</u> ata; /ʎ/: <u>b</u> ol <u>h</u> a	
NÃO	/p/: <u>p</u> ato	/f/: <u>f</u> aca	
VOZE-	/t/: <u>t</u> ora	/s/: <u>s</u> ede	
ADAS	/k/: <u>c</u> alo	/ʃ/: <u>ch</u> ã	

2.6.1- As vogais faladas no Brasil podem ser classificadas em sete tipos<sup>2</sup>:

/a/: pa; /ē/: fē; /ē/: vē; /i/: si; /ō/: bola;  
/ô/: ovo; /u/: nu; podendo-se acrescentar as semivo-



gais /y/: pai; /w/: água.

2.6.2- Conforme a zona de articulação, as vogais podem ser classificadas em<sup>2</sup>:

(a) anteriores: i, ē, ê;

(b) posteriores: ō, ô, u.

### III - MÉTODO DE TRABALHO

3.1 - *Introdução* - O método escolhido para obtenção das características desejadas pode ser esquematizado na seguinte sequência:

*Gravação do Sinal  
de Voz em Fita  
Magnética*

*Conversão do Sinal  
Análogo para Discreto, Digitalizado*

*Processamento em  
Computador  
Digital*

*Análise e Interpreta  
ção dos  
Resultados*

3.2 - *Inicialmente, foi realizada a gravação do sinal em fita, utilizando-se gravador UHER-1200-REPORT SYNCHRO, especial para gravar trilha sonora em películas ci*

nematográficas, cedido pelo Serviço Cine Fotográfico do Exército.

3.2.1- A gravação foi feita em câmara anecôica cedida pelo Instituto Nacional de Tecnologia.

3.2.2- Foram gravadas duas vozes, masculina (adulto de 45 anos) e feminina (jovem de 21 anos) e o texto consistiu em 37 "segmentos" abrangendo:

- (a) sons das vogais orais e nasais;
- (b) sons de ditongos;
- (c) sons dos fonemas consonantais;
- (d) fala dos dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8 e 9 (como são referidos na telefonia);
- (e) frases simples, com palavras mais frequentes na conversação comum, possíveis de serem ouvidas em diálogos telefônicos, p. ex.:

- "A Regina casa com o Raul sábado que vem".

- "Alô, papai! Quem está falando aqui é o Pedro (Maria)".

- "O Júlio quebrou o dedo jogando vôlei".

- "Sô posso viajar na semana depois da Páscoa",  
e outras do mesmo tipo.

3.2.3- Cada segmento foi articulado em intervalo de tempo de, aproximadamente, 3 segundos, sempre sobrando margem antes de o ponteiro completar o intervalo: procu-

rou-se, nas frases típicas, dar um tom coloquial, de uma conversa telefônica comum em família ou entre amigos.

3.2.4- Entre um segmento e o adjacente foi deixado intervalo de silêncio igual ou superior a 3 segundos, tempo esse exigido pelo conversor análogo-digital utilizado. Assim, a fita magnética do gravador ficou com a sequência de informações (cada uma durando 3 s), separadas por silêncios (> 3 s).

3.3 - Estrutura de acoplamento - Foi concebida estrutura intermediária entre o gravador de fita e o conversor análogo/digital (ver Figura 2) com as seguintes finalidades:

- (1) filtrar o sinal na saída do gravador, de modo a limitar a faixa em 4 kHz;
- (2) comprimir o sinal filtrado de modo a garantir que sinais de baixa amplitude no início da frase possam disparar o gerador de pulso retangular e evitar que níveis de maior amplitude ultrapassem os limites aceitos pelo conversor análogo/digital (-1,0 V a +1,0 V);
- (3) gerar um pulso retangular negativo, de amplitude >4V e largura igual a 3 segundos, responsável pelo comando do conversor análogo/digital. (O projeto eletrônico dessa estrutura, segundo essas especificações mínimas, foi realizado, detalhadamente, pelo Cap. Engenheiro de Comunicações Álvaro Simões, professor

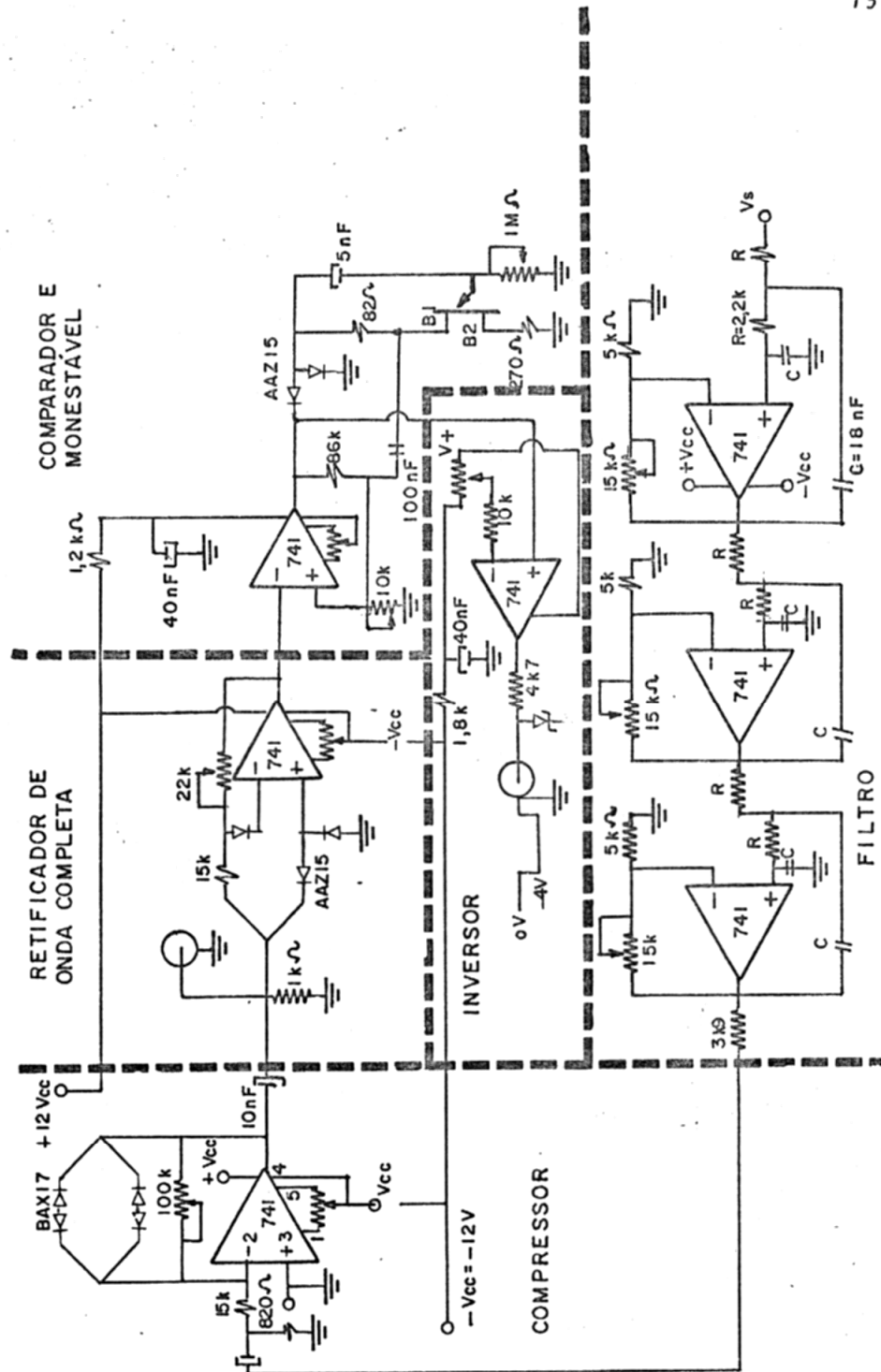


Figura 2 - Circuito da estrutura de acoplamento.

do I M E).

3.4 - Conversão análoga/digital - Foi realizada no minicomputador HEWLETT-PACKARD modelo 2116-B, instalado no laboratório BETA do INPE (Instituto Nacional de Pesquisas Espaciais - S. José dos Campos - SP), sendo o programa da conversão (escrito em ASSEMBLER) elaborado especialmente pelo Eng<sup>o</sup>. Francisco Viola, responsável pelo computador.

3.4.1- A conversão utilizou a frequência de amostragem de 11000 Hz (11kHz), superior, portanto, à taxa de Nyquist compatível com o sinal (8 kHz). O sinal amostrado foi quantizado em 256 níveis e codificado em código binário de 8 bits (o sinal + ou - dado pelo bit mais à esquerda).

3.4.2- O sinal convertido foi gravado em fita de computador, à taxa de 800 bytes/polegada.

3.4.3- Após a conversão análoga/digital, e utilizando-se conversor digital/análogo do próprio computador, foi injetado o sinal reconstruído em amplificador de áudio constatando-se, de modo prático e direto, a precisão e fidelidade da conversão análoga/digital.

3.5 - A fita magnética contendo os segmentos codifi-

dados foi processada no computador IBM/370 - Mod. 145 da Gerência de Processamento de Dados do Centro Brasileiro de Pesquisas Físicas - CNPq (Rio de Janeiro - RJ).

3.5.1- Para obtenção dos parâmetros desejados foram utilizados os seguintes programas e subrotinas:

- GRAFG - subrotina feita pelo Cap. Engº Com. M.Sc. Carlos Alberto Reinert de Lima; permite o traçado de gráficos através da impressora, além de tabular os valores das coordenadas.
- SIMQ - subrotina convencional do "package" científico da IBM - resolve sistema de equações simultâneas.
- POLRT - (ib.) calcula as raízes de um polinômio.
- FFT - (Fast Fourier Transform) - algoritmo citado por Rabiner & Gold<sup>18</sup>.
- AUTOQ - subrotina para cálculo de correlação.
- SPWCH - programa para traçado do espectro estimado; baseado no método descrito por Welch<sup>10</sup>.
- ALISA - programa para traçado do espectro estimado; baseado em simples operação de "smoothing" citada por Schwartz & Shaw<sup>13</sup>.
- FRMNT - programa para obtenção dos formantes; baseado no método descrito por Itakura & Shuzo<sup>12</sup>.
- FQREL - programa para obter o gráfico das frequências relativas das amplitudes do sinal de voz.
- VARVA - programa para obter o gráfico da variação da

variância de segmentos curtos (13,6 ms) comparada com a variância geral de uma frase.

- ZEROS - programa para obter o gráfico da ocorrência de cruzamentos do nível zero.
- VARCO - programa para obter o gráfico da variação da correlação de retardo unitário normalizada.
- EXAME - programa para traçar o desenho do sinal de voz amostrado.
- BASE2 - subrotina que lê o sinal da fita magnética, converte o código de 8 bits para o código do computador, expande e filtra o sinal.

3.5.2- As listagens dos programas FFT, SPWCH, ALISA, FRMNT, VARVA, ZEROS, VARCO e FQREL estão em anexo neste compêndio.

3.5.3- As figuras de nº 3 a 8 representam algumas das formas de onda dos fonemas vocálicos analisados.





Figura 3 - Forma de onda de um /a/, v.fem.

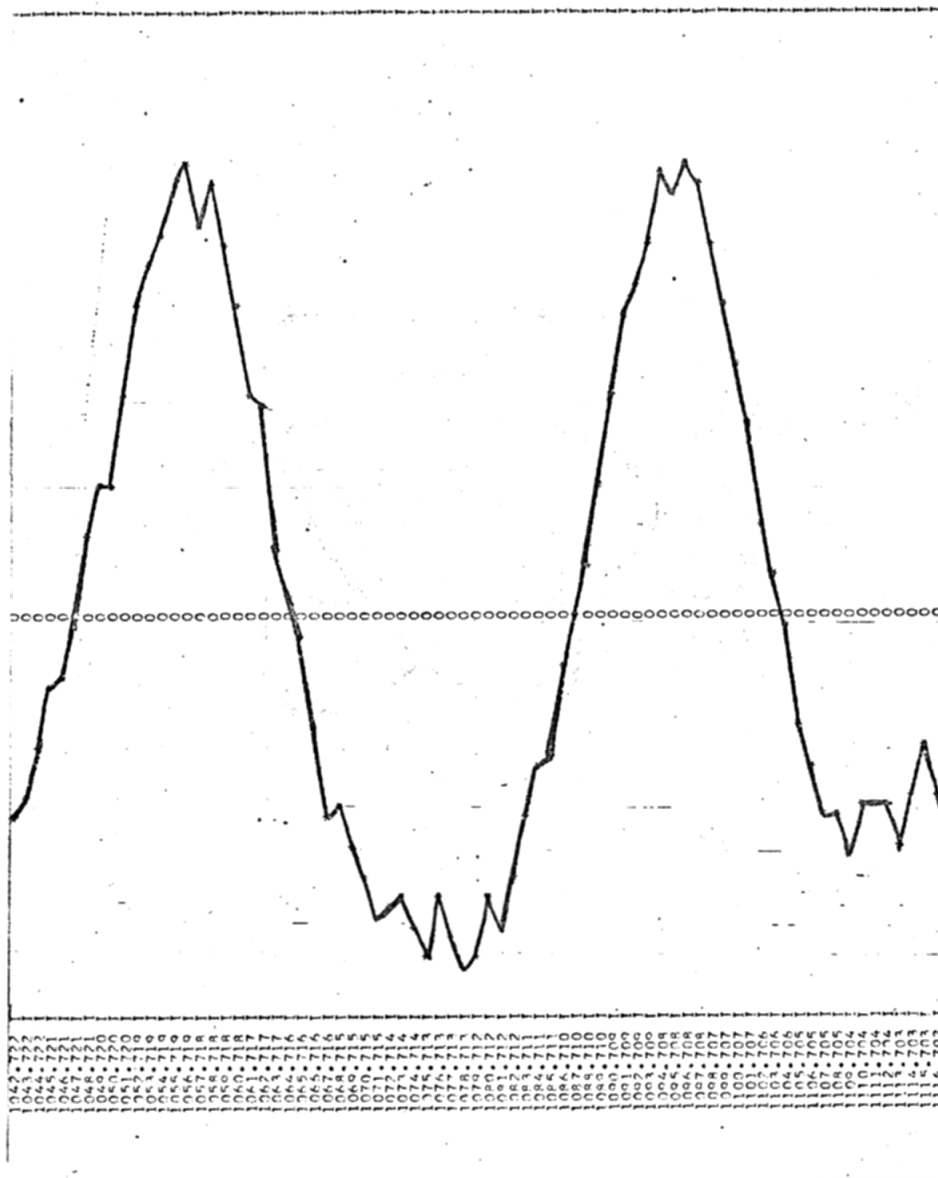


Figura 4 - Forma de onda de um /i/, v.fem.

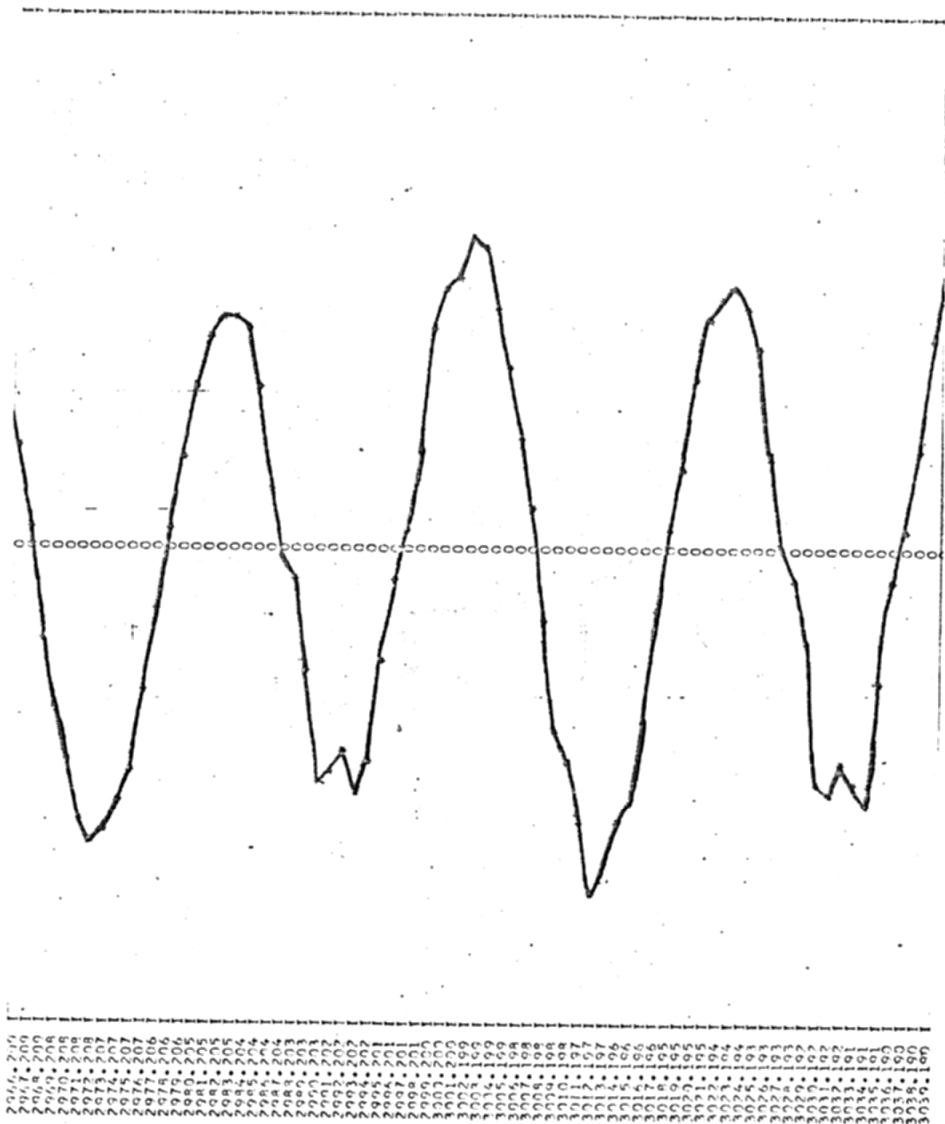


Figura 5 - Forma de onda de um  $|\hat{e}|$ , v.fem.

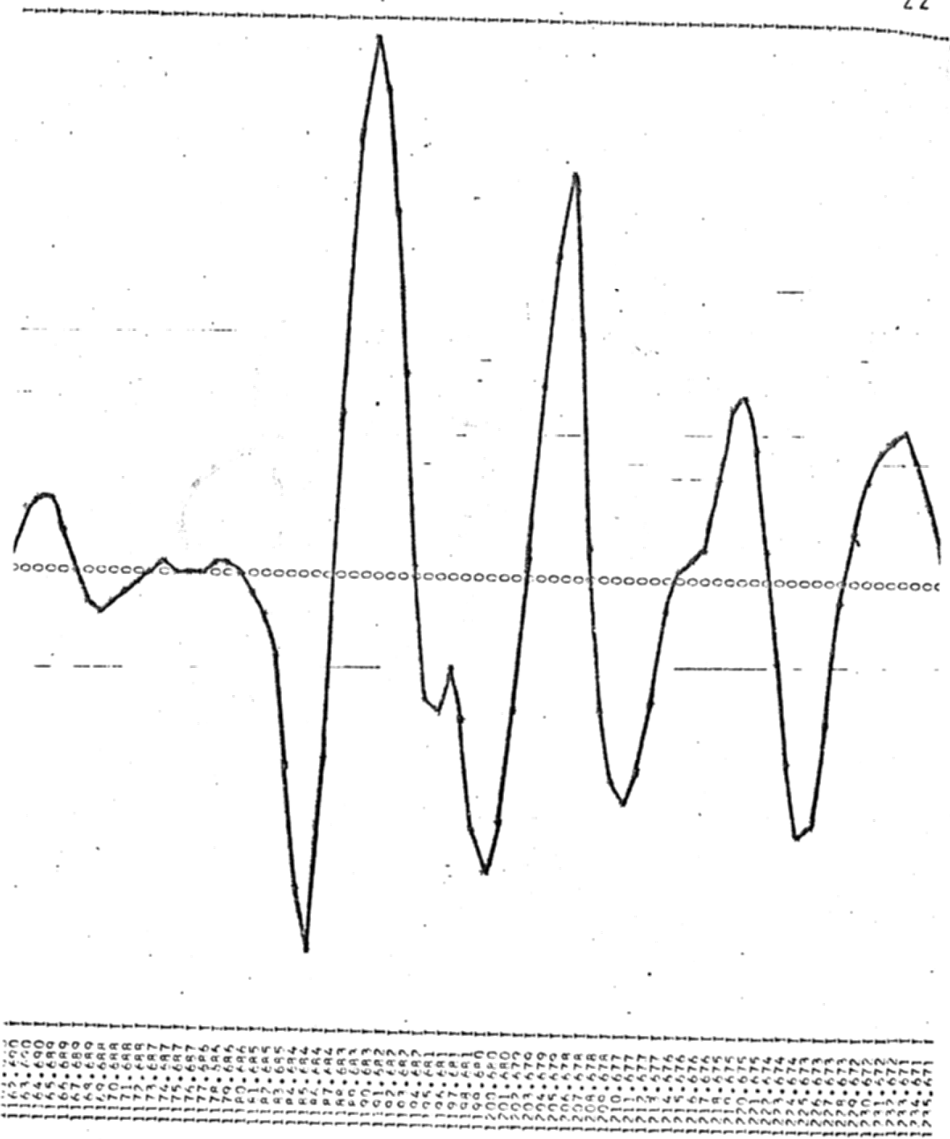


Figura 6 - Forma de onda de um /a/, v.masc.

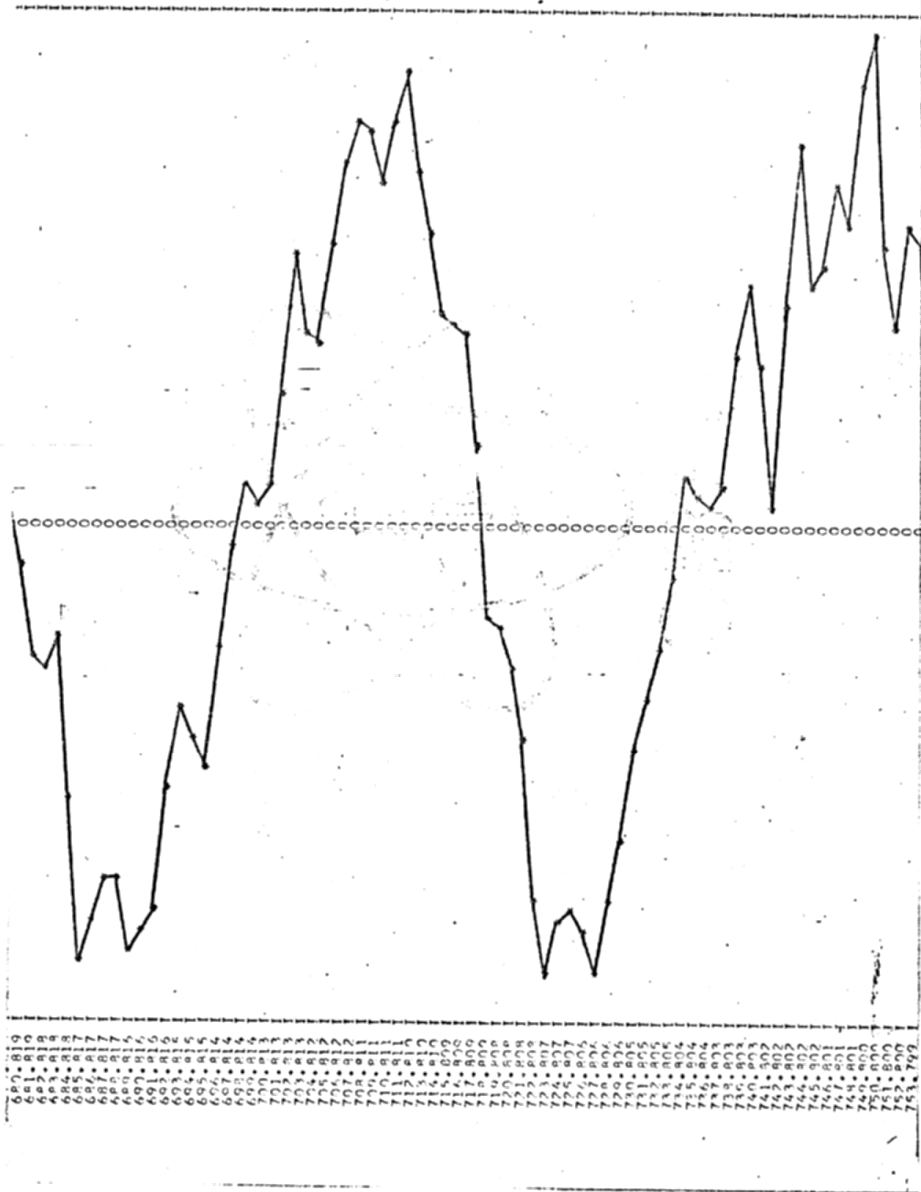


Figura 7 - Forma de onda de um  $\frac{1}{4}$ , v.masc.

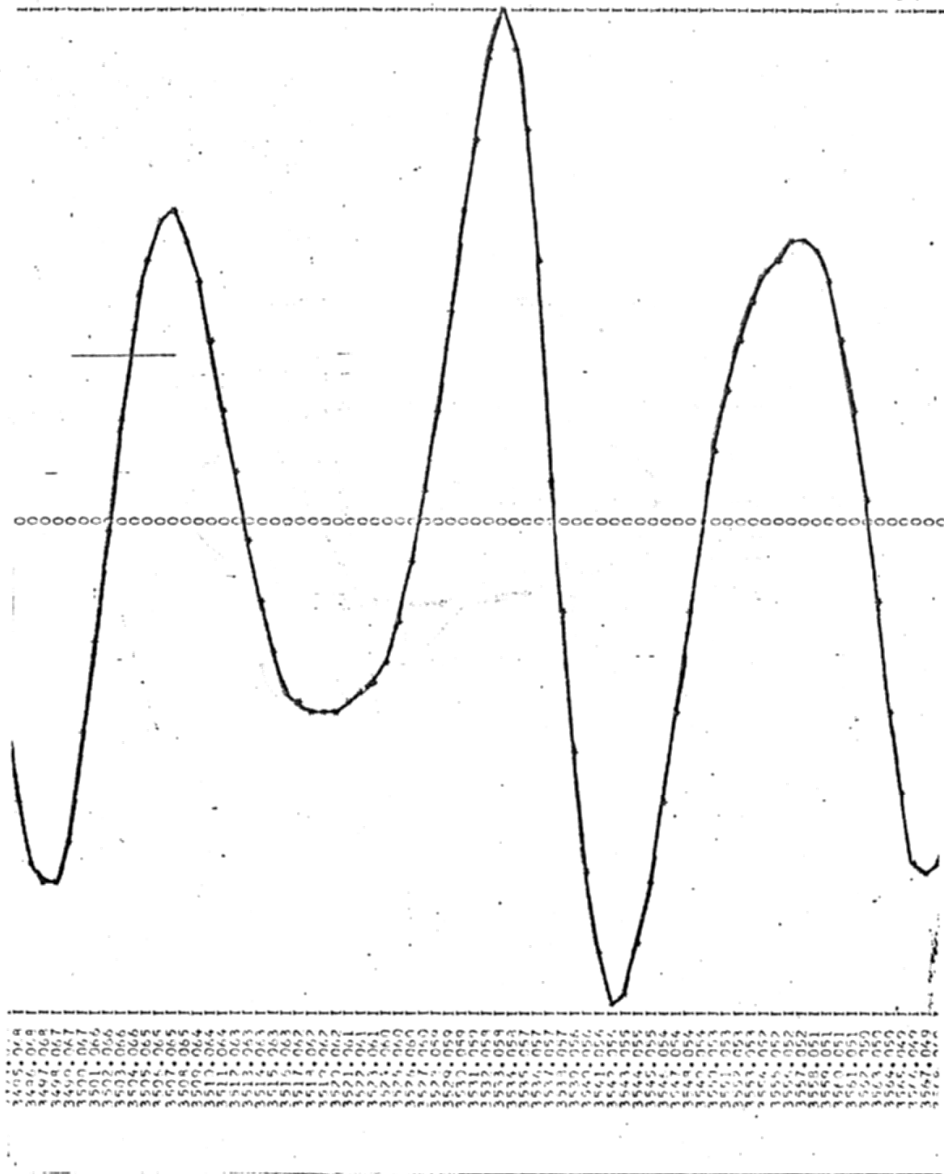


Figura 8 - Forma de onda de um  $\hat{v}_o$ , v.masc.

#### IV - RESULTADOS OBTIDOS

4.1 - Os resultados estão divididos em dois grupos sem que isso signifique a impossibilidade de existirem áreas comuns. Assim, existem os:

- levantamentos feitos no domínio da frequência;
- levantamentos feitos no domínio do tempo.

4.2 - Domínio da frequência - Trabalhou-se basicamente com os programas EXAME, SPWCH, FRMNT e ALISA.

4.2.1- O programa SPWCH utiliza o método descrito por Welch<sup>10</sup> para obter a estimação do espectro de potência.

Considere-se uma amostra  $x(n)$ ,  $n = 0, 1, 2, \dots, (N-1)$ , da sequência estocástica em estudo.

A amostra é dividida em segmentos superpostos de comprimento  $L$ , com os pontos iniciais separados de  $D$  unidades. Seja  $x_1(n)$ ,  $n = 0, 1, 2, \dots, (L-1)$  o primeiro segmento. Assim:

$$x_1(n) = x(n) \quad n = 0, 1, \dots, (L-1) \quad (1)$$

Semelhantemente:

$$x_2(n) = x(n+D) \quad n = 0, 1, \dots, (L-1) \quad (2)$$

etc. até:

$$x_k(n) = x(n + (K-1)D) \quad n = 0, 1, \dots, (L-1) \quad (3)$$

Isso é, tem-se um total de  $K$  segmentos cobrindo inteiramente a amostra (o programa SPWCH usou 7 segmentos, cada um com  $L = 128$ , cobrindo a amostra de 512 pontos).

4.2.1.1- Para cada segmento de comprimento  $L$ , calcula-se a transformada de Fourier:

$$X_i(k) = \frac{1}{L} \sum_{n=0}^{L-1} x_i(n) W(n) \exp(-j \frac{2\pi}{L} nk) \quad (4)$$

onde

$$X_i(k) = X_i(k \frac{2\pi}{L T_s}) \quad i = 1, 2, 3, \dots, K \quad (5)$$

$T_s$  = período de amostragem

$W(n)$  é uma janela para modificar o periodograma. Welch sugere duas janelas, uma do tipo  $1-t^2$  para  $-1 < t < 1$ , e outra do tipo  $1-|t|$ , para  $-1 < t < 1$ . Optou-se pela primeira, com a equação:

$$W(n) = 1 - \left[ \frac{n - \frac{L-1}{2}}{\frac{L+1}{2}} \right]^2, \quad n = 0, 1, 2, \dots, (L-1) \quad (6)$$

4.2.1.2- Welch demonstra que a máxima redução na variância do periodograma é, aproximadamente, atingida quando  $D = L/2$ , obtendo-se, nesse caso:  $\hat{\sigma}_s^2 = \frac{11}{18} \sigma_s^2$  ou seja 61% do valor real. Por essa razão, fez-se no progra-



ma  $D = \frac{L}{2} = 64$  pontos.

4.2.1.3- Obtida a transformada de Fourier,  $X_i(k)$ , são calculados os  $K$  periodogramas modificados

$$I_i(k) = \frac{L}{U} |X_i(k)|^2, \quad i = 1, 2, \dots, K \quad (7)$$

onde

$$U = \frac{1}{L} \sum_{n=0}^{L-1} \omega^2(n) \quad (8)$$

e, finalmente, obtém-se a estimação espectral através da média

$$S(k) = \frac{1}{K} \sum_{i=1}^K I_i(k) \quad (9)$$

4.2.1.4- Aplicou-se o programa SPWCH aos seguintes fonemas:

/a/, /ê/, /i/, /ô/, /u/, /an/, (v. fem. e masc.)

/pa/, /ta/, /sa/, /fa/, /ka/, (v. fem. e masc.)

/ba/, /ra/, /va/, /za/, (v. fem. e masc.)

(alguns espectros estão representados nas Figuras de nº 9 a 21).

4.2.1.5- Os espectros levantados, induzem às seguintes observações:

- nos espectros dos sons femininos, as frequências predominantes são mais altas que as predominantes nos sons masculinos correspondentes;

- o fonema /a/ tem os formantes mais grupados, o que explica as formas de onda mais "limpas", isso é, sem recortes;
- os fonemas /ê/ e /i/ têm os principais formantes afastados o que explica as formas de onda "recortadas", isso é, como se estivessem moduladas por ruído<sup>4</sup>; observem-se as Figuras de nº 4, 5 e 7;
- os fonemas nasalizados (/an/, no caso) apresentam diferentes formantes (polos e zeros) explicados pela corrente de ar passando pelos tratos oral e nasal<sup>1</sup>;
- comparando o espectro de fonema puramente vocálico (/a/, em nosso caso) com os espectros de fonemas consonantais correspondentes, observa-se nesses últimos, o aparecimento de desníveis maiores entre os sucessivos picos e vales da linha espectral, sugerindo um processo de "modulação de amplitude" do espectro do som básico (vogal) pelo espectro do fonema consonantal;
- quase todos os espectros de fonemas consonantais mostram diminuição do valor da frequência de amplitude máxima quando comparado ao valor da frequência análoga no espectro da vogal pura, algo como se a articulação do fonema consonantal alterasse a função de transferência,  $H(s)$ , do trato, p. e.: introduzindo um zero próximo à origem;
- todos os espectros apresentam a tendência geral decrescente, levando-nos a considerar o trato vocal como uma estrutura passa-baixas.





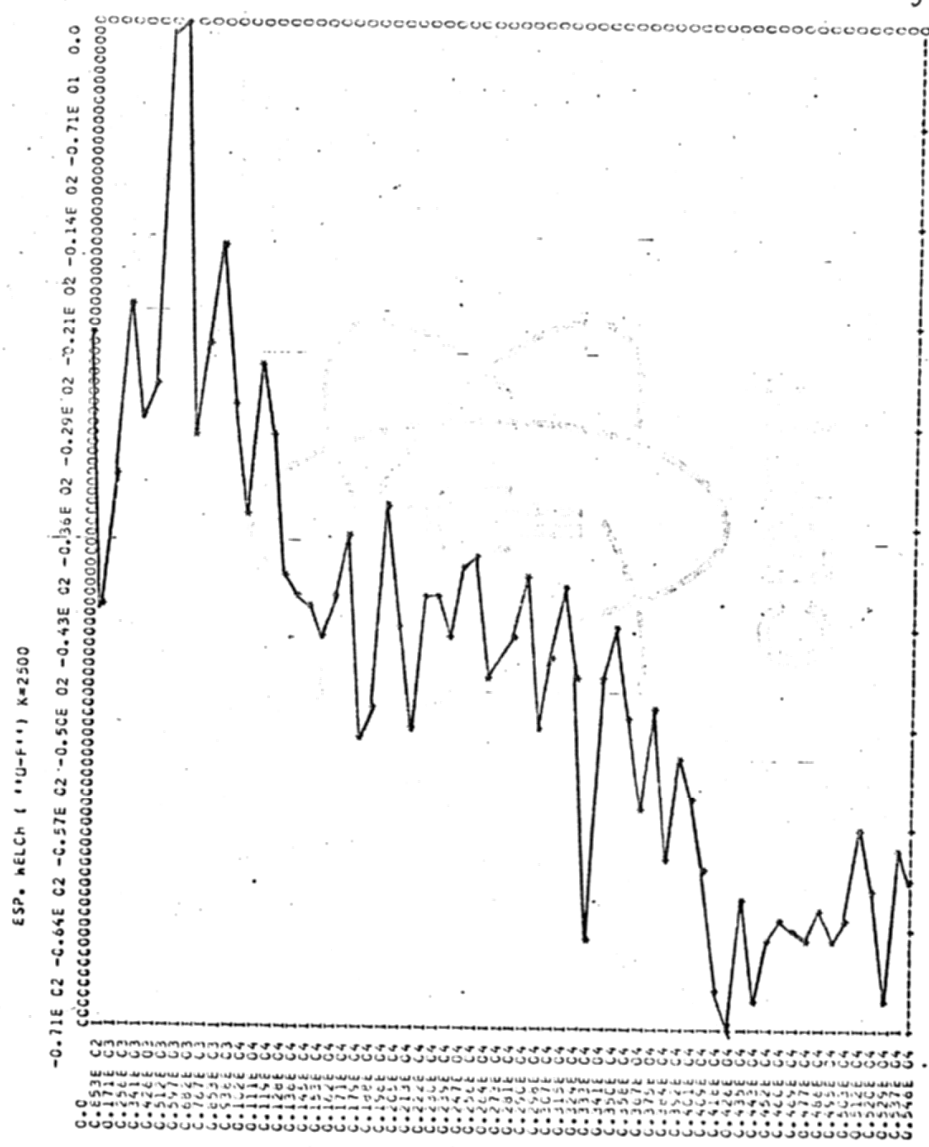


Figura 11 - Esp. (SPWCH) - / $\hat{o}$ /, v.fem.

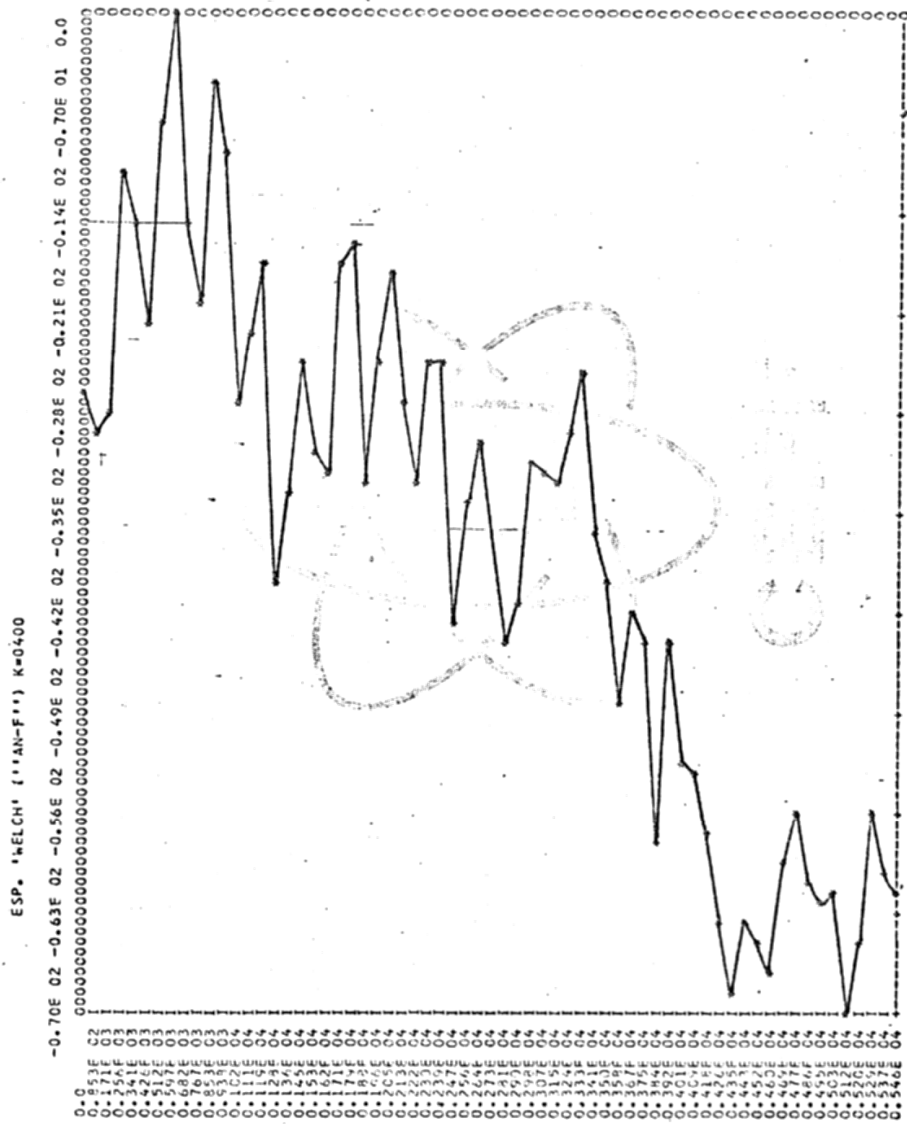


Figura 12 - Esp. (SPWCH) - /an/, v.fem.

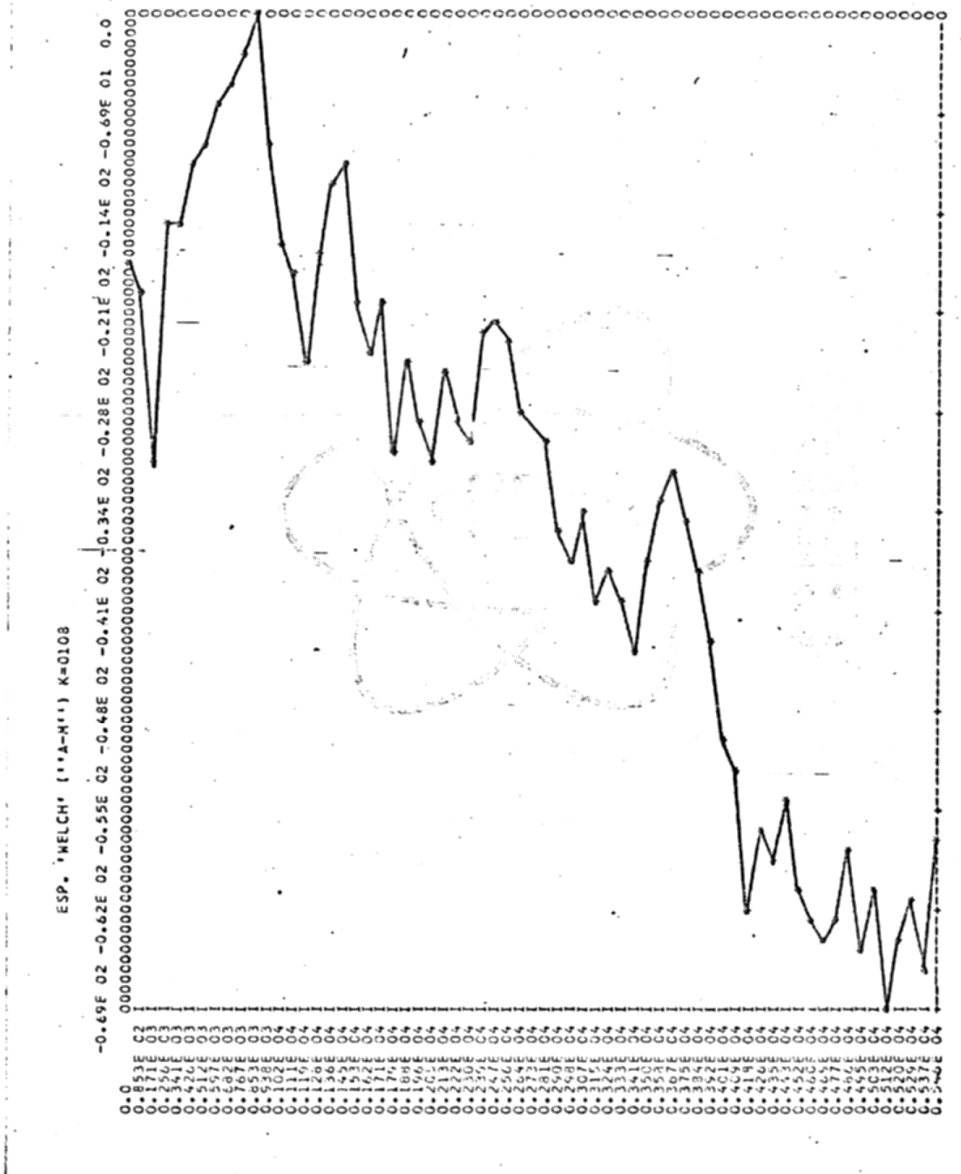


Figura 13 - Esp. (SPWCH) - /a/, v.masc.

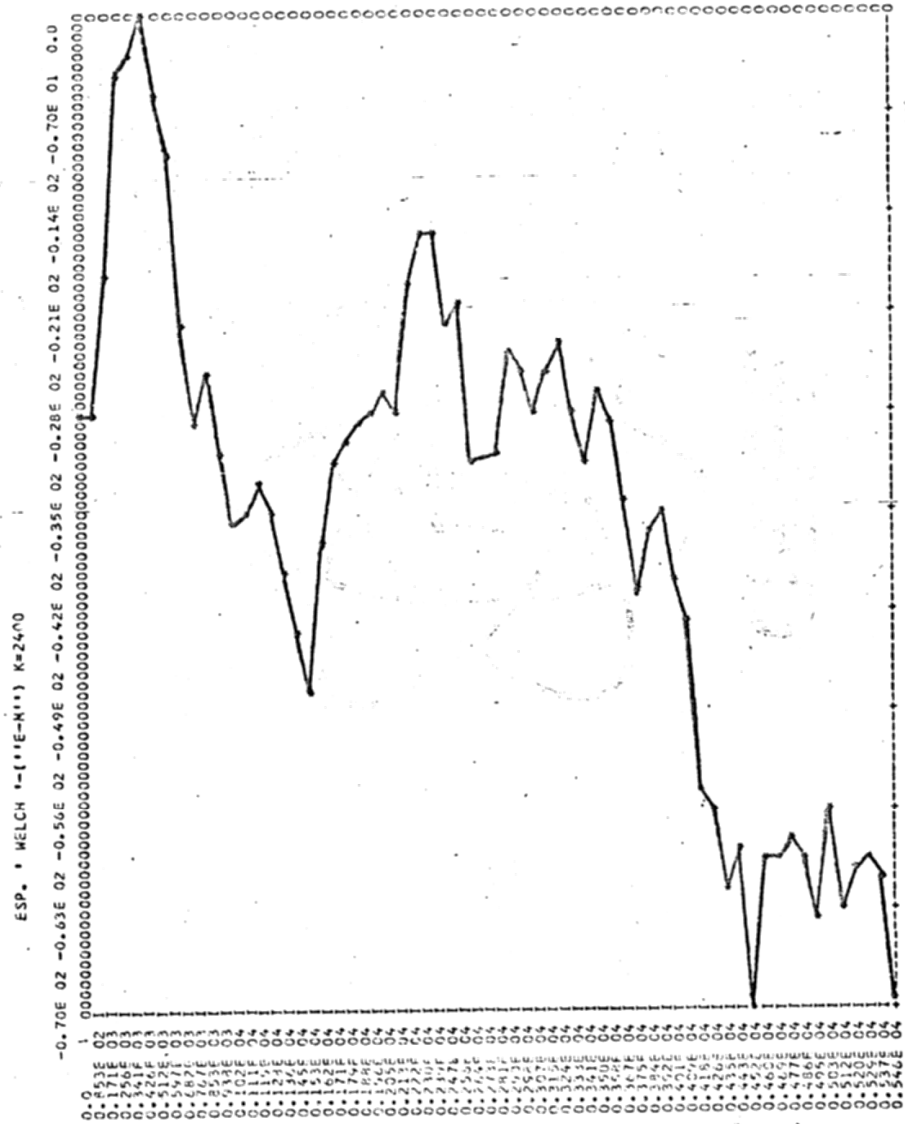


Figura 14 - Esp. (SPWCH) -  $|e|$ , v.masc.





Figura 15 - Esp. (SPWCH) - /i/, v.masc.

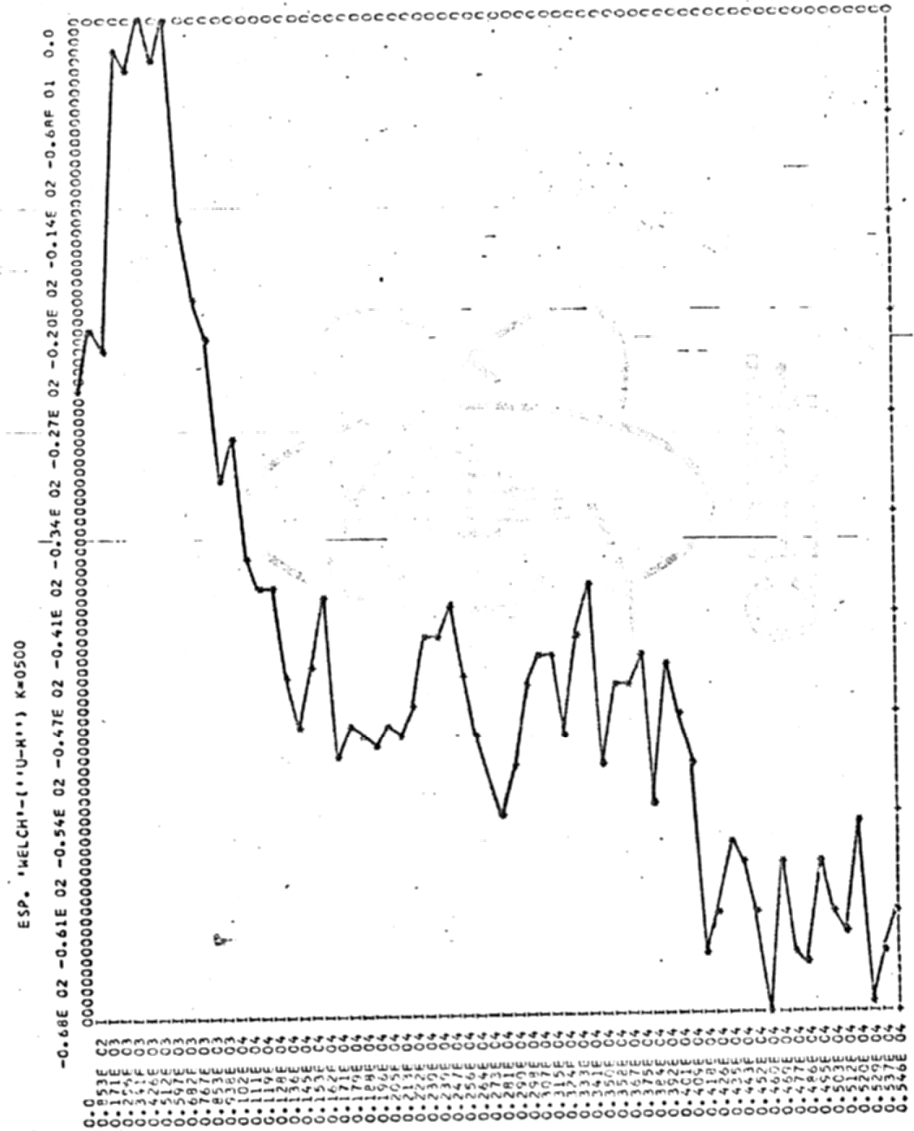


Figura 16 - Esp. (SPWCH) - /u/, y.masc.

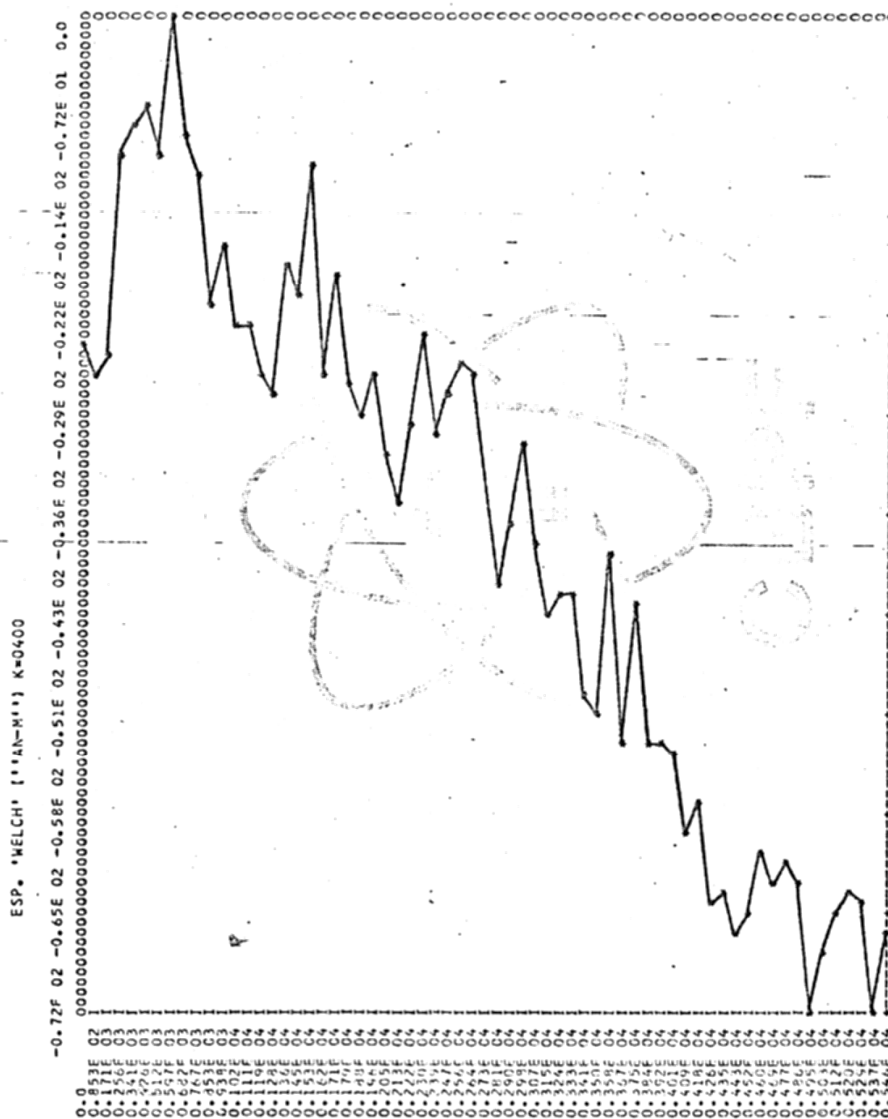


Figura 17 - Esp. (SPWCH) - /an/, v.masc.

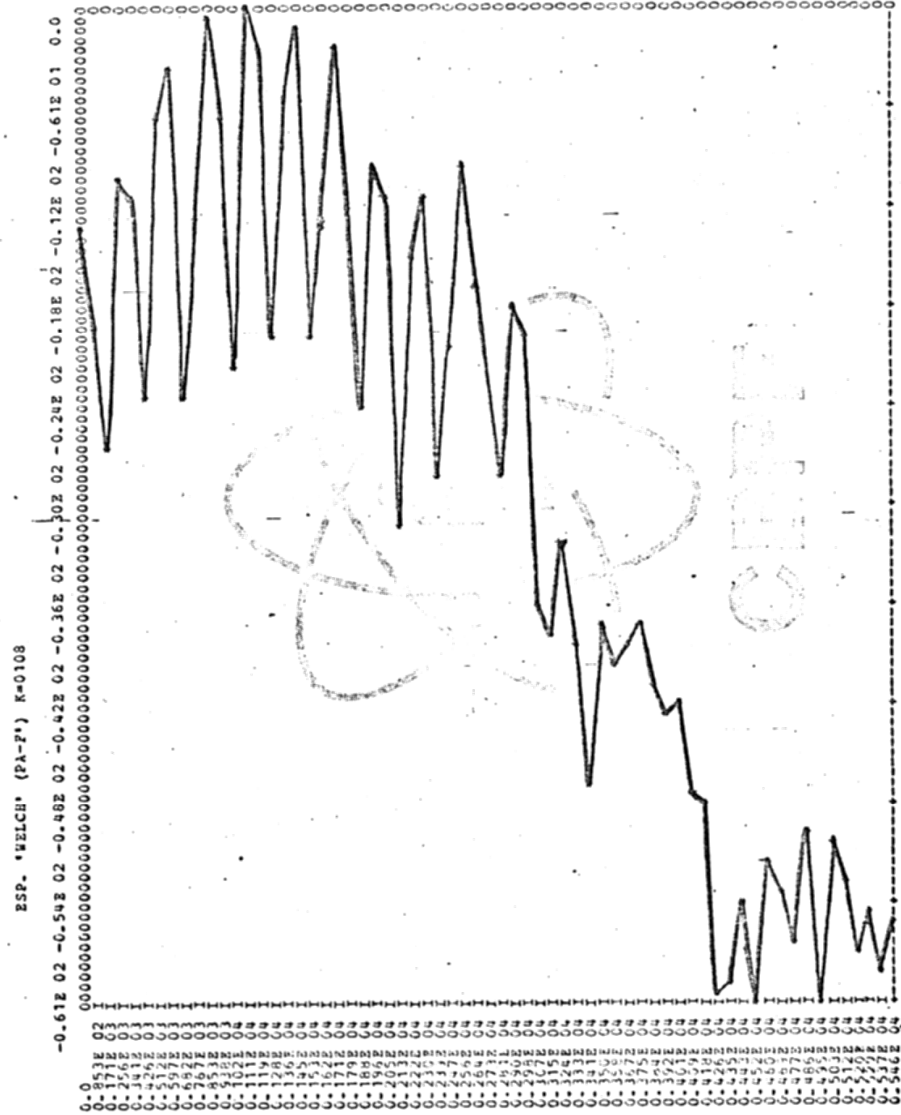


Figura 18 - Esp. (SPWCH) - /pa/, v. fem.

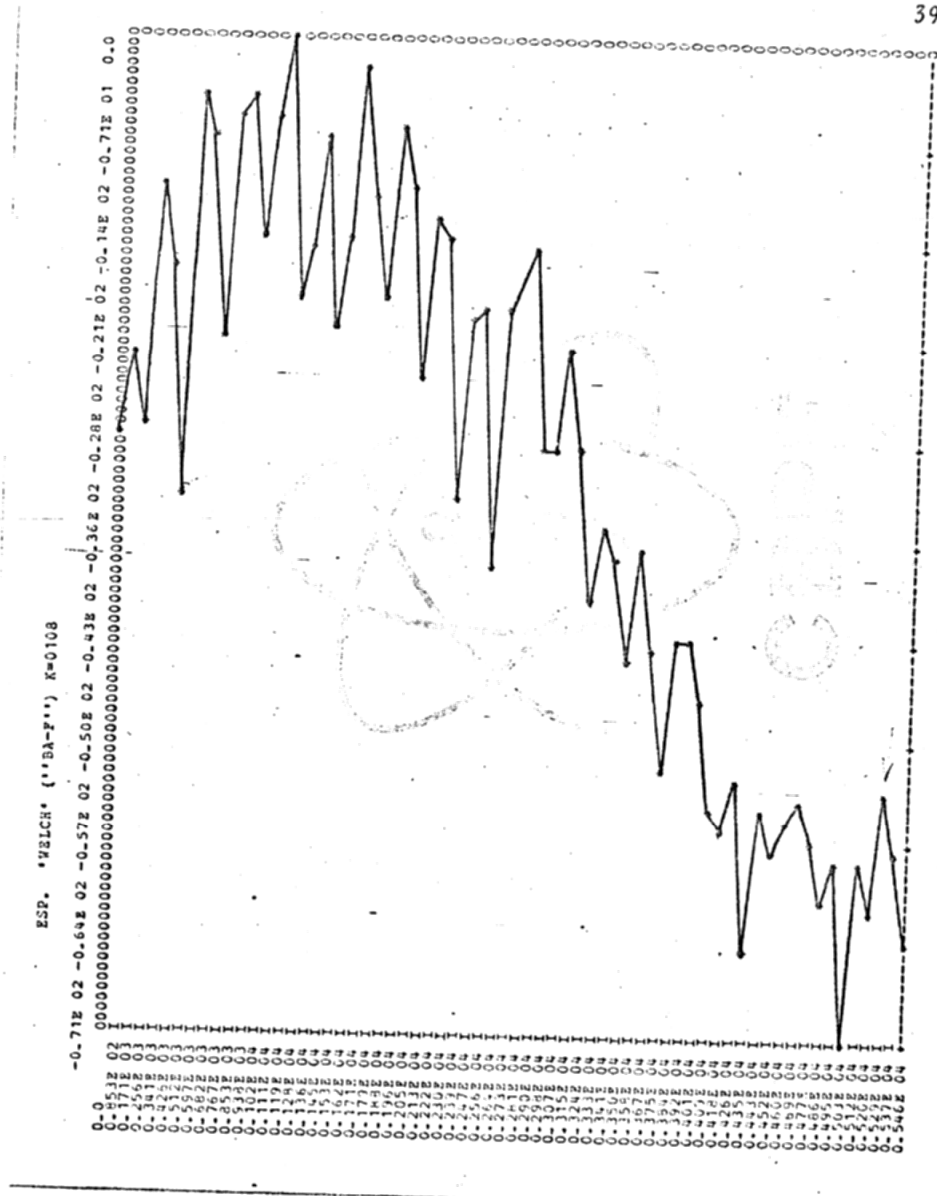


Figura 19 - Esp. (SPWCH) -/ba/, v. fem.



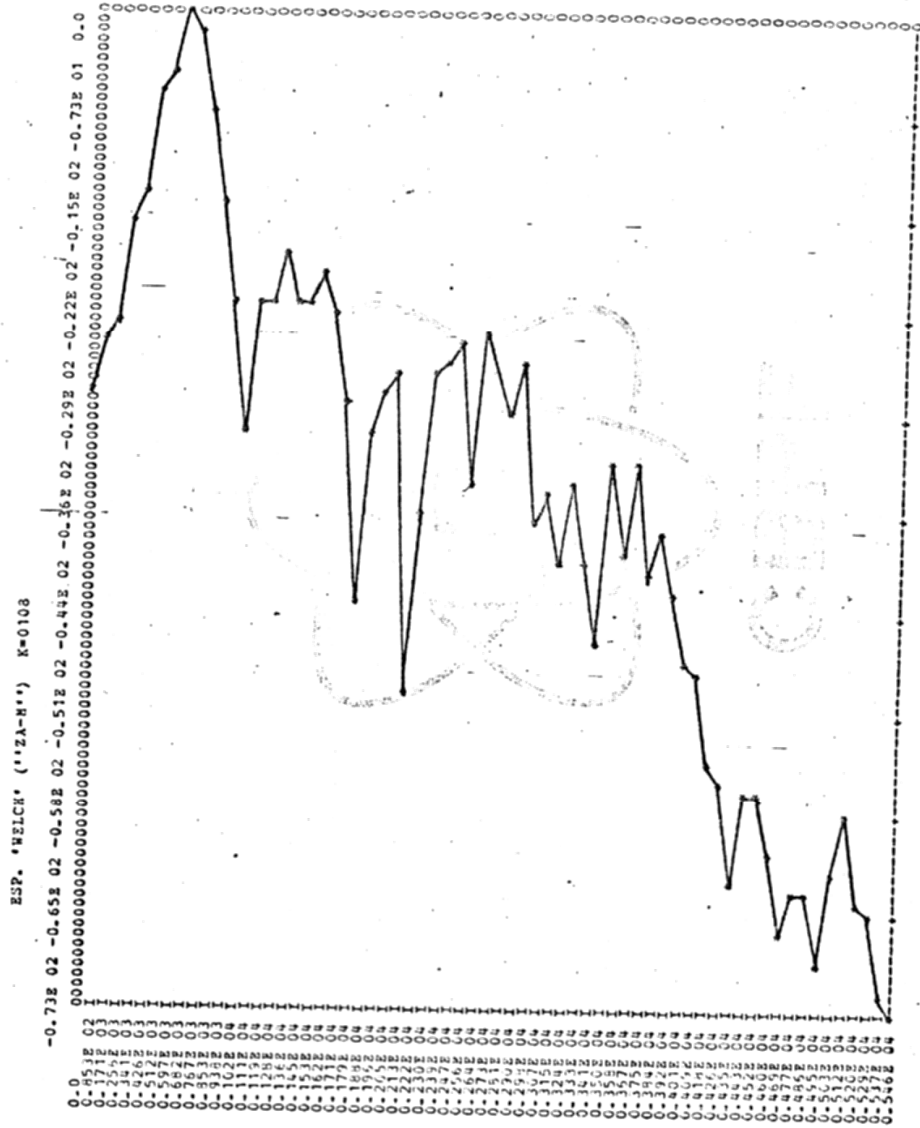


Figura 21 - Esp. (SPWCH) -  $|z̄a|$ , v.mas.c.

4.2.2- O programa ALISA utiliza um simples processo de alisamento ("smoothing"), descrito, p.ex.:, por .... Schwartz & Shaw<sup>13</sup>. O periodograma "alisado" corresponde à média de vários valores adjacentes,

$$\bar{S}(\omega_i) = \frac{1}{2L+1} \sum_{k=i-L}^{i+L} S_N(\omega_k) \quad (10)$$

isso é, temos a média de  $(2L+1)$  valores do periodograma próximos à frequência  $\omega_k$ , onde  $S_N(\omega_k)$  é o periodograma, obtido diretamente da transformada de Fourier de  $N$  valores do sinal discreto  $x_n$  (trabalhou-se com  $N = 256$ ).

4.2.2.1- Essa operação de obter a média é análoga à operação de filtrar sinais discretos no domínio do tempo. Ora, pode-se provar, facilmente, que a filtragem reduz a variância do sinal. Assim, a variância do periodograma "alisado",  $\bar{S}(\omega_i)$ , será inferior à do periodograma escabroso:

$$\text{var.} \left[ \bar{S}(\omega_i) \right] = \sum_{k=i-L}^{i+L} \text{var.} \left[ \frac{1}{2L+1} S_N(\omega_k) \right] \quad (11)$$

$$= \sum_{k=i-L}^{i+L} \frac{1}{(2L+1)^2} \text{var.} \left[ S_N(\omega_k) \right]$$

$$= \frac{1}{(2L+1)^2} \sum_{k=i-L}^{i+L} \text{var.} \left[ S_N(\omega_k) \right] \quad (12)$$

e, supondo o processo estacionário, tem-se, finalmente:



$$\begin{aligned} \text{var.} \left[ \bar{S}(\omega_i) \right] &= \frac{1}{(2L+1)^2} (2L+1) \text{var.} \left[ S_N(\omega_i) \right] \\ &= \frac{1}{(2L+1)} \text{var.} \left[ S_N(\omega_i) \right] \end{aligned} \quad (13)$$

ou seja, uma redução a  $1/(2L+1)$  do valor original.

4.2.2.2- Os dois menores valores de  $L$  oferecidos a opção eram  $L = 1$  e  $L = 2$ , causando, o primeiro, redução de  $1/3$  e o segundo redução de  $1/5$  no valor da variância.

4.2.2.3- Optou-se pelo segundo,  $L = 2$ , sabendo, a priori, que ele causaria maior tendenciosidade ("bias") no periodograma "alisado". Entretanto, intencionalmente, preferiu-se reduzir a variância para obter, nos espectros, contornos gerais mais facilmente observáveis.

4.2.2.4- Poder-se-á, caso desejado reduzir a tendenciosidade aumentando o número de pontos,  $N$ , de 256 para 512.

4.2.2.5- Aplicou-se o programa ALISA aos seguintes fonemas: /a/, /ê/, /i/, /ô/, /u/, /an/ (voz masculina e feminina) e os espectros obtidos permitem fazer as seguintes observações (ver Figuras 22 a 27):

- o aspecto geral da envoltoria dos espectros obtidos com

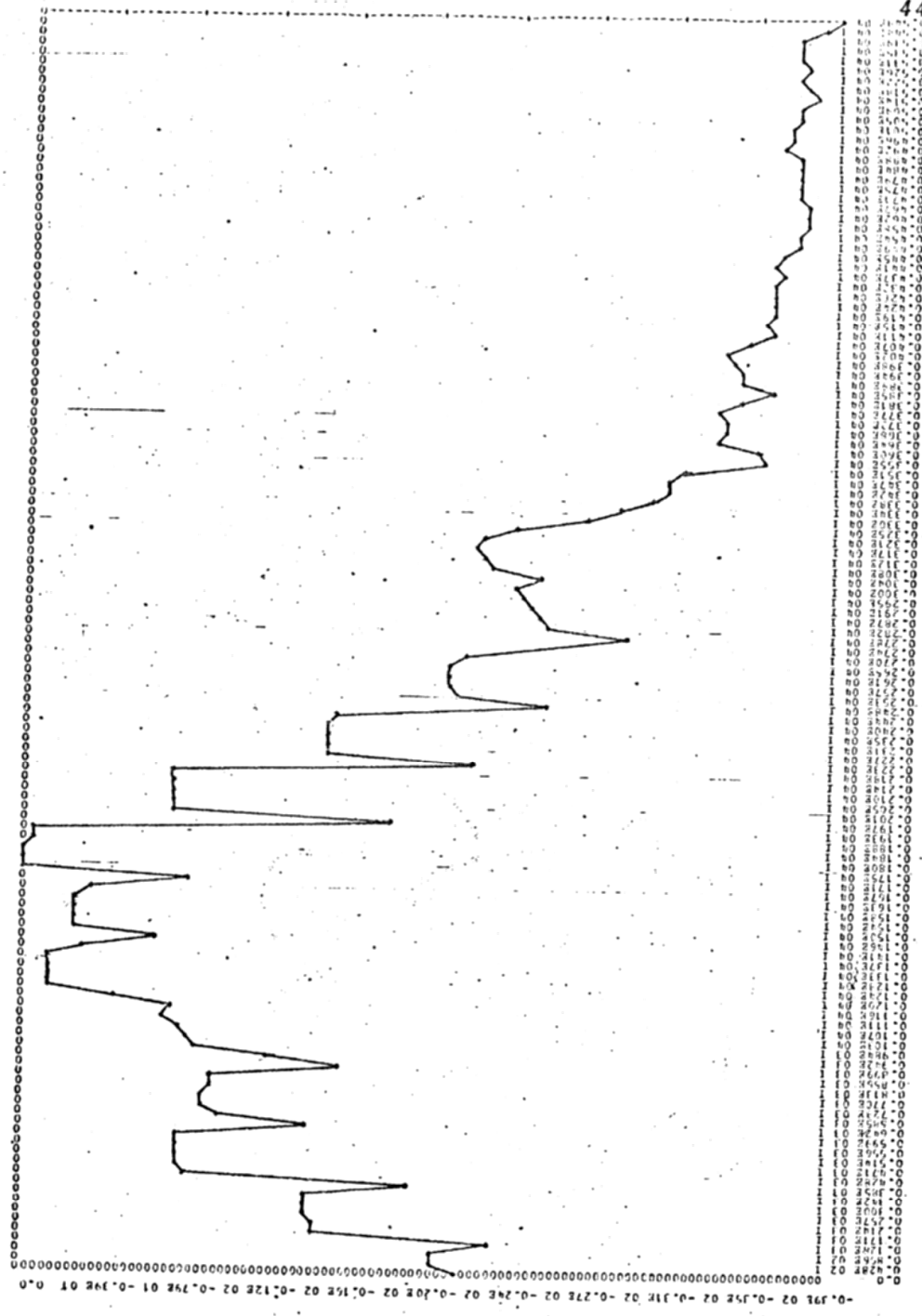


Figura 22 - Esp. (ALISA) - /a/, v. 6cm.

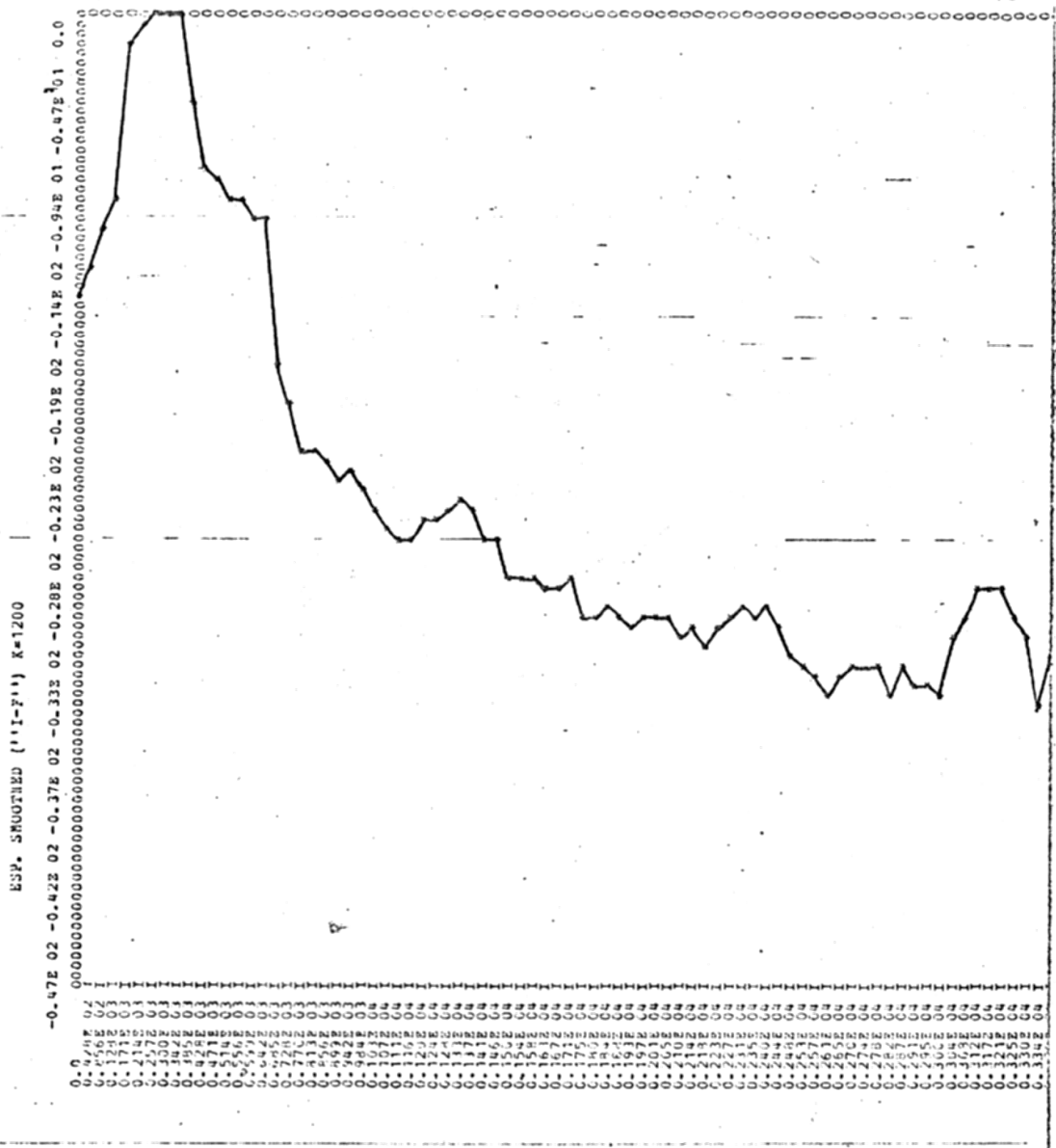


Figura 23 - Esp. (ALISA) - /i/, v.6em.

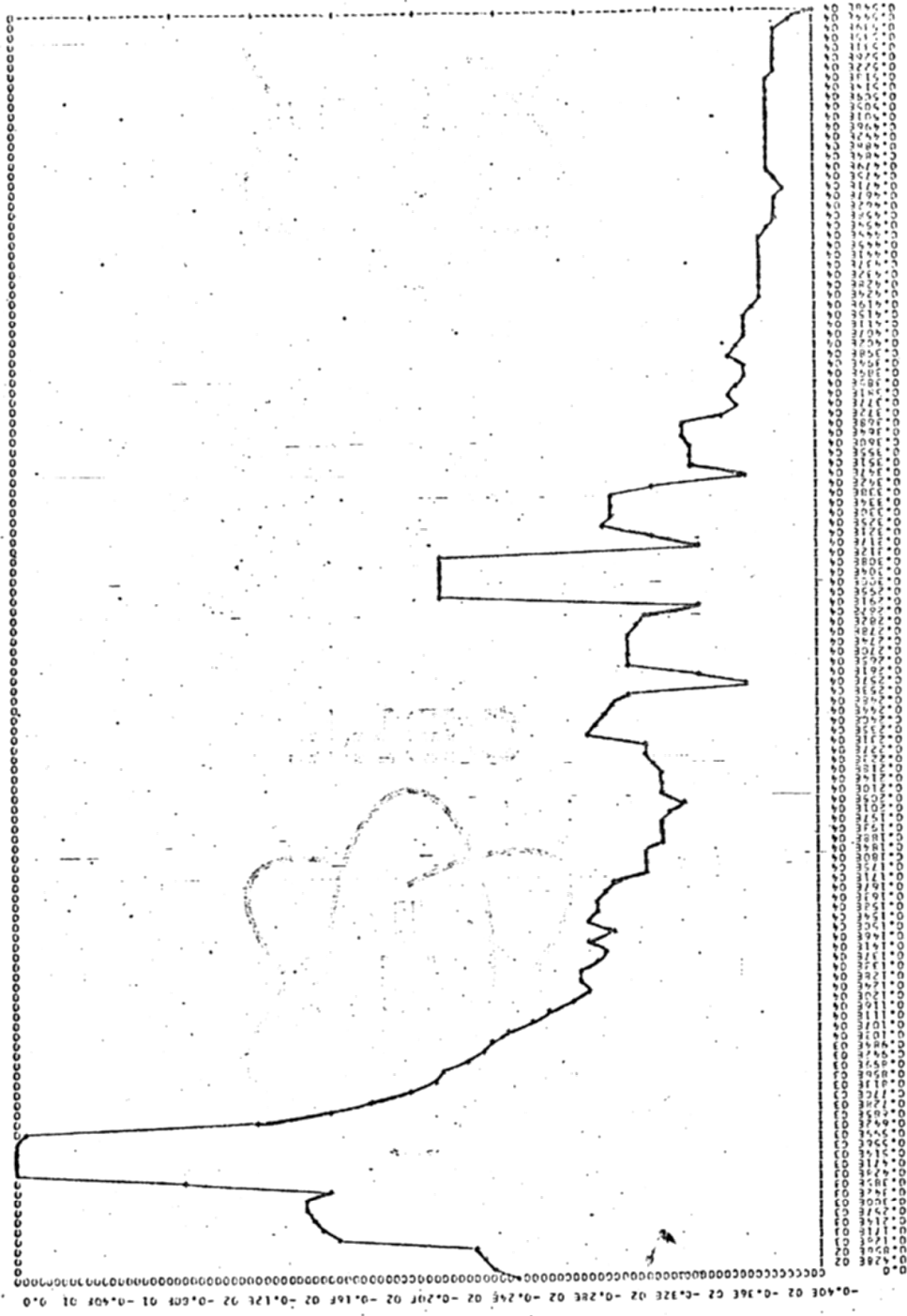


Figura 24 - Esp. (ALISA) - 12/, v. gem.

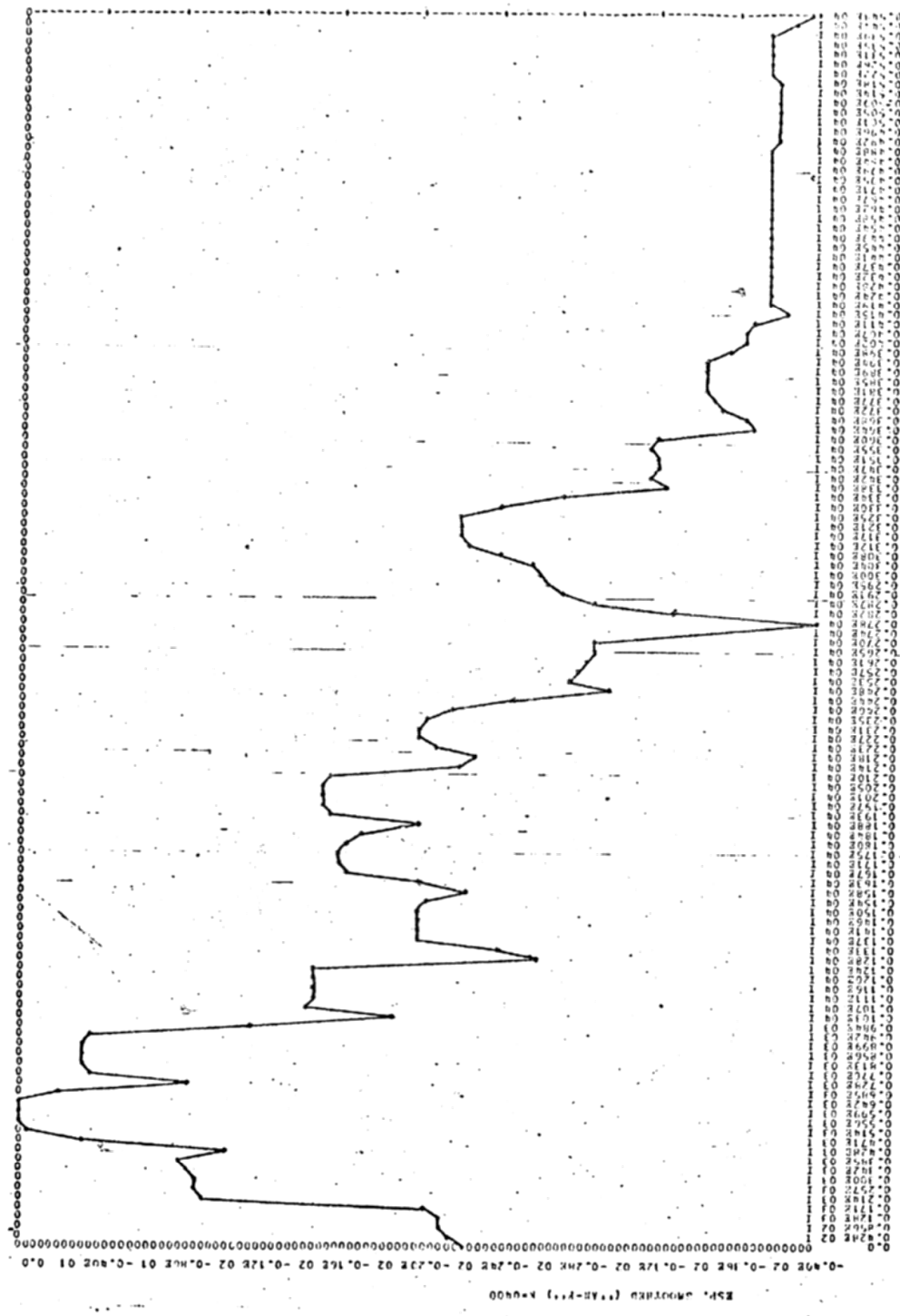


Figura 25 - Esp. (ALISA) - /an/, v. gem.

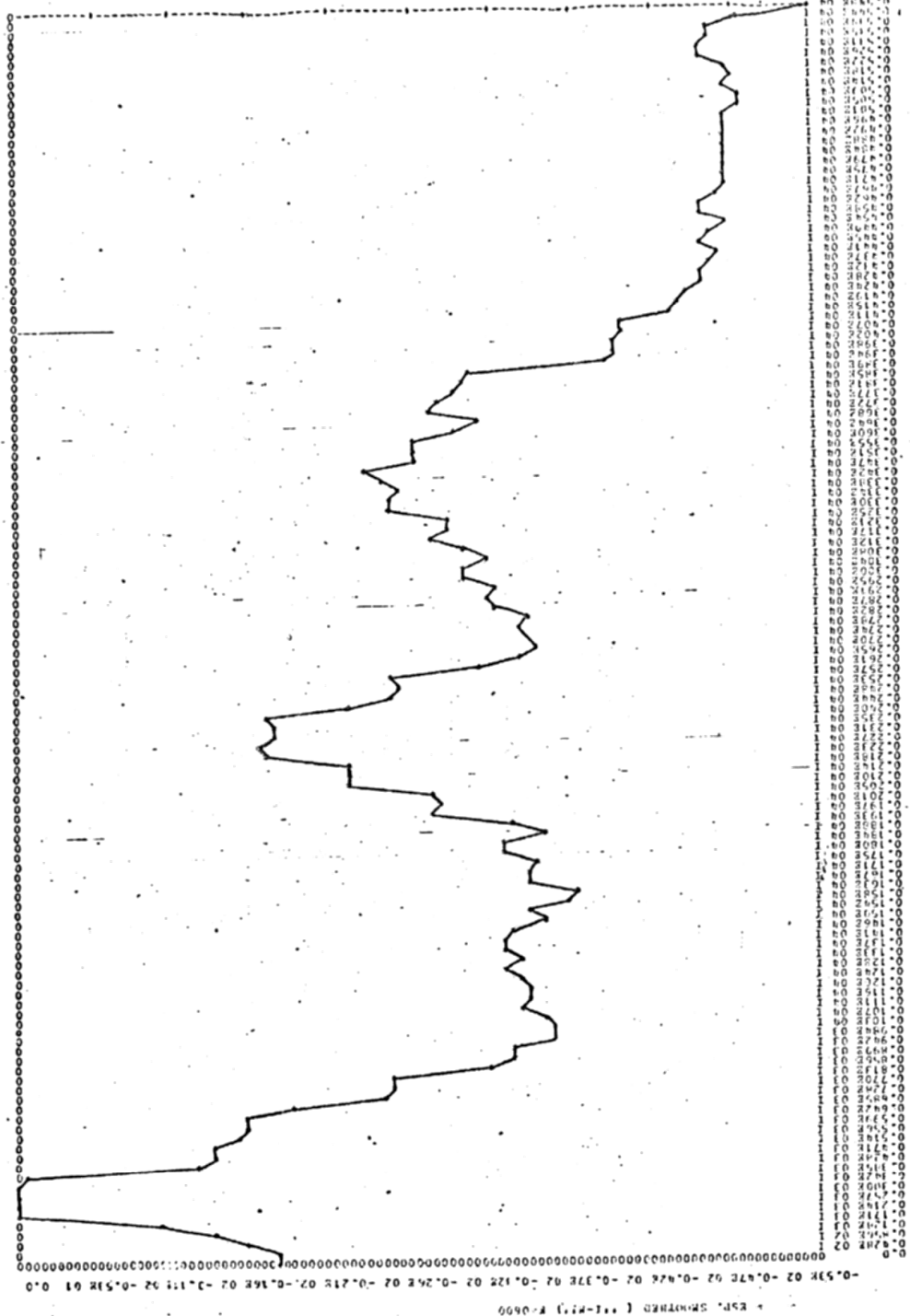


Figura 26 - Esp. (ALISA) - /λ/, v.mad.с.

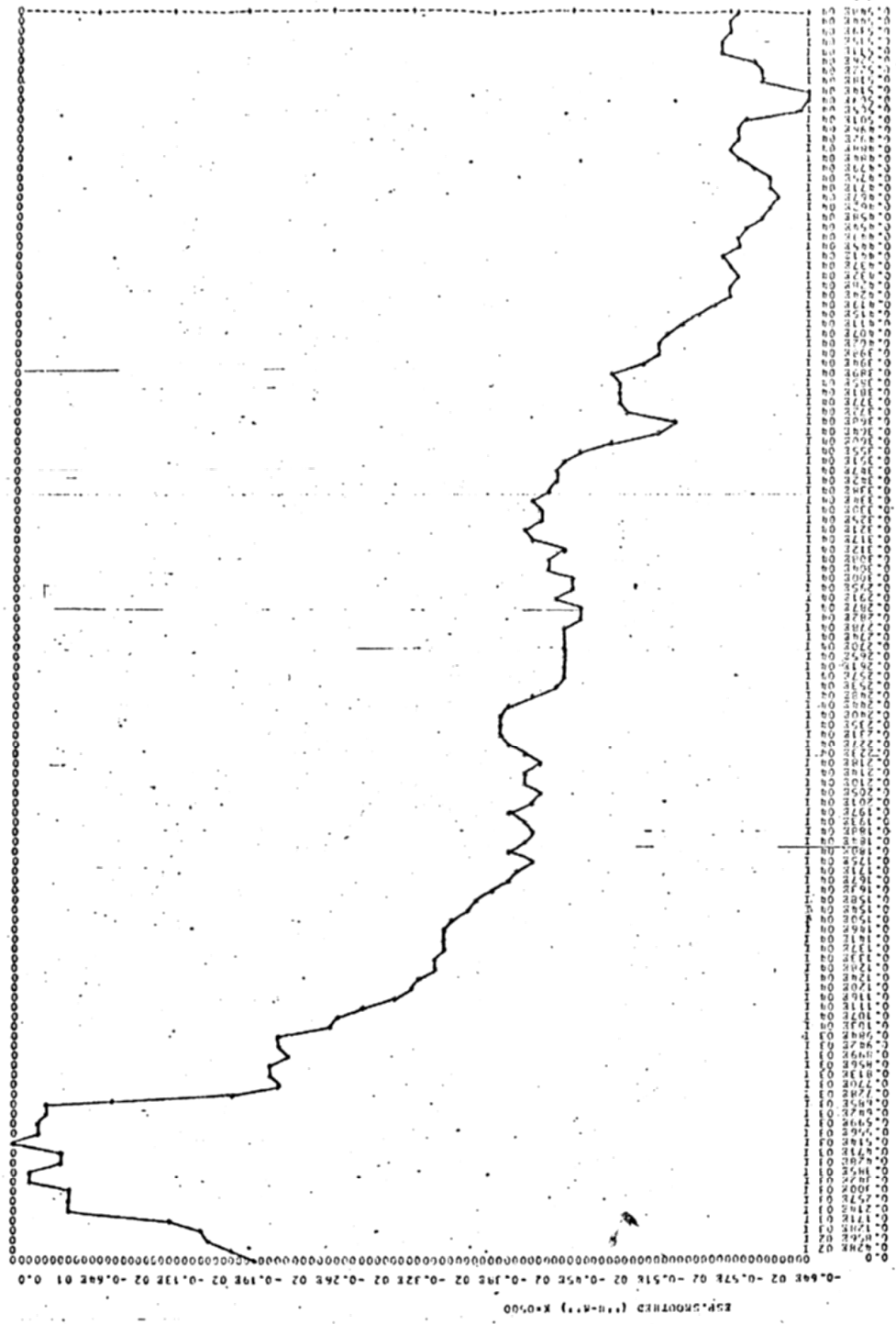


Figura 27 - Esp. (ALISA) - /u/, v.masc.

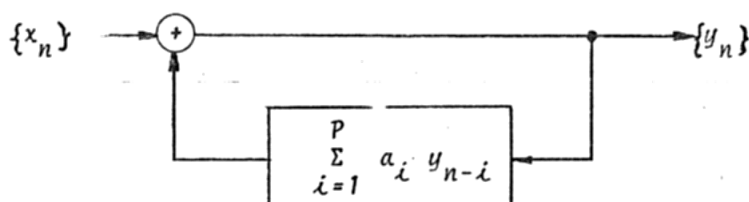
- o programa ALISA é muito semelhante ao análogo dos espectros SPWCH;
- tanto no som /a/ feminino quanto no /a/ masculino, os formantes estão mais agrupados e mais nitidamente definidos; por exemplo: o /a/ masculino mostra, claramente, três formantes de amplitude acentuada; (em torno de 685 Hz, 1460 Hz e 2440 Hz);
  - os espectros do /ê/ e do /i/ (femininos e masculinos) revelam uma brecha, uma separação nítida entre o primeiro e o segundo formante, separação essa já notada no espectro SPWCH e, conforme citado, responsável pela forma de onda recortada, "ruidosa" daqueles sons no domínio do tempo;
  - os espectros do /ô/ e do /u/ (femininos e masculinos) mostram o primeiro formante bem nítido e os outros bem atenuados, o que deve explicar as formas de onda mais "limpas", com periodicidade mais observável que as dos sons /ê/ e /i/;
  - o espectro do som nasalado (/an/ - masculino e feminino) apresenta formantes diferentes dos observados no espectro da vogal oral /a/, fato esse, já citado, decorrente da passagem do ar pelo trato nasal.

4.2.3- Enquanto os programas SPWCH e ALISA fornecem estimativas não-paramétricas dos periodogramas, o programa FRMNT proporciona estimativa paramétrica baseada na estimação de máxima verossimilhança ("maximum-likelihood"),



apresentada de modo simplificado, nos próximos itens.

4.2.3.1- Suponhamos o sinal de voz constituindo processo estocástico, ergódico e, discretizado, aparecendo como resultado (saída) do sistema abaixo esquematizado:



isso é, um sistema discreto, linear, invariável no tempo, caracterizado pela equação de diferenças finitas:

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + a_3 y_{n-3} + \dots + a_p y_{n-p} + x_n$$

e condições iniciais

$$y_{-1} = y_{-2} = y_{-3} = \dots y_{-p} = 0 \quad (14)$$

Esse sistema, portanto, terá uma função de transferência do tipo "sô-de-polos"

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - a_3 z^{-3} \dots - a_p z^{-p}} \quad (15)$$

Essa primeira simplificação está coerente com a natureza da percepção acústica caracterizada pela maior sensibilidade do ouvido humano às variações de amplitude dos po-  
los, comparada com a sensibilidade à variação dos zeros<sup>12</sup>.

4.2.3.2- Trata-se, agora, de encontrar, para os parâmetros  $a_1, a_2, a_3, \dots, a_p$ , valores que tornem os dados observados mais prováveis que quaisquer outros. O primeiro passo é estabelecer a f.d.p. (função densidade da probabilidade) conjunta das amostras observadas em termos dos parâmetros desconhecidos<sup>13</sup>.

4.2.3.3- Seja a função densidade conjunta:

$$P_{y_0, y_1, \dots, y_{N-1}}(y_0, y_1, \dots, y_{N-1}; a_1, a_2, \dots, a_p) \quad (16)$$

das  $N$  observações. Considerando as probabilidades condicionais, teremos:

$$\begin{aligned} P_{y_0, y_1, \dots, y_{N-1}} &= \\ &= P_{y_0} P_{y_1|y_0} P_{y_2|y_1, y_0} \dots P_{y_{N-1}|y_0, y_1, \dots, y_{N-2}} \quad (17) \end{aligned}$$

4.2.3.4- Como segunda simplificação, suponha-se que na entrada existe um processo  $\{x_n\}$  estacionário, Gaussiano, de média nula, variância  $\sigma_x^2$  e correlação  $\sigma_x^2 \delta(n)$ , isso é, as amostras são totalmente independentes. Sendo a estrutura linear e estável, na saída haverá um processo estacionário Gaussiano<sup>16, 17</sup>. Assim:

$$P_{y_0}(y_0) = \frac{1}{\sigma_y \sqrt{2\pi}} \exp\left(-\frac{y_0^2}{2\sigma_y^2}\right) \quad (18)$$

onde  $\sigma_y^2$  é a variância de  $y_n$ .

4.2.3.5- A função  $P_{y_1|y_0}$  pode ser deduzida da equação do sistema fazendo:

$$y_1 = a_1 y_0 + x_1 \quad (19)$$

Como  $x_1$  é independente de  $x_0$  e, portanto, de  $y_0$ , segue-se

$$P_{y_1|y_0} = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{(y_1 - a_1 y_0)^2}{2 \sigma_x^2} \right] \quad (20)$$

4.2.3.6- Para  $P_{y_2|y_1, y_0}$  tem-se, semelhantemente

$$y_2 = a_1 y_1 + a_2 y_0 + x_2 \quad (21)$$

e, como  $x_2$  independe de  $x_1$  e de  $x_0$ ,

$$P_{y_2|y_1, y_0} = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{(y_2 - a_1 y_1 - a_2 y_0)^2}{2 \sigma_x^2} \right] \quad (22)$$

Generalizando, virã:

$$P_{y_k|y_{k-1} \dots y_0} = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[ -\frac{(y_k - a_1 y_{k-1} \dots a_k y_0)^2}{2 \sigma_x^2} \right] \quad (23)$$

4.2.3.7 - Substituindo a expressão (18) e as expressões semelhantes à (23) na relação (17), virã

$$P_{y_0, y_1, y_2, \dots, y_{N-1}; a_1, a_2, \dots, a_p} =$$

$$= \frac{1}{\sigma_y \sqrt{2\pi}} \exp\left(-\frac{y_0^2}{2\sigma_y^2}\right) \times$$

$$\times \frac{1}{\sigma_x^{N-1} (\sqrt{2\pi})^{N-1}} \exp\left[-\frac{\sum_{i=1}^{N-1} (y_i - a_1 y_{i-1} \dots a_i y_0)^2}{2\sigma_x^2}\right] \quad (24)$$

Derivando, sucessivamente, em relação a  $a_1, a_2, a_3, \dots, a_p$ , igualando a zero, e compondo os termos encontra-se o seguinte sistema:

$$a_1 \sum_{i=1}^{N-1} y_{i-1}^2 + a_2 \sum_{i=1}^{N-1} y_{i-1} y_{i-2} + \dots +$$

$$+ a_p \sum_{i=1}^{N-1} y_{i-1} y_{i-p} = \sum_{i=1}^{N-1} y_i y_{i-1}$$
  

$$a_1 \sum_{i=1}^{N-1} y_{i-1} y_{i-2} + a_2 \sum_{i=1}^{N-1} y_{i-2}^2 + \dots +$$

$$+ a_p \sum_{i=1}^{N-1} y_{i-2} y_{i-p} = \sum_{i=1}^{N-1} y_i y_{i-2}$$
  

.....

.....

$$a_1 \sum_{i=1}^{N-1} y_{i-p} y_{i-1} + a_2 \sum_{i=1}^{N-1} u_{i-p} y_{i-1} + \dots +$$

$$+ a_p \sum_{i=1}^{N-1} y_{i-p}^2 = \sum_{i=1}^{N-1} y_i y_{i-p} \quad (25)$$

4.2.3.8- Dividindo ambos os membros de todas as equações do sistema (25) por  $(N-1)$  e lembrando que o processo, por hipótese, é estacionário, tem-se, finalmente,

$$a_1 R_y(0) + a_2 R_y(1) + \dots + a_p R_y(p) = R_y(1)$$

$$a_1 R_y(1) + a_2 R_y(0) + \dots + a_p R_y(p-1) = R_y(2)$$

.....  
 .....

$$a_1 R_y(p) + a_2 R_y(p-1) + \dots + a_p R_y(0) = R_y(p) \quad (26)$$

onde  $R_y(k)$  é a autocorrelação de  $y_n$ ; resolvendo (26) e substituindo os valores de  $a_1, a_2, \dots, a_p$  na igualdade (15), tem-se a função de transferência do sistema. A substituição de  $z$  por  $e^{j\omega T_s}$ , onde  $T_s$  é o período de amostragem, permitirá calcular a resposta em função de  $\omega$ , frequência em rad/s. O programa FRMNT, basicamente, arma e resolve o sistema (26) e depois resolve a decorrente equação:

$$z^p - a_1 z^{p-1} - a_2 z^{p-2} - \dots - a_p = 0 \quad (27)$$

4.2.3.9- Aplicando o programa FRMNT aos fonemas /a/, /ê/, /i/, /ô/, /u/, /an/ (femininos e masculinos) determinaram-se espectros cujas envoltórias permitem fazer, entre outras, as seguintes observações (ver Figuras 28 a 35):  
 - a vogal /a/ tem bem caracterizados seus formantes ra-



Figura 28 - Env. espectral - |a|, v.fem.

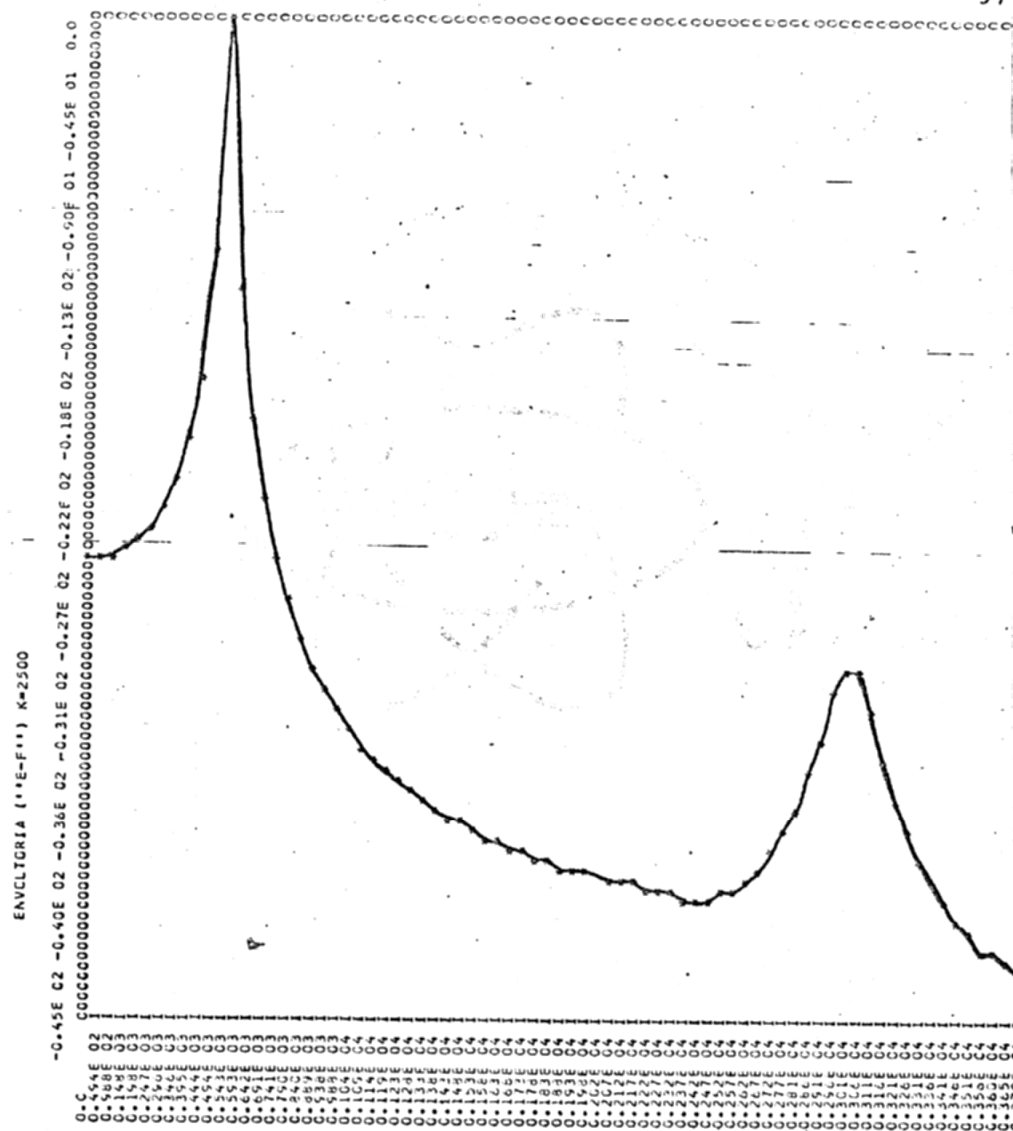


Figura 29 - Enf. espectral -  $|e|$ , v. fem.

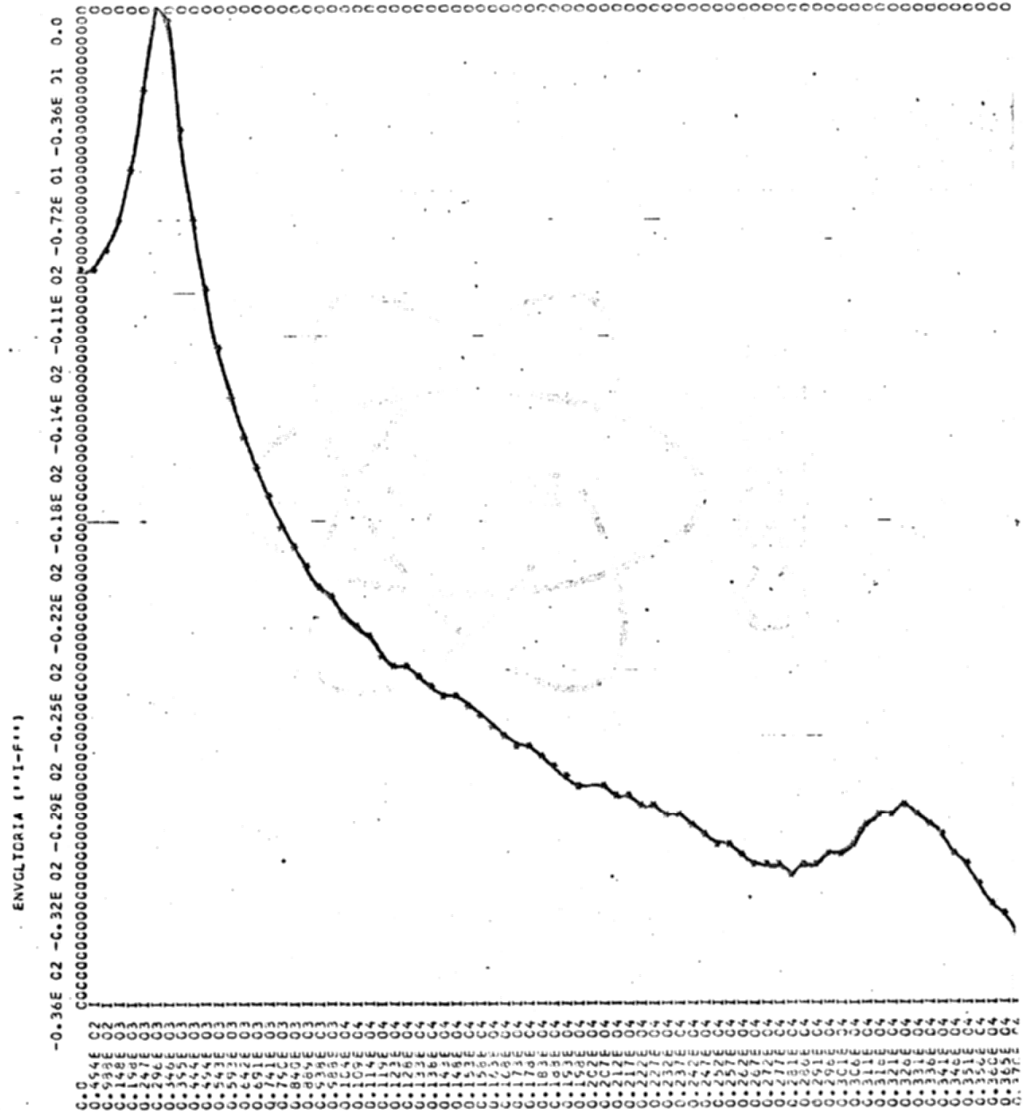


Figura 30 - Env. espectral - /i/, v.fem.



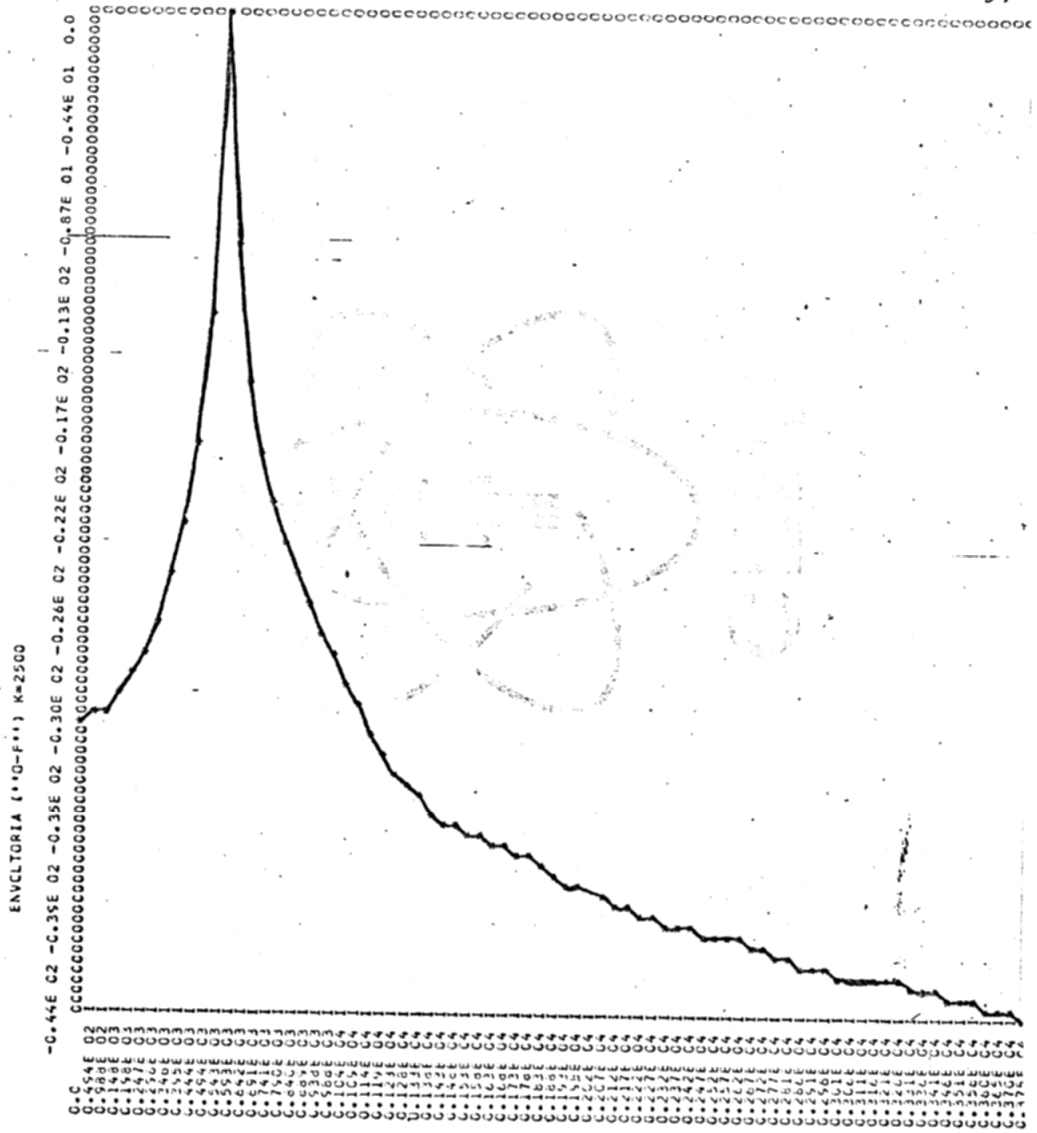


Figura 31 - Env. espectral -  $|\hat{\sigma}|$ , v.fem.



Figura 32 - Env. espectral - /u/, v.masc.

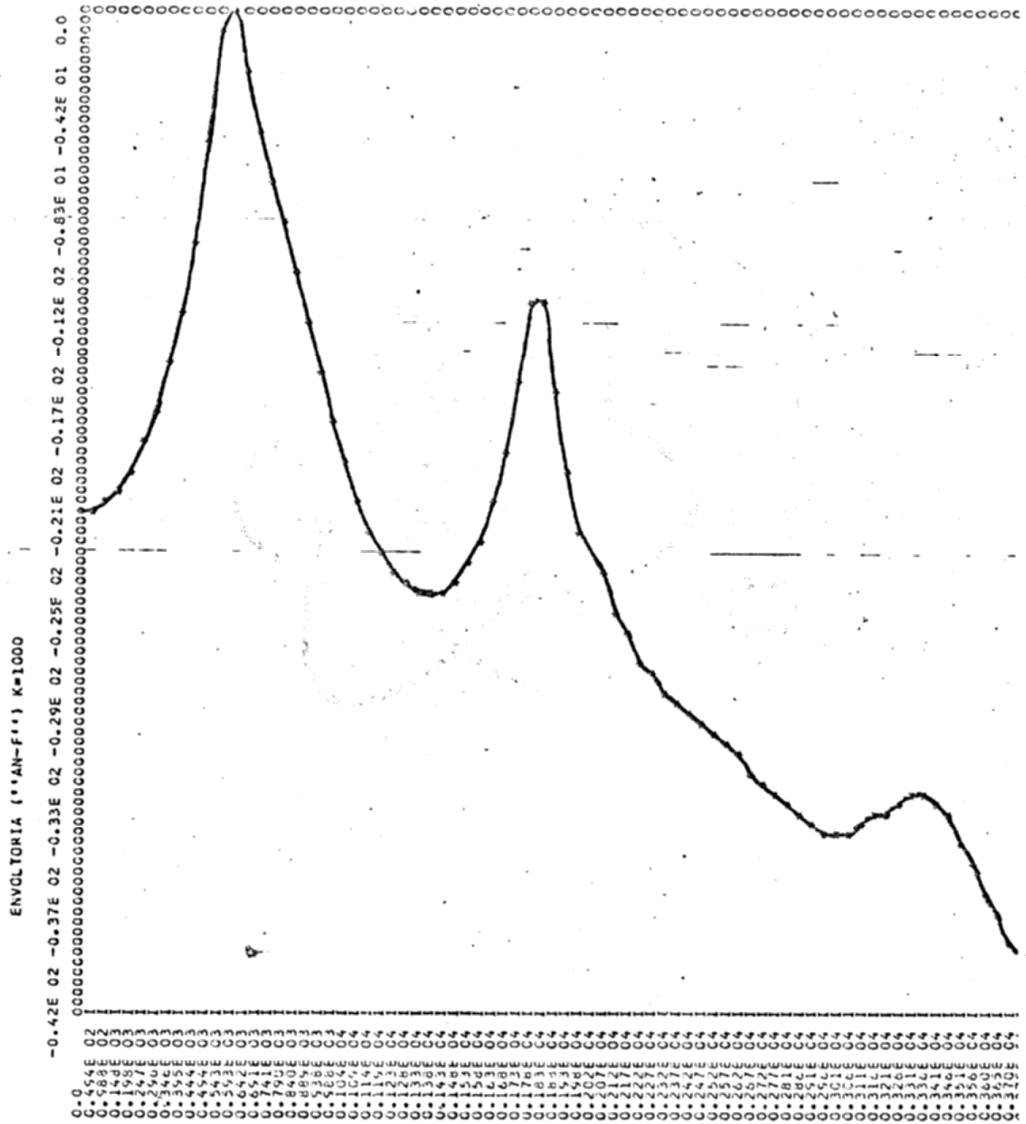


Figura 33 - Env. espectral - /an/, v.fem.



Figura 34 - Env. espectral - /a/, v.masc.

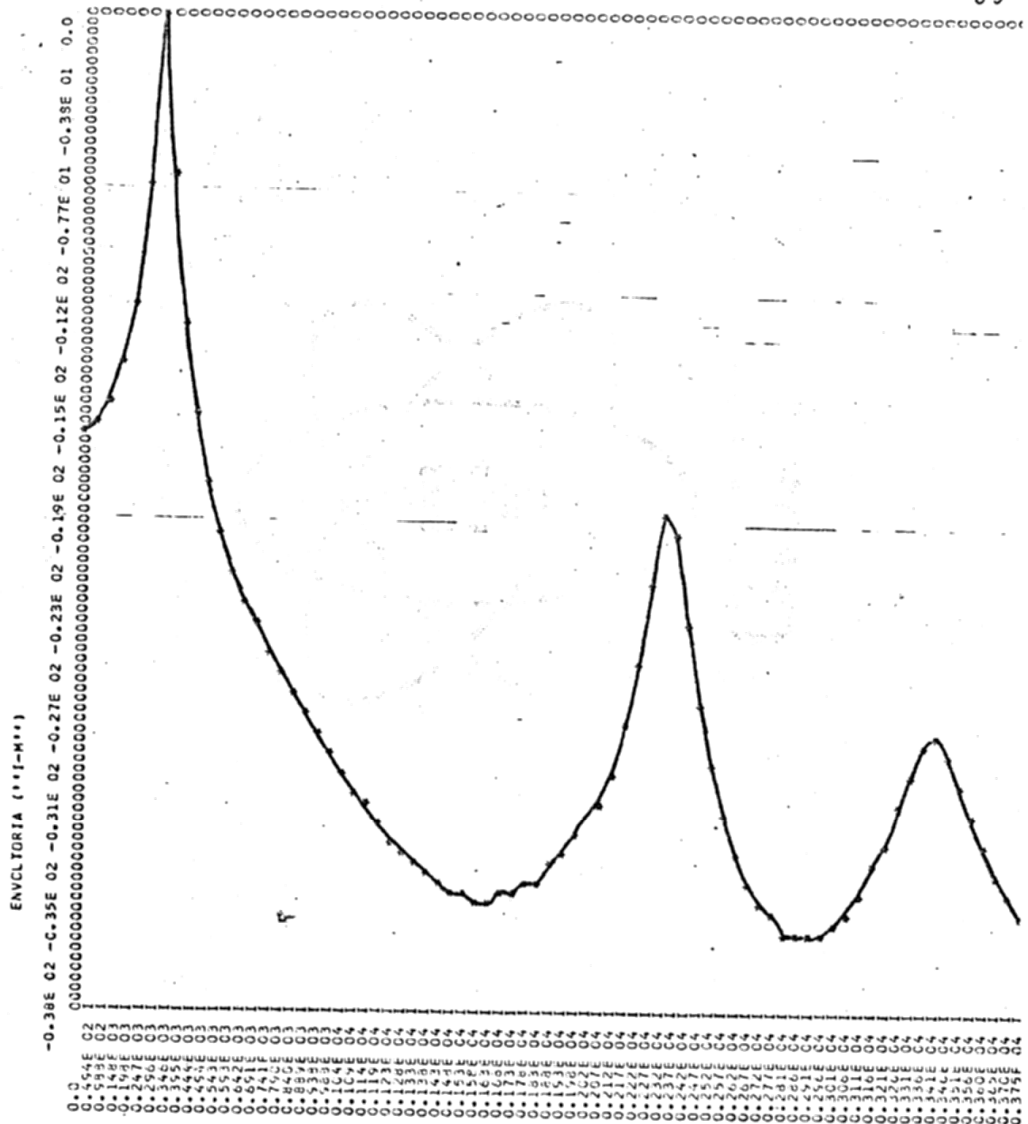


Figura 35 - Env. espectral - /i/, v.masc.

- ção pela qual é considerada a vogal mais autêntica<sup>1</sup>;
- confirmando os resultados obtidos com as estimativas não-paramétricas (SPWCH e ALISA), os fonemas /ê/ e /i/ apresentam nítida separação entre os primeiros formantes<sup>4</sup>;
  - também confirmando resultados anteriores, a nasalização da vogal introduz, na envoltória, diferentes formantes<sup>1</sup>;
  - o método paramétrico de estimação revela-se superior aos métodos não-paramétricos porquanto além de fornecer envoltórias espectrais bem definidas, não exige, como os outros, o aumento do número de amostras para aumentar a resolução espectral; esse aumento, quando exagerado, entra em conflito com uma característica básica do sinal de voz que é a de ser estatisticamente estacionário em intervalos curtos ("short-time").

4.2.3.20- O programa FRMNT, além de traçar a envoltória espectral, lista todos os polos da equação (27) e respectivas faixas passantes dadas pela equação:

$$B_i = (\ln |z_i|) / \pi T_s \quad (28)$$

onde  $|z_i|$  é o módulo do polo complexo,  $T_s$  é o período de amostragem e  $\ln$  representa logaritmo Neperiano. Assim, convencionando-se uma determinada faixa passante relativa,  $B_i/F_i$ , podem-se obter os valores dos principais formantes<sup>12</sup>.

4.2.3.11- Através de uma iteração o mesmo programa permite obter envoltórias espectrais de segmentos sucessivos de frases inteiras. Assim, foi aplicado à frase: "O João quebrou o dedo", em voz feminina e masculina e, levantando-se os formantes sucessivos dentro de intervalos de, aproximadamente, 18,1 milisegundos (correspondendo à taxa: 55 amostras/s), obtiveram-se os espectros da "variação dos formantes" (ver Figuras de nº 36 a 41).

4.2.3.12- Os espectros de variações dos formantes mostram nitidamente, serem essas variações bastante lentas, admitindo, portanto, taxas de amostragem muito baixas, entre 60 e 100 Hz, por exemplo. Esses valores são coerentes com as baixas taxas de transmissão usadas nos sistemas de voz sintética operando com formantes, p. ex. taxas de 600 bits/s<sup>15,3</sup>. Possivelmente, enlaces telefônicos ou radio-telefônicos onde a qualidade pode ser prejudicada em benefício da inteligibilidade venham a admitir o uso de tais sistemas. A disponibilidade atual de meios de transmissão de elevada capacidade, tais como os cabos de fibra ótica, atenuou o esforço na pesquisa do emprego dessas taxas muito reduzidas nos serviços telefônicos comerciais.

4.2.3.13- O programa FRMNT permite, ainda, fazer-se o registro dos sucessivos valores dos coeficientes  $a_k$  ( $k = 0, 1, 2, \dots, p$ ) da equação (27) ao longo de uma

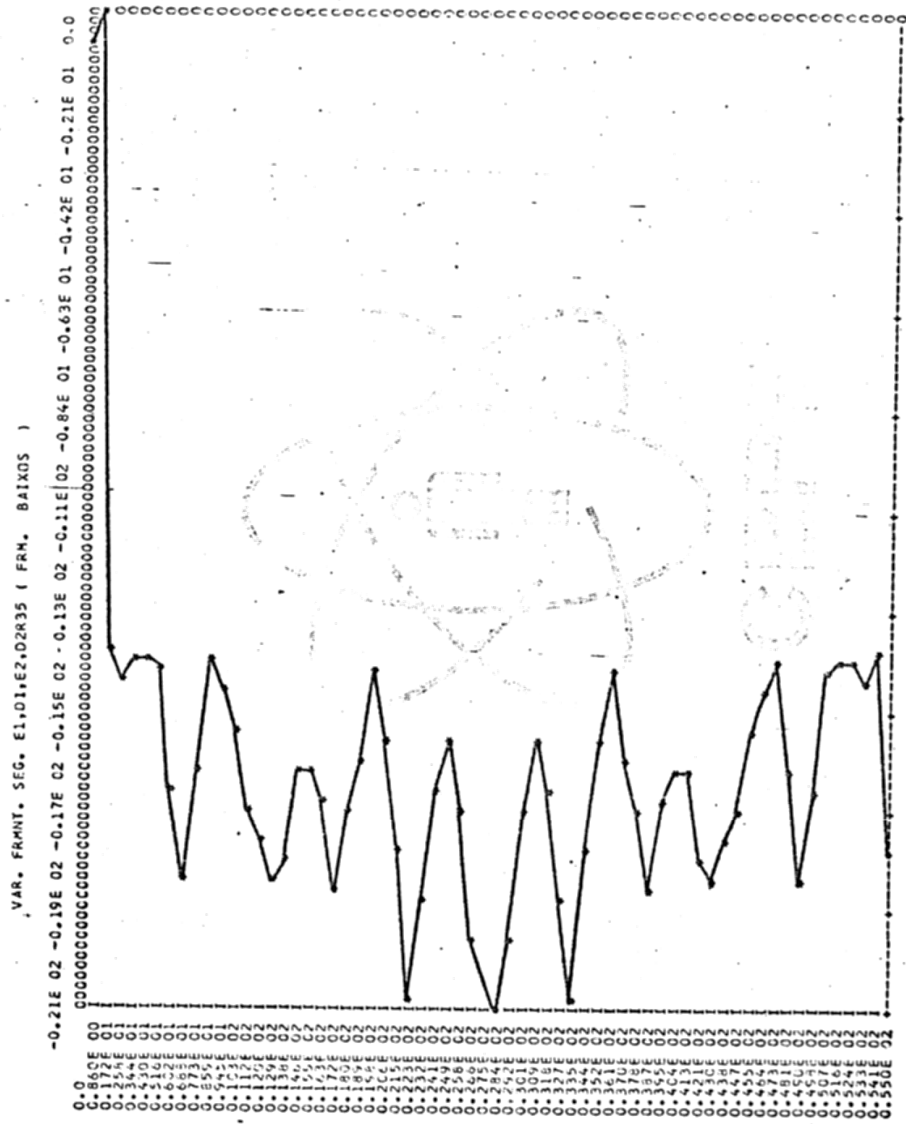


Figura 36 - Esp. var. formantes (baixos-v.fem.)



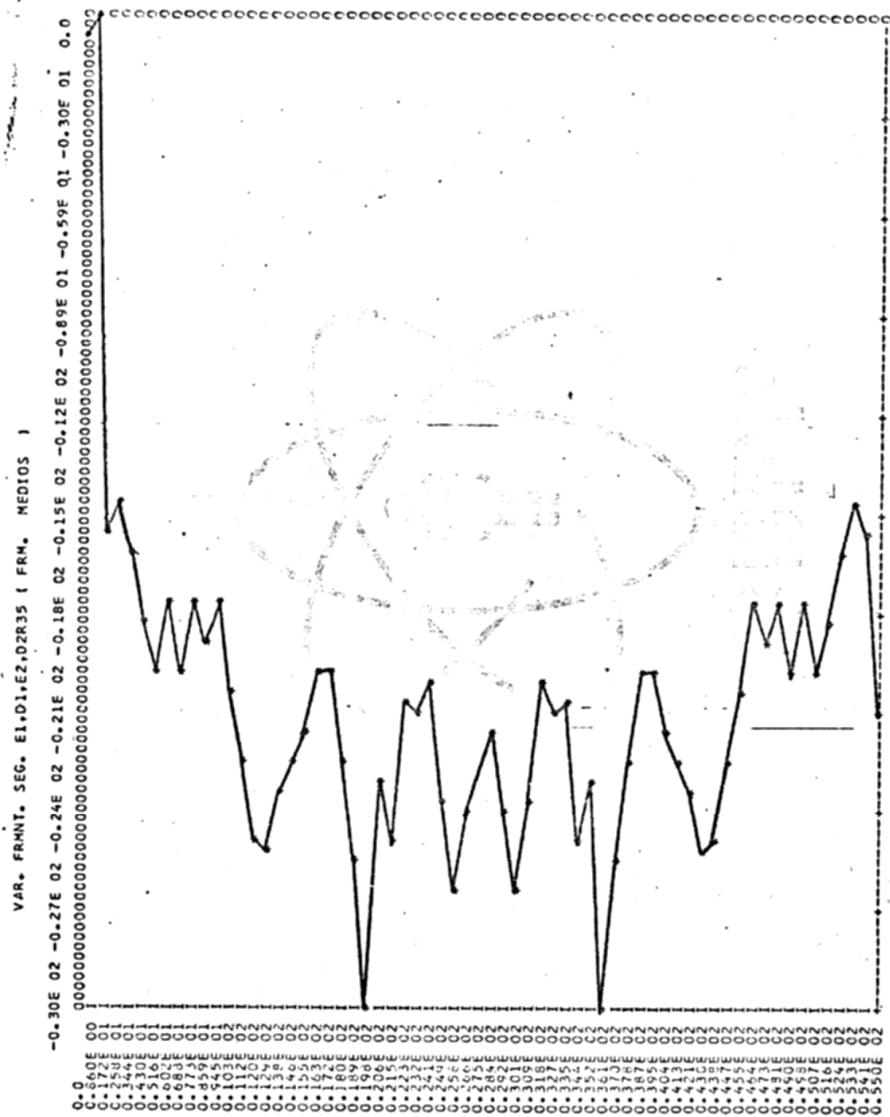


Figura 37 - Esp. var. formantes (médios-v.fem.)

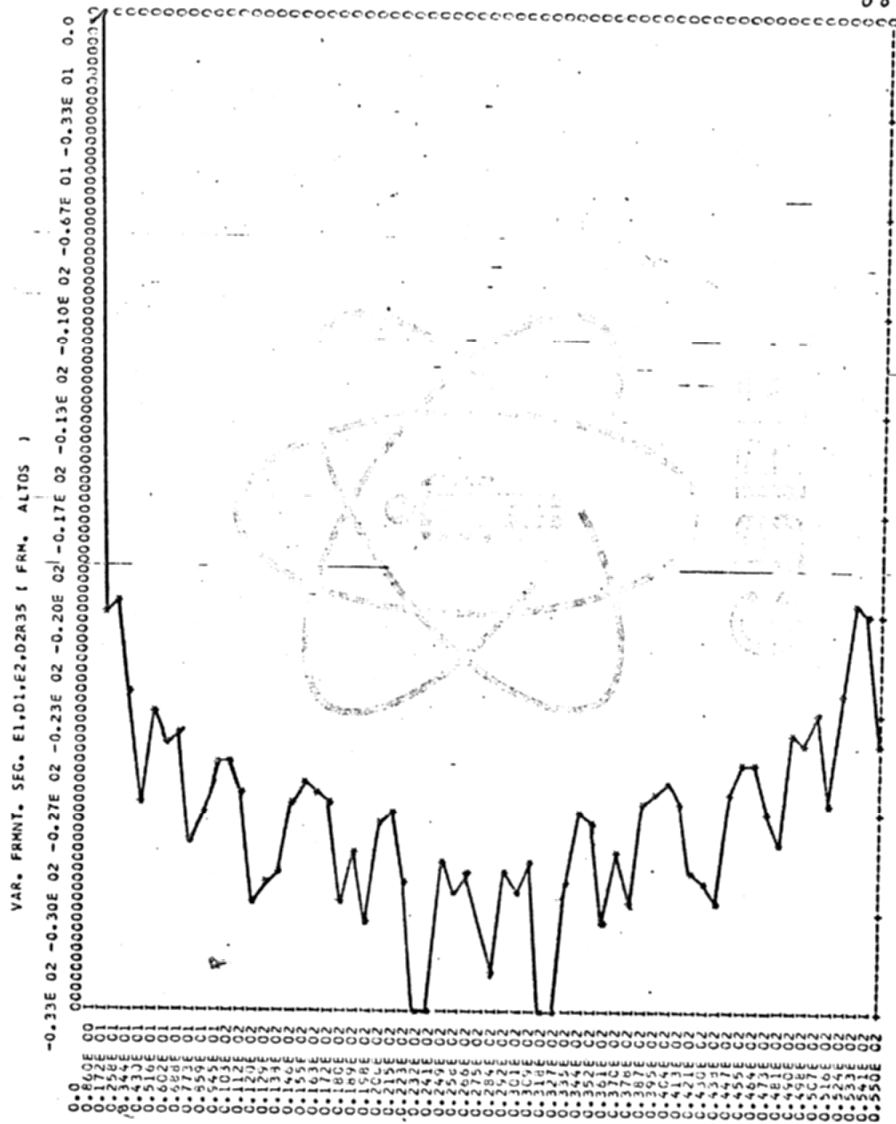


Figura 38 - Esp. var. formantes (altos-v.fem.)

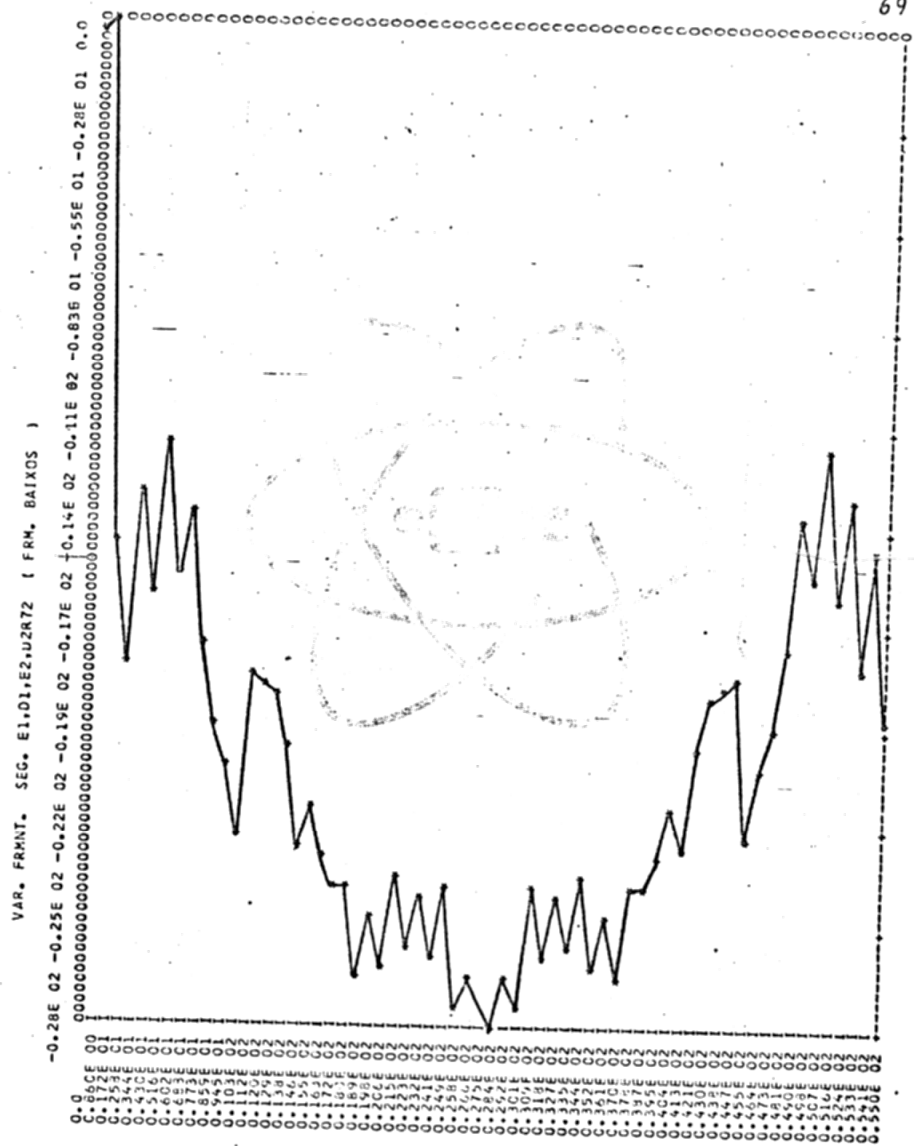


Figura 39 - Esp. var. formantes (baixos-v.masc.)

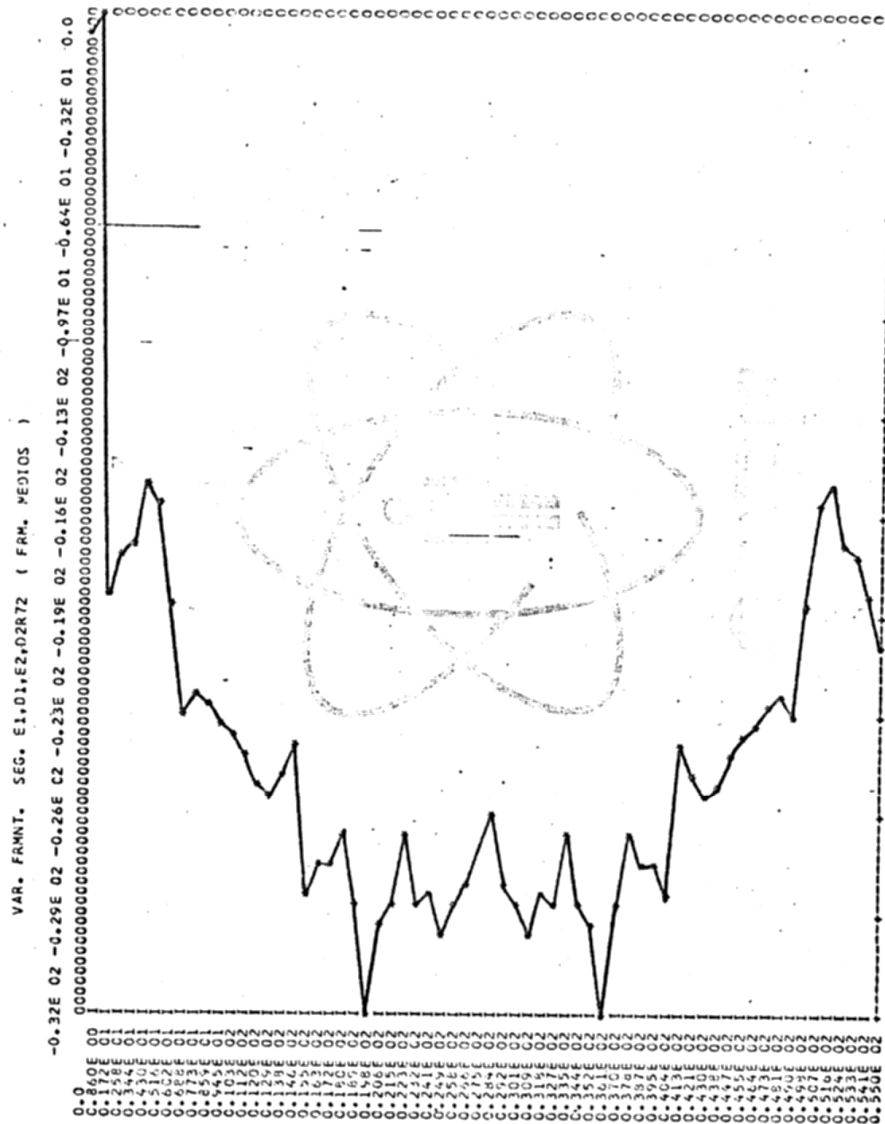


Figura 40 - Esp. var. formantes (médios-v.masc.)

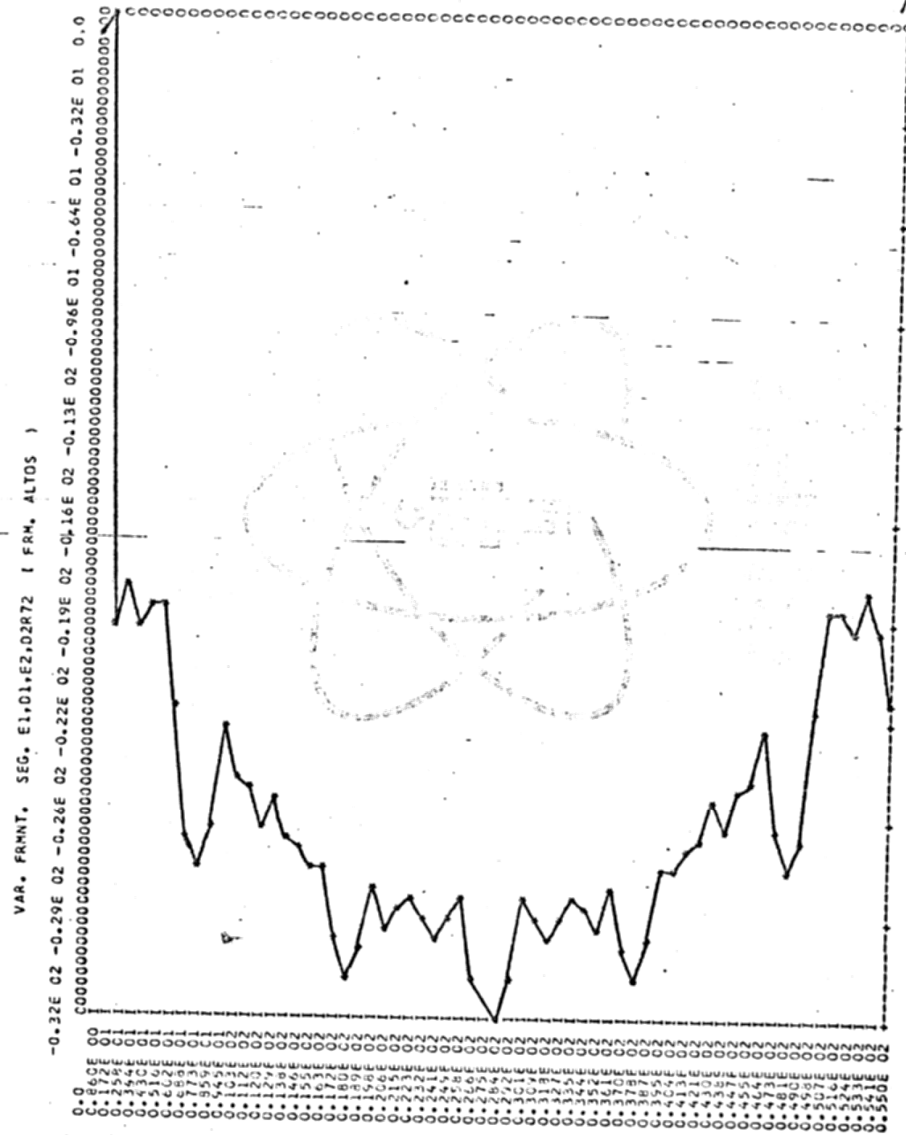


Figura 41 - Esp. var. formantes (altos-v.masc.)

frase dividida em segmentos iguais, isso é, com o mesmo número de amostras. Escolhendo-se adequadamente o comprimento do segmento (p. ex.:  $L = 10$  ms ou sejam  $N = 110$  amostras na taxa de amostragem, 11kHz, usada neste trabalho), pode-se estimar a ordem de grandeza e a taxa de variação dos pesos (coeficientes) de um preditor linear adequado à codificação digital da voz<sup>15</sup> em sistemas de codificação direta da forma de onda, tais como o DPCM, ... ADPCM etc. Em sistemas de codificação da voz usando predição adaptiva, os pesos são atualizados a cada ..... 10 — 30 ms<sup>15</sup>.

4.2.4- Domínio do tempo - Trabalhou-se, basicamente, com os programas EXAME, VARVA, FQREL, ZEROS e VARCO.

4.2.4.1- O programa VARVA foi sugerido pelo resultado apresentado em artigo versando sobre quantização adaptiva<sup>14</sup>, particularmente a curva de variação da amplitude do sinal de voz ao longo de uma frase (a pesquisa divulgada no artigo utilizou a frase alemã "DAS HERZ MUNCHEN").

4.2.4.2- Na referência 14, é apresentada curva correspondente à equação:

$$P_S(k) = 10 \log_{10} \left( \frac{1}{128} \sum_{j=1}^{128} x_j^2 / \sigma^2 \right) \quad (29)$$

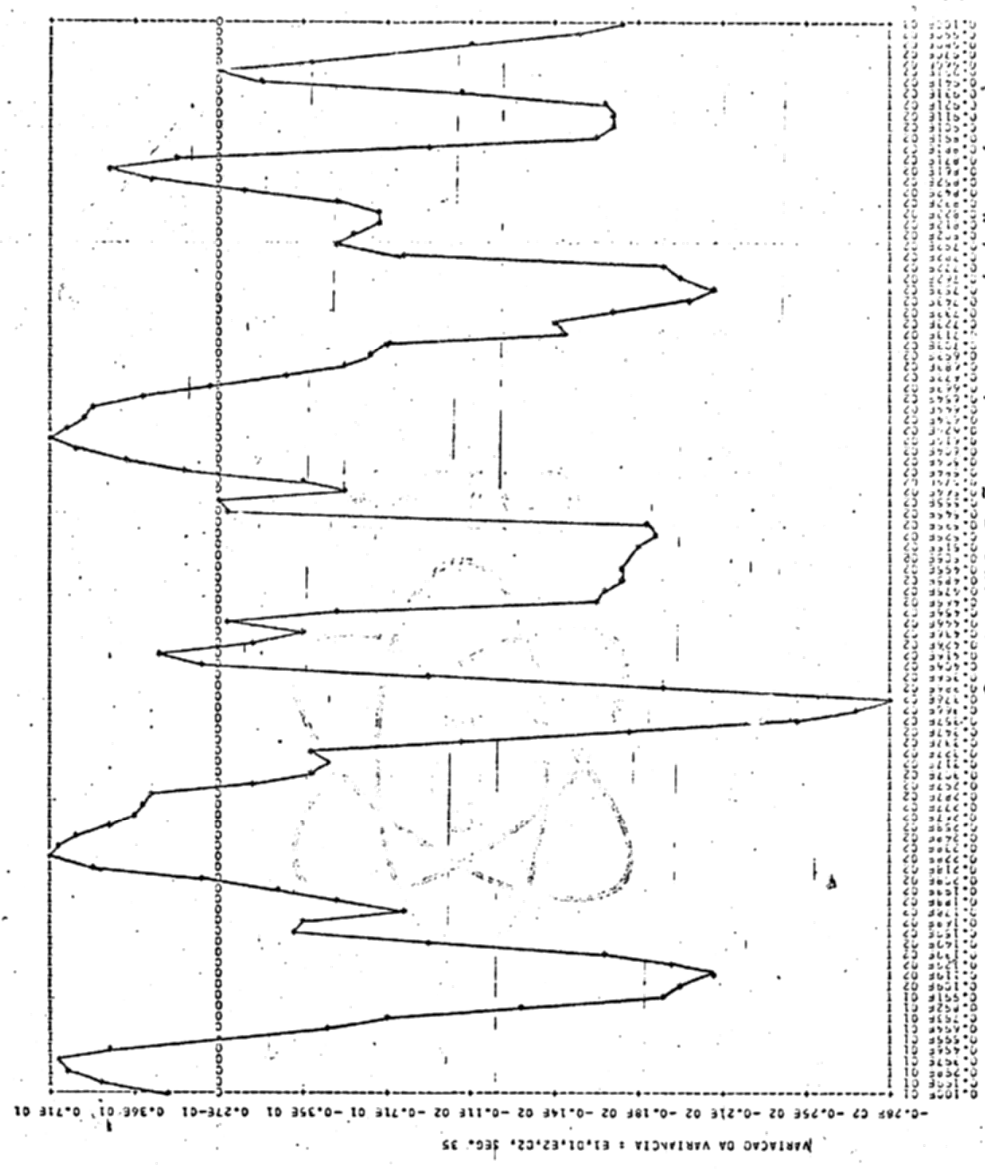


Figura 42 - Var. variância ("O Joao quebrou o dedo" - v.fem.)

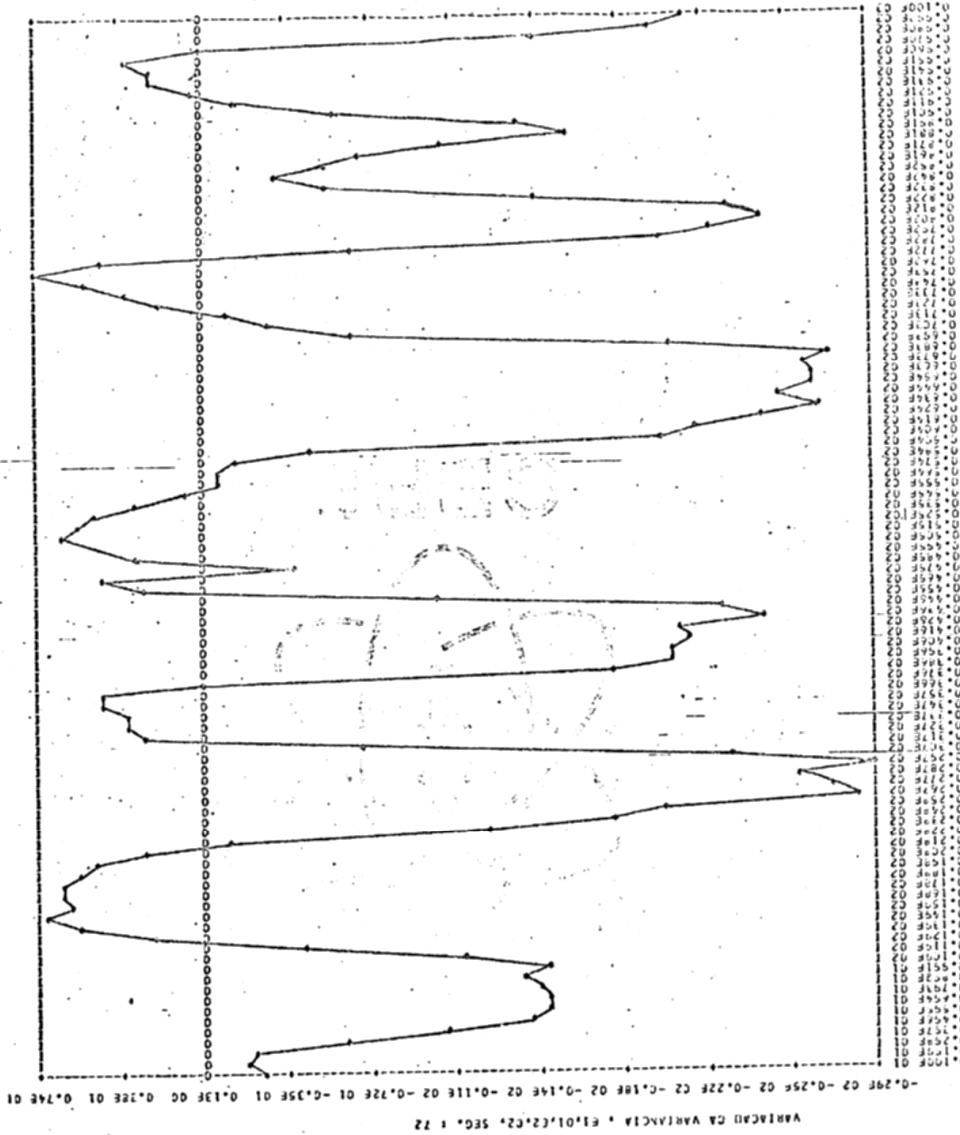


Figura 43 - Var. variância ("O João quebrou o dedo" - v.masc.)



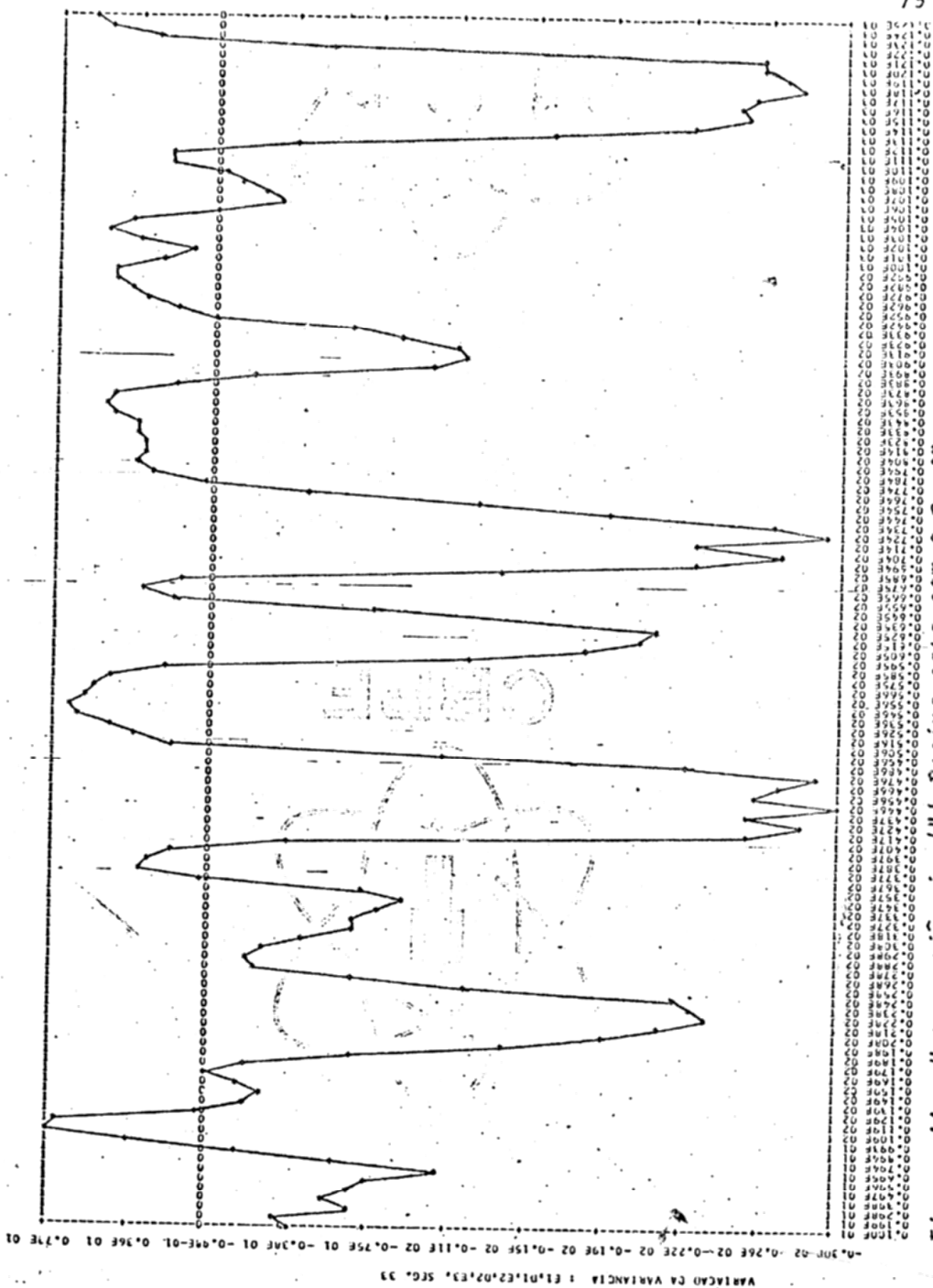


Figura 44 - Var. variância ("A Regina casa com o Raul" - v. fem.)

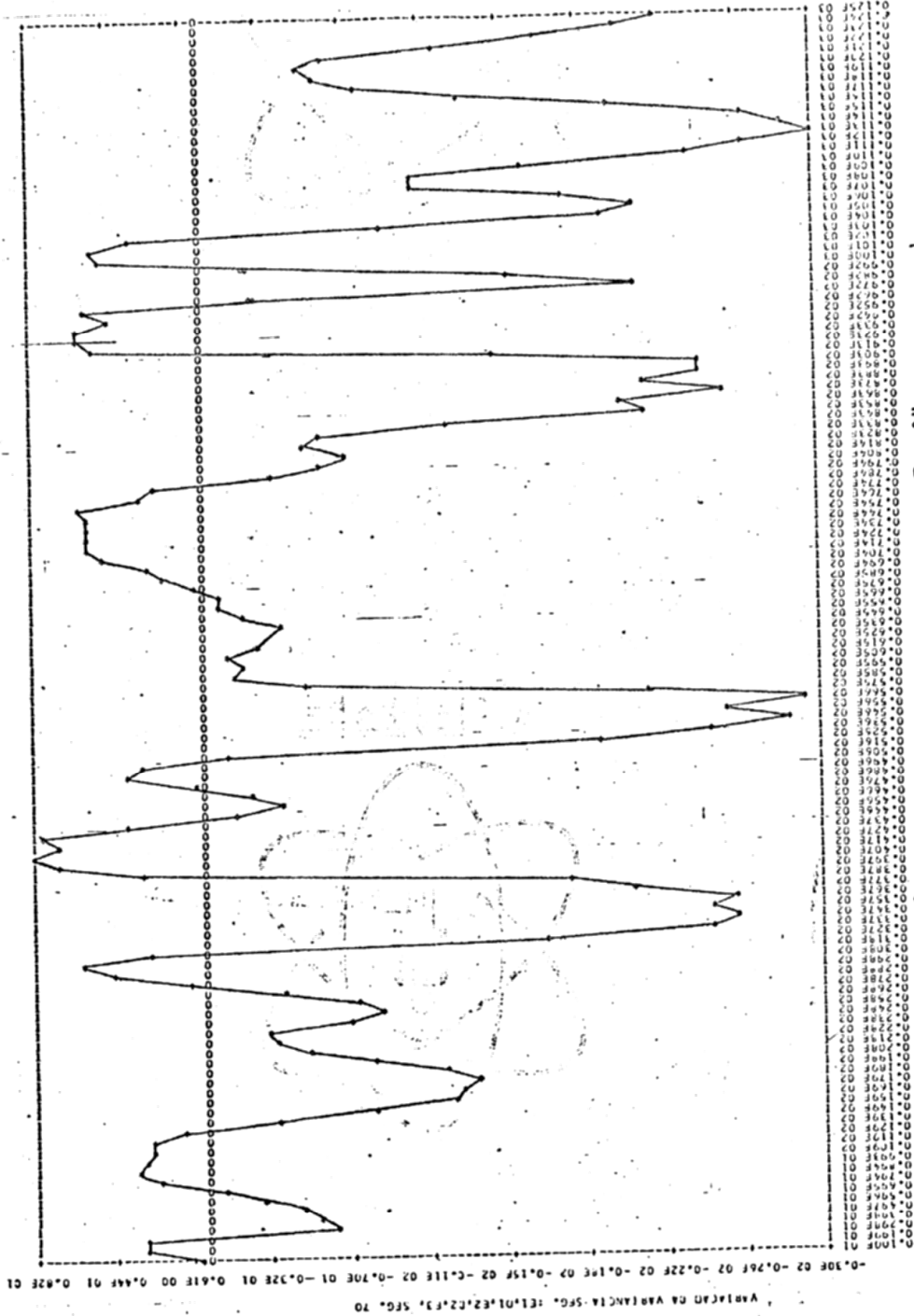


Figura 45 - Var. variância ("A Regina casa com o Raul" - v.masc.)

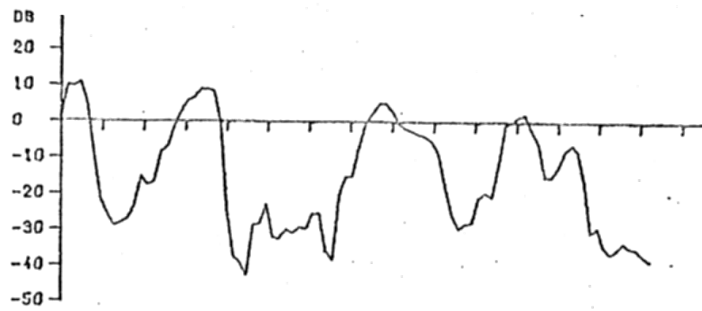


Figura 46 - Var. variância "Das Herz München<sup>14</sup>")

ou seja, a relação entre a variância amostral das amplitudes de um segmento e a variância geral ( $\sigma^2$ ) da frase dividida em segmentos iguais de 128 amostras cada um (16 ms).

4.2.4.3- O programa VARVA, com segmentos de 150 amostras, cada um (13,6 ms), foi aplicado às frases: "A Regina casa com o Raul" e "O João quebrou o dedo" (voz feminina e masculina) obtendo-se as curvas das Figuras 42, 43, 44 e 45.

4.2.4.4- Observando tais figuras verifica-se:

- são muito semelhantes ao diagrama incluso na referência<sup>14</sup> (ver figura 46);
- ao longo da frase, a variância permanece, na maior parte do tempo, inferior à variância geral; nas frases testadas registram-se as relações 67%, 64%, 72% e 69% chegando os níveis mínimos a -30 dB e os máximos a +8 dB. Tais levantamentos são importantes na escolha de algoritmos para a quantização adaptiva.

4.2.4.5- Várias referências (p. ex.: 20, 21) apresentam a densidade de probabilidade de Laplace

$$P_x(x) = (1/2 \sigma) \exp\left(-\frac{|x|}{\sigma}\right),$$

onde  $\sigma = \sqrt{2} B$ ,

como a mais adequada à caracterização estatística dos sinais de voz.

4.2.4.6- Embora seja discutível o aproveitamento da citada densidade no projeto de quantizadores ótimos, ela oferece, pelo menos, valioso subsídio aos trabalhos de simulação de sinais de voz.

4.2.4.7- O programa FQREL se destina à comparação da frequência relativa das amplitudes existentes em segmentos de frase (cada segmento com 360 ms, aproximadamente) com a f.d.p de Laplace. As curvas foram levantadas no intervalo  $[-4\sigma, +4\sigma]$ ; aplicando, simplesmente, a desigualdade de Chebyshev:

$$\text{Prob} \left[ |x - \mu| > k\sigma \right] < \frac{1}{k^2} \quad (31)$$

tem-se, para  $k = 4$ , 94% de probabilidade de o sinal estar dentro daquele intervalo.

4.2.4.8- As Figuras 47 e 48 representam algumas das muitas curvas levantadas, todas confirmando o seguinte:  
- a distribuição das amplitudes instantâneas do sinal de voz segue bem de perto a densidade de Laplace.

4.2.4.9- Uma das medidas aplicáveis na classificação de segmentos de voz em VOZEADOS, NÃO-VOZEADOS ou SI-

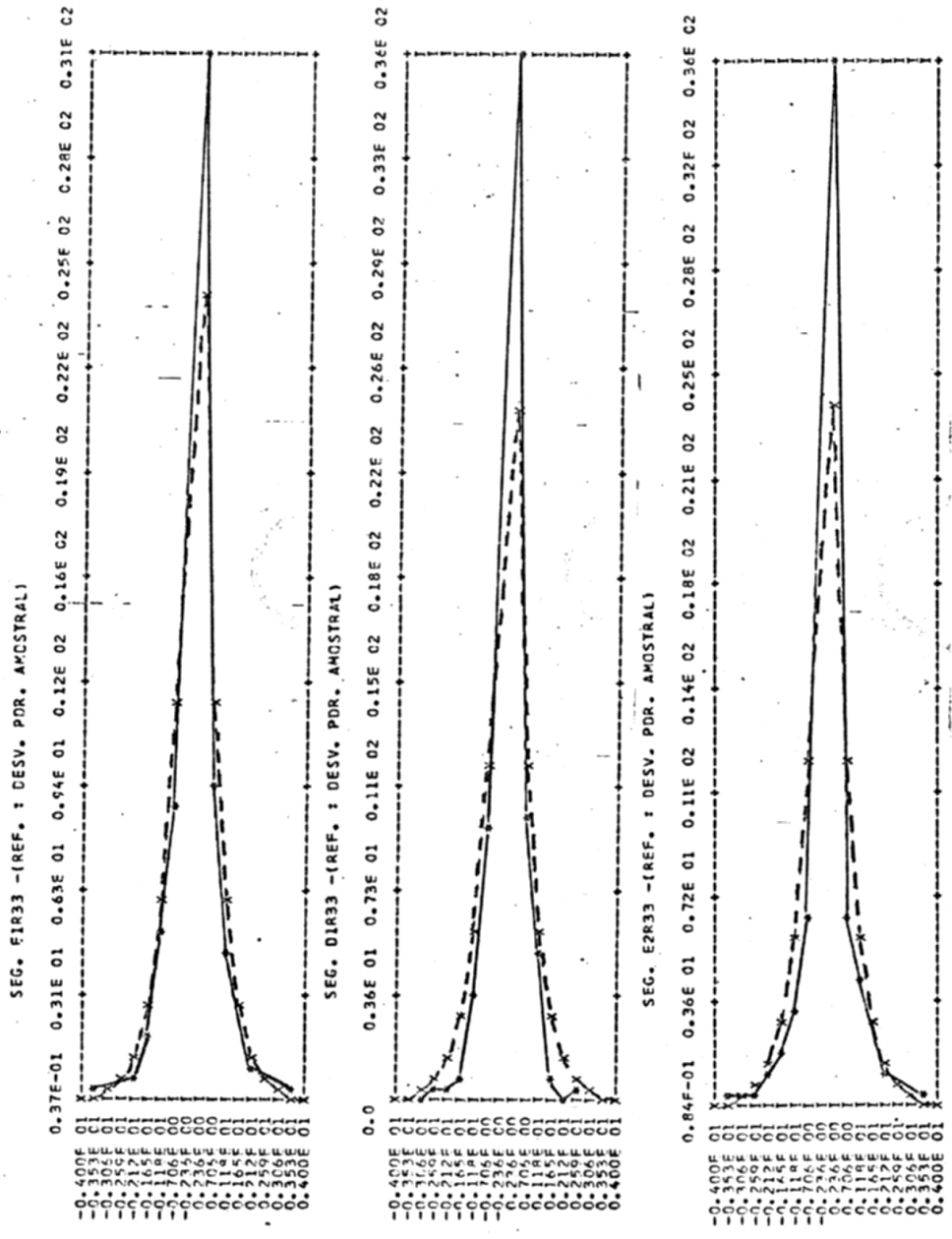


Figura 47 - Freq. relativa ("A Regina casa com o Raul" - v.fem. - três segmentos iniciais).

---X--- f.d.p. Laplace  
 —\*— freq. rel. sinal

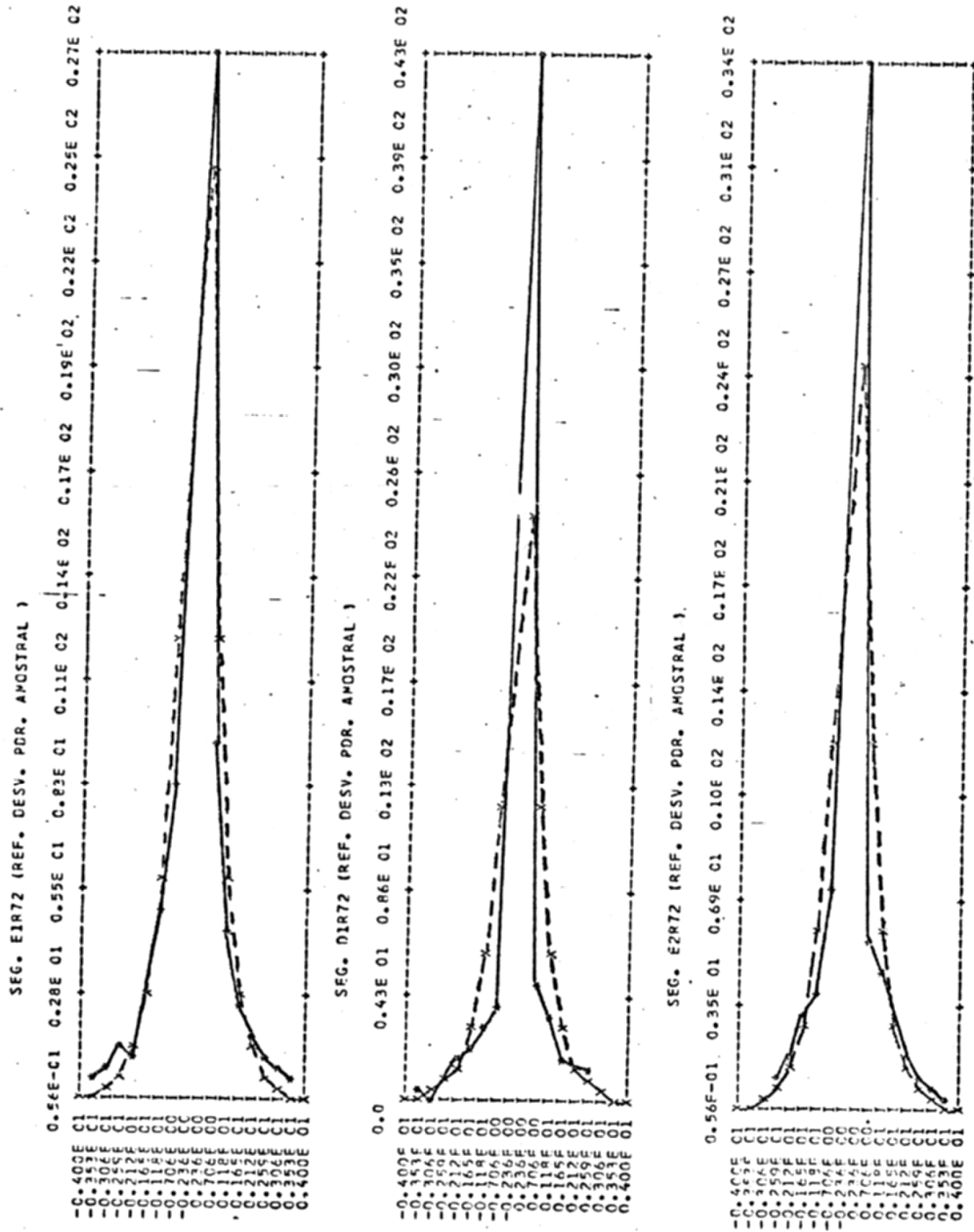


Figura 48 - Freq. relativa ("O João quebrou o dedo" - v.masc. - três segmentos iniciais).

---X--- f.d.p. Laplace  
 —\*— freq. rel. sinal

LÊNÇIO é a contagem dos cruzamentos do nível zero, isso é do número de vezes que o sinal corta o eixo dos tempos<sup>4,11</sup>. Essa contagem tem sido sugerida como algoritmo para sistemas precisando decidir sobre a presença da voz em meio ao ruído<sup>20</sup>.

4.2.4.10- O programa ZEROS executa essa contagem em segmentos consecutivos de 110 amostras (11 ms), analisando intervalos de frase de, aproximadamente, 340 ms cada um.

4.2.4.11- Aplicando o ZEROS às frases "O João quebrou o dedo" (voz feminina e voz masculina), "A Regina casa com o Raul" (id.) e "Alô, papai!" (id.) obtiveram-se os seguintes valores médios:

Tipo do Segmento	Cruzamentos de zero (valores médios/intervalo de 11 ms)	
	Voz feminina	Voz masculina
Vozeado	18	13
Não-vozeado	16	18
Silêncio	14	14

4.2.4.12- O artigo em referência 11 apresenta as seguintes médias para os cruzamentos de zero (sem especi<sup>f</sup>



car o tipo de voz: masculina ou feminina):

- vozeados: 12;
- não-vozeados: 50;
- silêncio: 25;

e admite distribuição Gaussiana para os três segmentos.

4.2.4.13- A grande disparidade entre as médias encontradas (16; 18) e a média citada (50) para os segmentos não-vozeados pode ser explicada pelos seguintes motivos:

- os fonemas não vozeados têm forma de onda semelhante a ruído e de baixa amplitude; sendo o número (256) de níveis de quantização relativamente pequeno, torna-se difícil delimitar com precisão onde se iniciam os segmentos não-vozeados;
- a população testada (3 frases para cada voz) provavelmente é insuficiente para obter-se a média real de cruzamentos de zero nos fonemas não-vozeados em frases faladas em português.

4.2.4.14- A quantização adaptiva permite ganhos consideráveis na relação SINAL/RUÍDO (pico) quando a correlação<sup>23</sup>

$$C_1 = \frac{E [X_n X_{n-1}]}{E [X_n^2]} \quad (32)$$

se aproxima de 1.

4.2.4.15- O programa VARCO, montado para traçar o gráfico da variação do valor do parâmetro  $C_1$ , foi aplicada a frase "O João quebrou o dedo" (voz masculina e voz feminina), dividida em segmentos iguais de 10 ms (110 amostras) cada um, obtendo-se as curvas das Figuras 49 e 50.

4.2.4.16- A observação das curvas mostra nitidamente que a correlação de retardo igual a 1 permanece, na maior parte do tempo, próxima de 1 revelando, assim, a forte correlação entre uma amostra e sua consecutiva. Esse resultado parece estar coerente com o obtido na aplicação do programa VARVA.

4.1.4.17- Uma consequência da forte correlação entre amostras consecutivas é a redução da variância da diferença entre as duas<sup>23</sup>. Assim:

$$D_n(1) = x_n - x_{n-1} \quad (33)$$

Virã:

$$\begin{aligned} E \left[ D_n^2(1) \right] &= E \left[ (x_n - x_{n-1})^2 \right] \\ &= 2 (1 - C_1) E \left[ x_n^2 \right] \end{aligned} \quad (34)$$

Sendo  $C_1$  superior a 0,5, obviamente a variância de  $D_n(1)$  será inferior à de  $x_n$ . Relações desse tipo constituem o princípio em que se baseiam os sistemas DPCM<sup>23</sup>.

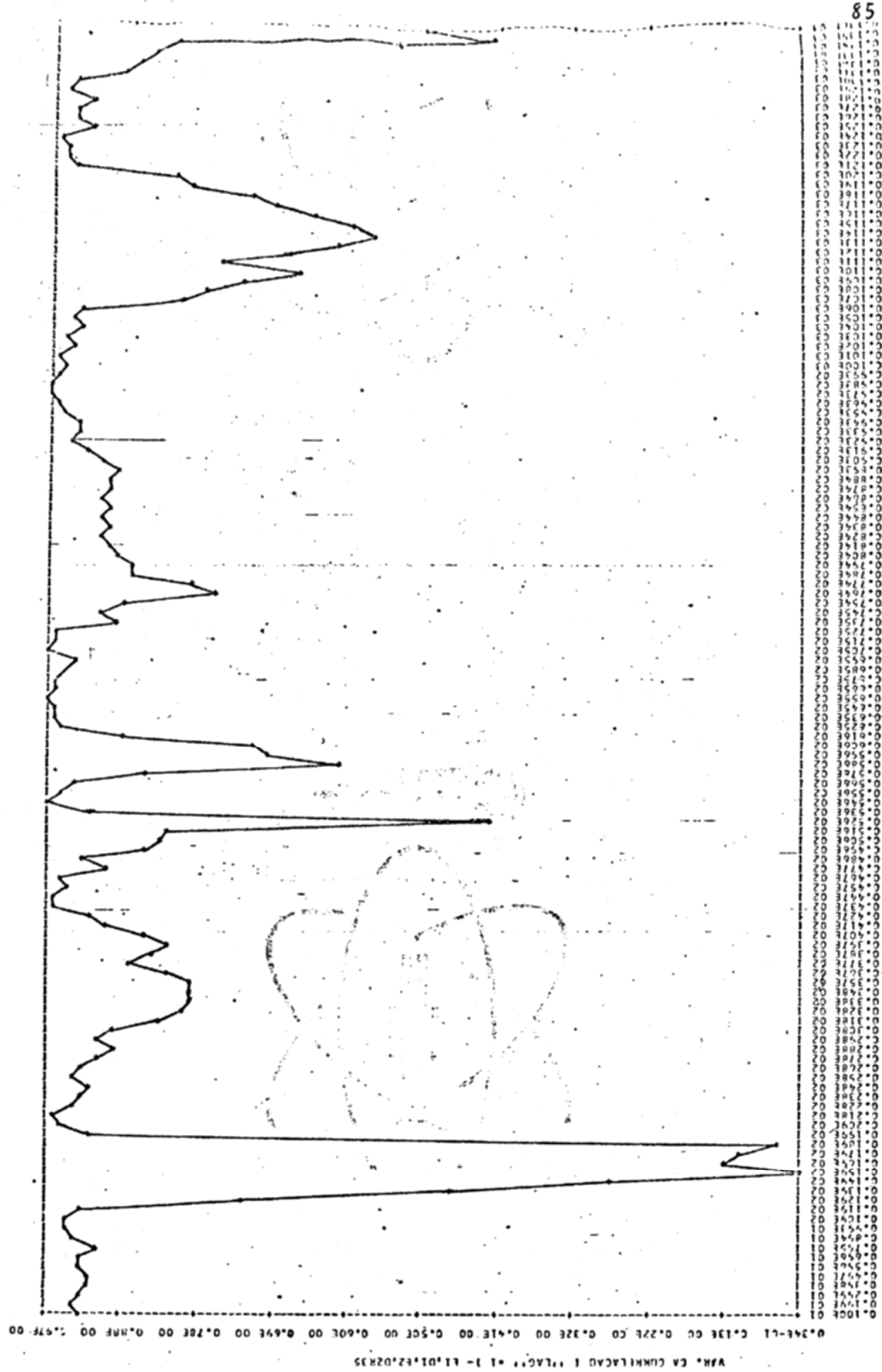


Figura 49 - Var. correlação, ret.=1 ("João quebrou o dedo" - v. fem.).

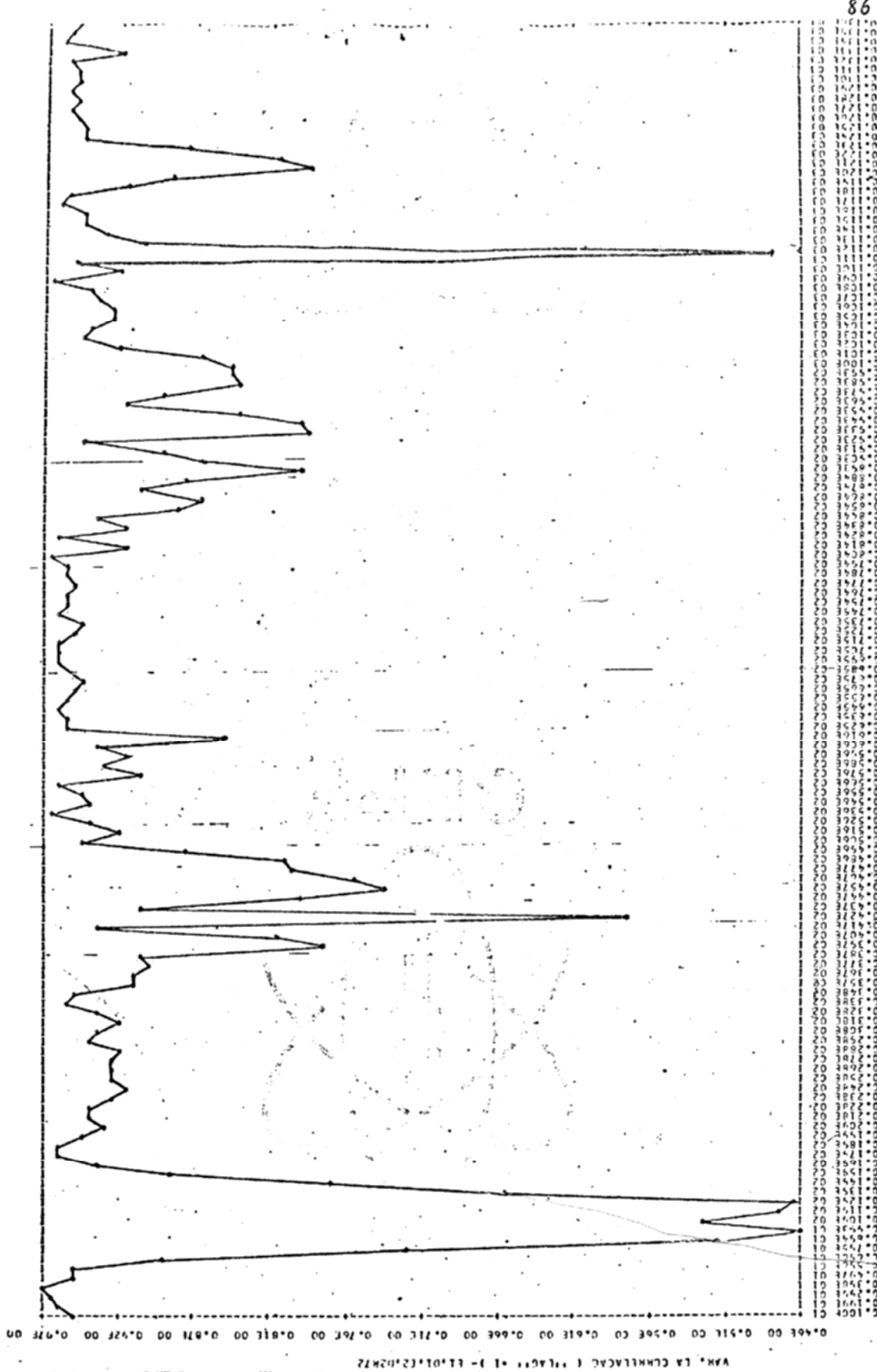


Figura 50 - Var. correlação, met. = 1 (1<sup>o</sup> João quebrou o dedo - v.masc.).

## V - CONCLUSÕES

5.1 - Em 1967, o Dr. J. R. Pierce, dos Laboratórios Bell, "previa que os computadores digitais logo se tornariam tão comuns nos laboratórios de pesquisa como os osciloscópios"<sup>8</sup>.

5.2 - O presente trabalho consistindo, apenas, na aplicação de algoritmos simples e programas escritos em FORTRAN acessível, mostrou realmente a grande potencialidade do computador como instrumento de pesquisa na área dos sinais de voz.

5.3 - Os resultados obtidos, em virtude do grande número de parâmetros e propriedades analisados, não permitem, ainda, afirmações categóricas tais como: "Os formantes característicos do fonema /ê/ feminino são:  $f_1$ ,  $f_2$  e  $f_3$ ". Para chegar a esse ponto, seriam necessários levantamentos extensos, utilizando população superior à utilizada na gravação (duas pessoas) e aumentando, também, o número de frases para teste de modo a garantir a ocorrência da real proporção dos diferentes tipos de fonemas da língua portuguesa, da real proporção dos cruzamentos de zero etc.

5.4 - Verificou-se, desde o início do trabalho, a im

portância de se dotarem os centros de pesquisa universitários com equipamento conversor análogo/digital, anexo ao computador existente. Sem aquele equipamento, qualquer processamento digital realizado em sinal real, isso é, não simulado, torna-se muito difícil, obrigando o pesquisador a se deslocar de sua universidade à procura de apóio alhures, com o rotineiro entrave administrativo e a desnecessária perda de tempo decorrente.

5.5 - Além das limitações citadas em 5.3, acresce o fato de ter sido utilizada, neste trabalho, codificação binária de 8 bits/amostra, ou seja um total de 256 níveis de quantização. Pesquisas realizadas em grandes centros estrangeiros costumam empregar 12 bits/amostra, ou sejam 4096 níveis<sup>11</sup>, aumentando, sem dúvida, a precisão dos levantamentos. Esse fator deverá ser levado em conta quando se equiparem os centros de processamento de dados com conversor análogo/digital, conforme sugerido em 5.4.

5.6 - Os levantamentos feitos no domínio da frequência evidenciaram a superioridade de técnicas paramétricas sobre técnicas não-paramétricas de análise espectral, particularmente na estimação do espectro dos sinais de voz. Com a técnica paramétrica consegue-se boa resolução espectral sem necessidade de aumentar demais o número de amostras; esse aumento é inconveniente na análise

de sinais estacionários a prazo curto, como é o caso dos sinais de voz. Além disso, é sempre atraente a possibilidade de usar expressão analítica no traçado do espectro de sinais aleatórios.

5.7 - Foi verificada a variação lenta dos formantes ao longo de uma frase, permitindo taxas de amostragem tão baixas quanto 60 Hz/100 Hz. A observação é válida para os formantes baixos, médios e altos, voz feminina ou masculina.

5.8 - A variação lenta dos formantes, naturalmente, corresponde a lenta variação dos parâmetros  $\{a_i\}$ . Esse fato levaria a cogitar-se transmitir  $\{a_i\}$  para efeito de síntese no lado da recepção. Em termos de circuito ... ("hardware"), teria de se pensar em estrutura digital capaz de armar a matriz  $R_y(k)$ , resolver o sistema (26), quantizar (com suficiente precisão) os parâmetros  $\{a_i\}$  e codificá-los em binário. A alternativa de transmitir diretamente os formantes exige o emprego de filtros ou algoritmos capazes de extrair informação referente às frequências.

5.9 - Verificou-se, também, a tendência geral de o sinal de voz permanecer com as amplitudes próximas da média; ou seja, a variância a prazo curto é quase sempre muito inferior à variância geral. O programa FQREL (com

paração da frequência relativa com a f.d.p. de Laplace) confirmou os resultados do programa VARVA (variação da variância). Resultados desse tipo podem ser úteis no projeto de quantizadores adaptivos, por exemplo: na escolha judiciosa dos multiplicadores do degrau de quantização<sup>23</sup>.

5-10- Verificou-se a possibilidade de identificação dos períodos de silêncio na frase através da contagem dos cruzamentos do zero. No caso mais geral, a média de cruzamentos para segmentos vozeados é inferior à dos segmentos de silêncio; e a média dos segmentos não-vozeados é bem maior que a média dos segmentos de silêncio. Além disso, embora a forma de onda dos segmentos não-vozeados seja semelhante à dos segmentos de silêncio, normalmente as amplitudes da primeira são maiores; o uso dessa característica poderá reduzir a probabilidade de uma decisão errada.

5.11- Se a correlação do retardo unitário normalizada permanece quase sempre elevada ou muitas vezes superior a 0,5, significa existir forte dependência entre sucessivas amostras do sinal de voz. Ocorrendo esse fato, a quantização adaptiva proporcionará considerável ganho na relação sinal/ruído.

5.12- Finalmente, sugerem-se para prosseguimento do trabalho as seguintes tarefas:



- (a) exame da possibilidade de formular o problema da identificação das pausas como problema de teste de hipóteses onde a observação seja o número de cruzamentos de zero. Lembre-se entretanto, que o projeto de sistemas identificadores não deve usar um único algoritmo<sup>11</sup>, caso contrário haverá sempre maior probabilidade de erro;
- (b) exame da possibilidade de realização de transmissão binária de voz usando métodos de análise-síntese, por exemplo: transmitindo identificação do segmento: vozeado, não-vozeado, silêncio e os parâmetros  $\{a_i\}$  correspondentes ao segmento.

#### REFERÊNCIAS

- 1 - A. Rosetti, Introdução à Fonética, Publicações Europa-América, (1962).
- 2 - Celso Cunha - Gramática da Língua Portuguesa, M E C (1972).
- 3 - J.N. Holmes, Speech Synthesis, Mills & Boon (1972).
- 4 - W.A. Ainsworth, Mechanisms of Speech Recognition, Pergamon Press, (1976).
- 5 - James L. Flanagan, The Synthesis of Speech, Sci. Am. - Fev. (1972).
- 6 - W.H. Saunders, The Larynx, Clinical Symposia da CIBA.
- 7 - J.L. Flanagan, C.H. Coker, L.R. Rabiner, R.W. Schafer, e N. Umeda, Synthetic Voices for Computers, IEEE - Spectrum, Out. (1970).
- 8 - James L. Flanagan, Focal Points in Speech Communication Research, IEEE - Transactions on Communications Technology - Dez. (1971).
- 9 - R.W. Schafer, L.R. Rabiner, Digital Representation of Speech Signals, Proc. IEEE - Abr. (1975).
- 10 - Peter D. Welch, The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method

- Based on Time Averaging Over Short Modified Periodograms, IEEE. Trans. Audio e Electroac. - Jun. (1967).
- 11 - B.S. Atal, L.R. Rabiner, A Pattern Recognition Approach to Voiced - Unvoiced - Silence Classification with Applications to Speech Recognition, IEEE - Trans. on Ac., Speech, Sig. Proc., ASSP-24, 201 a 204, (1976).
- 12 - Fumitada Itakura, S. Saito, A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies, Electronics and Communication in Japan, Vol. 53-A - No 1, (1970).
- 13 - M. Schwartz, L. Shaw, Signal Processing, Mc-Graw Hill, (1975).
- 14 - Peter Noll, Adaptive Quantizing in Speech Coding Systems, Proc. IEEE Zurich Seminar on Digital Commun. (1974).
- 15 - J.W. Bayless, S.J. Campanella, A.J. Goldberg, Voice Signals: bit-by-bit, IEEE - Spectrum, Out., (1973).
- 16 - H. Taub, D.L. Schilling, Principles of Communication Systems, Mc-Graw Hill, (1971).
- 17 - G.M. Jenkins, D.G. Watts, Spectral Analysis and its Applications, Holden-Day, (1968).

- 18 - L.R. Rabiner, B. Gold, Theory and Application of Digital Signal Processing, Prentice Hall, (1975).
- 19 - S.A. Treffer, Introduction to Discrete-Time Signal Processing, J. Wiley & Sons, (1976).
- 20 - J.A. Jankowski, Jr., A New Digital Voice-Activated Switch, COMSAT Technical Review - Vol. 6, Nº 1, 160, (1976).
- 21 - M.R. Sambur, An Efficient Linear Prediction-Vocoder, B.S.T.J., Vol. 54 - Nº 10, 1700, 1701, (1975).
- 22 - J. Pines, O. Barradas, Sistemas Multiplex, Livros Técnicos e Científicos Editora S.A., (1977).
- 23 - N.S. Jayant, Digital Coding of Speech Waveforms: PCM, DPCM and DM Quantizers, Proc. IEEE, Vol. 62 - Nº 5, 617, (1974).

```

SUBROUTINE FFT4(AVC,M,N,MODE)
COMPLEX AVC(N),U,W,T
C
C FFT- ALGORITMO CITADO POR RABINER E GOLD
C NO 1 (VRO " THEORY AND APPLICATION OF
C DIGITAL SIGNAL PROCESSING "
C
N=2**M
NV2=N/2
NM1=N-1
J=1
DO 7 I=1,NM1
IF(I .GE. J) GO TO 5
T=AVC(J)
AVC(J)=AVC(I)
AVC(I)=T
5 K=NV2
6 IF(K .GE. J) GO TO 7
J=J-K
K=K/2
GO TO 6
7 J=J+K
PI=4.*ATAN(1.0)
DO 20 L=1,M
LE=2**L
LE1=LE/2
U=(1.0,0.0)
IF(MODE .GT. 1) GO TO 8
W=CMPLX(COS(PI/LE1),-SIN(PI/LE1))
GO TO 9
8 W=CMPLX(COS(PI/LE1),SIN(PI/LE1))
9 DO 20 J=1,LE1
DO 10 I=J,N,LE
IP=I+LE1
T=AVC(IP)*U
AVC(IP)=AVC(I)-T
10 AVC(I)=AVC(I)+T
20 U=U*W
IF(MODE .LT. 2) GO TO 40
DO 30 I=1,N
30 AVC(I)=AVC(I)/N
40 RETURN
END

```



CBE

```

DIMENSION AV(7500), IAUX(7500), ISIMB(1)
DIMENSION AMP(128), SUM(128), A(54,2), ERZ(64)
COMPLEX AVC(128)
DATA LCART/5/

C
C
C   PROG. BASFACD NO METODO SUGERIDO POR P.O. WELCH, NA
C   REV. "IEEE TRANS. ON AUDIO AND ELECTROAC.",
C   VOL. AU-15, NR. 2 ,JUN. 1967
C
C
C   PREPARO DOS DADOS
C
C   KRPT=0
C   CALL BASE2 (KRPT,METAD,NRREG,NRSEG,IAUX,AV)
C   READ(LCART,777) KAMSI
777 FORMAT(I4)
C
C   DADOS PARA CALCULO DA FFT
C   M=7
C   N=128
C   TS=1.0/11000.0
C   PI= 4.0* ATAN(1.0)
C   MODO =1
C
C   REMANEJAR AS AMOSTRAS
C
C   INIC=KAMSI-100
C   IFIM=KAMSI+511
C   DO 39 I=INIC,IFIM
C   K=I-INIC+1
39 AV(K)=AV(I)
C
C
C   FILTRAR
C
C   JSEG=512
C   CALL FILFL (JSEG,AV)
C   CALCULO DO PERIODOGRAMA
C
C   J=-63
C   DO 35 I=1,128
35 SUM(I)=0.0
C   DO 80 I1=1,7
C   J=J+64
C   K= J
C   DO 70 L=1,128
C   FL = FLOAT(L)
C   ARSAV=ABS(AV(K))
C   IF(ARSAV .GT. 1.E-15) GO TO 65
C   AV(K)= 0.000
65 AVW=AV(K)*WIN(FL)
C   AVC(L)=CMPLX(AVW,0.0)
70 K=K+1
C   CALL FFT4(AVC,M,N,MODO)
C   DO 75 J=1,128
75 AMP(J)=(REAL(AVC(J)))**2+(AIMAG(AVC(J)))**2
C   DO 80 K=1,128
80 SUM(K)= SUM(K)+ AMP(K)
C
C
C   CALCULO DA CONSTANTE

```

```
C      SOMA= 0.00
      DO 81 I=1,128
      FL= FLOAT(I)
81     SOMA= SOMA + ( WIN(FL) )**2
      CTE=129.0/(7.0*SOMA)
C      TERMINO DO PERICODOGRAMA
      DO 82 I=1,64
      AMP(I)=CTE*SUM(I)
82     ERZ(I)=AMP(I)

C
C      PREPARO CAS COORDENADAS
C
C      PREPARO DAS ABSCISSAS
      FN= FLOAT(N-1)
      A(1,1)= 0.0
      ACRS= (1.0/TS)/FN
      ABC= 0.00
      DO 83 J=2,64
      ABC = ABC + ACRS
83     A(J,1)= ABC

C
C      PREPARO DAS ORDENADAS
C
84     K=0
      I=0
85     I=I+1
      IF(ERZ(I).LE.ERZ(I+1)) GO TO 85
      TEMP=ERZ(I)
      ERZ(I)=ERZ(I+1)
      ERZ(I+1)=TEMP
      K=1
86     IF(I.LT.63) GO TO 85
      IF(K.NE.0) GO TO 84
      DO 87 J=1,64
87     A(J,2)=10.0*(ALOG(AMP(J)/ERZ(64))/ALOG(10.0))

C
C      GRAFICO
      ISIMB(1)=10
      CALL GRAFG(A,1,1,64,64,2,2,1,1,ISIMB,2,1)
      STOP
      END
```

```

DIMENSION ISIMB(1), AV(7500), IAUX(7500)
DIMENSION SMO(128), A(128,2), AMP(128), ERZ(128)
COMPLEX AVC(256)
DATA LCART/5/

C
C   PROGRAMA 'ALISA'
C   ALISAMENTO SIMPLES DO PERIODOGRAMA
C
KRPT=0
CALL BASE2 (KRPT,METAD,NRREG,NRSEG,IAUX,AV)
READ(LCART,777) KAMSI
777 FORMAT(I4)
C   DADOS PARA CALCULO DA FFT
M=8
N=256
TS=1.0/11000.0
MODD=1

C
C   REMANEJAR AS AMOSTRAS
C
INIC= KAMSI-100
IFIM= KAMSI+ 255
DO 39 I= INIC,IFIM
K= I-INIC +1
39 AV(K)= AV(I)

C
C   FILTRAR
C
JSFG = 256
CALL FILFL(JSFG,AV)

C
C   CALCULO DO PERIODOGRAMA
C
DO 71 I=1,256
ABSAV =ABS(AV(I))
IF(ABSAV .GT. 1.E-15) GO TO 71
AV(I)= 0.000
71 AVC(I)= CMPLX(AV(I) ,0.0)
CALL FFT4(AVC,M,N,MODD)
DO 93 I=1,128
93 AMP(I)={REAL(AVC(I))**2 + (AIMAG(AVC(I)))**2}
J=3
SMO(I)={AMP(I)+AMP(2)+AMP(3)}/5.0
SMO(2)=SMO(1)+(AMP(4))/5.0
DO 99 I=3,126
SMO(I)={AMP(J-2)+AMP(J-1)+AMP(J)+AMP(J+1)+AMP(J+2)}/5.0
J=J+1
99 CONTINUE
SMO(127) = SMO(126) - (AMP(124))/5.0
SMO(128) = SMO(127) - (AMP(125))/5.0
DO 101 I=1,128
101 ERZ(I)= SMO(I)

C
C   ORDENACAO DE ERZ
C
104 K= 0
I= 0
105 I= I+1
IF(ERZ(I) .LE. ERZ(I+1)) GO TO 106.
TEMP = ERZ(I)

```



```
ERZ(I)= ERZ(I+1)
ERZ(I+1)= TEMP
K=1
106 IF(I .LT. 127) GO TO 105
    IF(K .NE. 0) GO TO 104
C
C   PREPARO CAS COORDENADAS
C
DO 107 I=1,128
107 A(I,2)= 10.0 *(ALOG( SMO(I)/ERZ(128))/ALOG(10.0))
    A(I,1)= 0.00
    ACRS =(1./ TS)/255.0
    DO 108 I= 2,128
108 A(I,1) = ACRS* FLOAT(I-1)
C
C   GRAFICO
C
ISIMB(1)=10
CALL GRAFG(A,1,1,128,128,2,7,1,1,ISIMB, 1)
STOP
END
```

```
DIMENSION AA(81,2),F(14),AM(14,14),R(14),ALFA(15),VC(15),
* BW(14),ROOTI(14),ROOTR(14),X(81),ACHAP(15),ENV(81),ERZ(81)
DIMENSION ISIMB(1), KAMSI(18), GAMA(15)
DIMENSION AV(7500), IAUX(7500)
DATA IMPR/6/
DATA LCART/5/
C
C
C   PROGRAMA 'FERMNT'
C   DETERMINACAO DOS FORMANTES
C   PROGRAMA BASEADO NO ARTIGO DOS ENG. F. ITAKURA E S. SAITO,
C   NA REV. 'ELECT. AND COMM. IN JAPAN' - VOL.53, NR.1, 1970
C
C   PREPARO DOS DADOS
C
C   KRPT=0
C   CALL BASE2 (KRPT,METAD,NRREG,NRSEG,IAUX,AV)
C
C   CONSTANTES
C   PI=4.0*ATAN(1.0)
C   TS=1.0/11000.0
C   HZPRD=2.0*PI*TS
C
C   GERACAO DO SINAL DE TESTE
C
C   JSEG REPRESENTA O NUMERO DE AMOSTRAS DO SEGMENTO
C   NP REPRESENTA O GRAU DA EQ. DAS FREQUENCIAS
C   NRGRF REPRESENTA O NUMERO DE GRAFICOS
C   READ(LCART,110) JSEG,NP,NRGRF
110 FORMAT(3I3)
   READ(LCART,330) (KAMSI(KR),KR=1,NRGRF)
330 FORMAT(7I4)
   DO 2000 KR=1,NRGRF
   LCNT=0
65   INIC= KAMSI(KR)-100
   IFIM=KAMSI(KR)+JSEG-1
   DO 67 I=INIC,IFIM
   K=I-INIC+1
67   AV(K)=AV(I)
C
C   FILFL E' UM FILTRO DE FASE LINEAR ('FIR')
C   FREQ. CORTE =4 KHZ
C
C   CALL FILFL (JSEG,AV)
C
C   GERACAO DOS V-CHAPEU
C
C   LFIM=NP+1
C   DO 5 J=1,LFIM
C   CALL AUTOC(J,AV,0,JSEG)
5   VC(J)=0
C
C   GERACAO DA MATRIZ 'AM'
C   ( A MATRIZ 'AM' E' SIMETRICA )
C
C   DO 10 J=1,NP
C   DO 10 K=1,NP
C   IF(J-K)7,8,9
7   L=K-J+1
   AM(J,K)=VC(L)
```

```

      GO TO 10
      8 AM(J,K)=VC(I)
      GO TO 10
      9 L=J-K+1
      AM(J,K)=VC(L)
10 CONTINUE
C
C   GERACAO DO VETOR ' B '
C
      DO 70 K=1, NP
      L=K+1
70 B(K)=VC(L)
C
C   CALCULO DAS RAIZES
C   SENDO 'AM' SIMETRICA, NAO E' PRECISO ARRUMAR O 'ARRAY'
C
      IORD=NP
      JA=IORD+1
      CALL SIMO(AM, B, NP, KS)
      IF(KS)1020, 71, 1020
C
C   GERACAO DOS COEFICIENTES DO POLINOMIO
C
71 ALFA(1)=1.000
   GAMA(1)=1.000
   DO 75 K=2, JA
   L=K-1
   ALFA(K)=B(L)
75 GAMA(K)=ALFA(K)
C
C   CALCULO DAS RAIZES
C
      CALL RAIZP(GAMA, W, IORD, ROOTR, ROOTI, IER)
      IER=IER+1
      GO TO (82, 2000, 2000, 81, 2000), IER
81 WRITE(IMP, 1700) KR
   WRITE(IMP, 1800) (ALFA(J), J=1, JA)
   GO TO 2000
C
C   GERACAO DOS ' A-CHAPEU '
82 DO 85 I=1, JA
   SOMA=0.000
   J=1
   L=I
83 SOMA=SOMA+ALFA(J)*ALFA(L)
   IF(L-JA) 84, 85, 85
84 J=J+1
   L=L+1
   GO TO 83
85 ACHAP(I)=SOMA
C
C   CALCULO DE SIGMA-2 'CHAPEU'
   SIGM2=ACHAP(1)*VC(1)
   DO 86 J=2, JA
86 SIGM2=SIGM2+2.0*ACHAP(J)*VC(J)
C
C   CALCULO DA ENVOLTORIA ESPECTRAL
C
   SNUM=SIGM2/(2.0*PI)
   SOMA=ALFA(1)
   DO 87 I=2, JA

```

```

87 SOMA=SOMA-ALFA(I)
DEN=SOMA**2
ENV(I)=1./DEN)*SNUM
IF( ENV(I) .GT. 1.E-61 ) GO TO 88
ENV(I)=1.E-61
88 ERZ(I)=ENV(I)
DO 90 I=2,31
PTR=ALFA(I)
PTI=0.0
ARC=( FLOAT(I)-1.0 )*50.*HZPRD
DO 89 K=2,JA
BETA=( FLOAT(K)-1.0 )*ARC
PTR=PTR-ALFA(K)*COS(BETA)
89 PTI=PTI+ALFA(K)*SIN(BETA)
DEN=PTR**2+PTI**2
ENV(I)=( 1./DEN )*SNUM
IF( ENV(I) .GT. 1.E-61 ) GO TO 90
ENV(I)=1.E-61
90 ERZ(I)=ENV(I)
C
C
C   ORDENACAO DA VARIAVEL AUXILIAR ERZ
C
91 K=0
I=0
93 I=I+1
IF(ERZ(I).LE.ERZ(I+1)) GO TO 95
TEMP=ERZ(I)
ERZ(I)=ERZ(I+1)
ERZ(I+1)=TEMP
K=I
95 IF( I .LT. 80 ) GO TO 93
IF(K.NE.0) GO TO 91
C
C
C   CALCULO DAS FREQUENCIAS
C
DO 99 J=1,NP
REL=ROOTI(J)/ROOTR(J)
TETA=ATAN(REL)
IF( ROOTR(J) .GE. 0.0 ) GO TO 98
IF( ROOTI(J) .GE. 0.0 ) GO TO 97
TETA=TETA-PI
GO TO 98
97 TETA=TETA+PI
98 F(J)= TETA/HZPRD
QM=(ROOTI(J))**2+(ROOTR(J))**2
VM=SQRT(QM)
ZH=ALOG(VM)
99 BW(J)=ABS(ZH)/(PI*TS)
C
C
C   REGISTRAR O SEGMENTO
C
WRITE(IMPR,1200) NP
WRITE(IMPR,853) JSEG
WRITE(IMPR,1525) METAD,NRREG,NRSEG
WRITE(IMPR,1575) KAMSI(KR)
WRITE(IMPR,1949) ERZ(31)
C
C
C   ESCREVER AS FREQUENCIAS E FAIXAS PASSANTES
C
WRITE(IMPR,1550)

```



CBF

```

WRITE(IMPR,1600) (J,F(J),BW(J),J=1,NP)
C
C   DESENHO DA ENVÓLTORIA
C
X(1)=0.00
DO 100 J=1,81
XM=FLOAT(J-1)
100 X(J)= XM*50.0
DO, 104 J=1,81
AA(J,1)=X(J)
104 AA(J,2)=10.0*(ALOG(ENV(J)/ ERZ(81))/ALOG(10.0))
C
C
ISIMB(1)=10
CALL GRAFG(AA,1,1,81,81,2,2,1,1,ISIMB,2,1)
GO TO 2000
1020 KAMSI(KR)=KAMSI(KR)+100
LCONT=LCONT+1
IF(LCONT-3) 65,65,1932
1932 LACRS=LCONT-1
WRITE(IMPR,1937) LACRS
2000 CONTINUE
853 FORMAT(36X,'QUANTIDADE DE AMOSTRAS = ',I4,///)
1200 FORMAT(36X,'GRAU DA EQUACAO = ',I2,///)
1525 FORMAT(36X,'METADE',I2,1X,',',',',REG.',1X,I2,1X,'SEG.',1X,I2,///)
1575 FORMAT(36X,'AMOSTRA INICIAL : K=',2X,I4,///)
1550 FORMAT(36X,'POLOS',14X,'FAIXA PASSANTE',//)
1600 FORMAT(30X,'F(',I2,')',1X,'=',E10.3,6X,'BW=',E10.3//)
1700 FORMAT(15X,' ULTRAPASSADAS 2000 ITERACOES NO SEG. NR.',I2,///)
1800 FORMAT(10X,F12.5)
1937 FORMAT(16X,' TENTADOS ',1X,I4,1X,'ACRESCIMOS EM KAMSI',//)
1948 FORMAT(36X,'NIVEL DE REFERENCIA =',2X,E11.4,///)
STOP
END

```

```

C
C
C
C
DIMENSION AV(7500), IAUX(7500), A(34,2)
DIMENSION ISIMB(1)

CONTAGEM DOS CRUZAMENTOS DO NIVEL ZERO
INTERVALOS DE 10 MILISEGUNDOS

KRPT=0
CALL BASE2 (KRPT,METAD,NRREG,NRSEG,IAUX,AV)
NPN=34
JSEG=110
DO 90 K=1,NPN
NZERO=0
INIC=1+(K-1)*JSEG
IFIM=K*JSEG-1
DO 70 I=INIC,IFIM
PROD=AV(I)*AV(I+1)
IF (PROD.GE.0.0) GO TO 70
NZERC=NZERO+1
70 CONTINUE
A(K,2)=FLOAT(NZERO)
90 A(K,1)=FLOAT(K)
ISIMB(1)=10
CALL GRAFG(A,1,1,NPN,NPN,2,2,1,1,ISIMB,2,2)
STOP
END
```

```

DIMENSION AV(7500), I AUX(7500),
DIMENSION SIG2N(50), A(50,2)
DIMENSION ISIMB(1)
DATA LCART/5/

C
C   PROGRAMA ** VARVA **
C   VARIACAO DA VARIANCIA
C
C   INTERVALOS DE 13,6 MILISEGUNDOS
C
C   SIG2G E A VARIANCIA GERAL ( FRASE INTEIRA )
C
191 READ(LCART,181) SIG2G
   FORMAT(E10.3)
   IG=0
   KRPT=0
   DO 706 KI=1,2
   CALL BASE2(KRPT,METAD,NRREG,NRSEG, I AUX, AV)

C
C   CALCULO DAS VARIANCIAS NORMALIZADAS
C
   KIN=-149
   KFI=0
   DO 705 I=1,25
   SOMA=0.0
   KIN=KIN+150
   KFI=KFI+150
   DO 703 J=KIN,KFI
703  SOMA=SOMA+(AV(J))**2
      AMQ=SOMA/150.
      SOMA=0.0
   DO 704 J=KIN,KFI
704  SOMA=SOMA+AV(J)
      AMQ=AMQ+AV(J)
      QMD=(AMQ)**2
      IG=IG+1
705  SIG2N(IG)=(AMQ-QMD)/SIG2G
706  CONTINUE

C
C   PREPARO DAS ABCISSAS
C
   DO 707 I=1,50
707  A(I,1)=FLOAT(I)

C
C   PREPARO DAS ORDENADAS
C
   DO 708 I=1,50
708  A(I,2)=10.0*(ALOG(SIG2N(I))/ALOG(10.0))

C
C   GRAFICO
C
   ISIMB(1)=10
   CALL GRAFG(A,1,1,50,50,2,2,1,1, ISIMB,2,1)
   STOP
   END

```

```
DIMENSION COEF(136), A(136,2), ISIMB(1)
DIMENSION AV(7500), IAUX(7500)
DATA LCART/5/
```

```
C
C
C
C
C
C
```

```
PROGRAMA 'VARCO'
VARIACAO DA COPRELACAO DE RET.=1, NORMALIZADA
INTERVALOS DE 10 MILISEGUNDOS
```

```
ISIMB(1)=10
KRPT=0
NP=0
LSEG=110
DO 20 IX=1,4
CALL BASE2(KRPT,METAD,NRREG,NRSEG,IAUX,AV)
JFIM=0
INIC1=-108
INIC2=-109
DO 60 K=1,34
INIC1=INIC1+LSEG
INIC2=INIC2+LSEG
JFIM=JFIM+LSEG
SUM1=0.0
SUM2=0.0
DO 30 L=INIC1,JFIM
30 SUM1=SUM1+AV(L)*AV(L-1)
DO 40 L=INIC2,JFIM
40 SUM2=SUM2+AV(L)**2
NP=NP+1
60 COEF(NP)=SUM1/SUM2
90 CONTINUE
DO 100 I=1,136
A(I,1)=FLOAT(I)
100 A(I,2)=COEF(I)
CALL GRAFG(A,1,1,136,136,2,2,1,1,ISIMB,2,1)
STOP
END
```



```

DIMENSION ISIMB(2)
DIMENSION RF(17), A(17,3)
DIMENSION AV(7500), IAU(7500)
DATA LCART/5/

C
C   PROGRAMA **FREL**
C   COMPARAR A FRQ. REL. DO SINAL COM A F.D.P. DE LAPLACE
C   SIG2G E A VARIANCIA GERAL, DA FRASE COMPLETA.
C
READ(LCART,777) SIG2G
777  FORMAT(9.3)
SIGMA=SQRT(SIG2G)

C
C   DISTRIBUICAO DE LAPLACE
C
COFF=1.0/((SQRT(2.0))*SIGMA)
DO 41 I=1,17
FLI=FLOAT(I)
ZY=2.0*COFF*ABS((-4.5+0.5*FLI)*SIGMA)
IF(ZY-174.0) 38,39,39
38  A(I,3)=COFF/EXP(ZY)
GO TO 41
39  A(I,3)=0.00
41  CONTINUE

C
C   CHAMADA DAS AMOSTRAS
KRPT=0
CALL BASE2 (KRPT,METAD,NRREG,NRSEG,IAU,AV)

C
C   PREPARO DA CONTAGEM
C
DO 96 I=1,17
96  A(I,2)=0.00

C
C   PREPARO DOS PONTOS DE REFERENCIA
C
DO 97 I=1,17
FLI=FLOAT(I)
97  RF(I)=SIGMA*(-4.25+0.5*FLI)
ALIM=4.25*SIGMA

C
C   CONTAGEM PARA O HISTOGRAMA
C
DEN=1875.0*SIGMA
DO 105 I=1,3750
IF(ABS(AV(I)).GT.ALIM) GO TO 105
DO 103 K=1,17
IF(AV(I).GT.RF(K)) GO TO 103
A(K,2)=A(K,2)+1.0
GO TO 105
103 CONTINUE
105 CONTINUE
DO 106 I=1,17
106 A(I,2)=A(I,2)/DEN

C
C   PREPARO DAS ABCISSAS
C
DO 107 I=1,17
FLI=FLOAT(I)
107 A(I,1)=-4.5+0.5*FLI

C
C   GRAFICO
C
ISIMB(1)=10
ISIMB(2)=13
CALL GRAFG(A,1,1,17,17,3,2,1,1,ISIMB,2,2)
STOP
END

```

Tese apresentada por

*Roberto Miscow Filho*  
ROBERTO MISCOW FILHO  
TC QEM *em Al*

e aprovada pelos Srs:

*José Paulo de Almeida Albuquerque*  
JOSÉ PAULO DE ALMEIDA E ALBUQUERQUE  
Ph.D.

*Luíz Costa da Silva*  
LUIZ COSTA DA SILVA  
Maj QEM-Ph.D.

*J. Wilson Marques Daudt*  
LUIZ WILSON MARQUES DAUDT  
Maj QEM-M.Sc.

*Antonio Julio Lossio Botelho*  
ANTONIO JULIO LOSSIO BOTELHO  
M.Sc.