

DETECTOR DE VOZ
PARA SISTEMAS DE
INTERPOLAÇÃO DIGITAL

JOSÉ SINDI YAMAMOTO

Orientador: FÁBIO VIOLARO

Tese apresentada à Faculdade de Engenharia de Campinas, da Universidade Estadual de Campinas - UNICAMP - como parte dos requisitos exigidos para obtenção do título de MESTRE EM CIÊNCIAS.

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA DE CAMPINAS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

DEZEMBRO 1982

UNICAMP
BIBLIOTECA

À

Vera e Patrícia

AGRADECIMENTOS:

Desejo agradecer o auxílio de inestimável valor, tanto na parte teórica como na parte prática, do meu orientador Dr. FÁBIO VIOLARO, sempre disposto a ler, analisar e sugerir modificações ao texto original, seja no próprio conteúdo como na maneira de apresentá-lo.

Meus agradecimentos aos engenheiros Antonio Vivaldi Rodrigues, Antonio Akira Terada e Gorou Kanno, de cuja vasta experiência e conhecimento resultaram grandes contribuições. Em especial, agradeço ao engenheiro Antonio Vivaldi Rodrigues, que tornou possível a idéia da realização deste trabalho. Ao futuro engenheiro Júlio Cesar registro os meus agradecimentos pelo eficiente auxílio prestado nos cálculos computacionais.

Desejo ainda agradecer à M. Júlia, pela primorosa datilografia, aos desenhistas Edson e Luis pela confecção dos desenhos, e a todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.

ÍNDICE

SUMÁRIO

CAPÍTULO I - INTRODUÇÃO	1
I.1 - CONSIDERAÇÕES SOBRE A ATIVIDADE DOS CA NAIS TELEFÔNICOS	2
I.2 - PRINCÍPIOS DE INTERPOLAÇÃO DE VOZ	3
I.3 - DESCRIÇÃO GERAL DO SISTEMA MCP-60	5
I.4 - OBJETIVO DO TRABALHO	9
CAPÍTULO II - CARACTERÍSTICAS DO SINAL DE VOZ E DO RUÍ DO	11
II.1 - INTRODUÇÃO	12
II.2 - POTÊNCIA MÉDIA DO SINAL DE VOZ TELEFÔ NICO E DO RUÍDO DE LINHA	12
II.3 - FUNÇÃO DENSIDADE DE PROBABILIDADE DO SINAL DE VOZ	18
II.4 - O MECANISMO DA PRODUÇÃO DA VOZ	21
II.5 - NÚMERO MÉDIO DE CRUZAMENTOS DE ZERO ..	24
II.6 - PERIODICIDADE E VARIAÇÃO DE AMPLITUDE DA ENVOLTÓRIA	33
II.7 - RUÍDO AMBIENTAL	34
CAPÍTULO III - ESTRATÉGIAS PARA DETECÇÃO DO SINAL DE VOZ E REJEIÇÃO DE RUÍDO	37
III.1 - INTRODUÇÃO	38
III.2 - OBJETIVOS DE PROJETO PARA O DETECTOR DE VOZ	38
III.3 - DETECTOR DE VOZ CONVENCIONAL	39
III.4 - DETECTOR DE VOZ NÃO-CONVENCIONAL	41

III.5 - DETECTOR DE VOZ Nº 1	46
III.6 - DETECTOR DE VOZ Nº 2	54
III.7 - DETECTOR DE OFF-SET	64
CAPÍTULO IV - IMPLEMENTAÇÃO	65
IV.1 - INTRODUÇÃO	66
IV.2 - DIAGRAMA DE CIRCUITOS DO DETECTOR DE VOZ PARA UM CANAL	66
IV.3 - DIAGRAMA DE CIRCUITOS DO DETECTOR DE VOZ PARA 60 CANAIS	73
CAPÍTULO V - TESTES DE DESEMPENHO E CONCLUSÕES	77
V.1 - TESTES DE DESEMPENHO	78
V.2 - CONCLUSÕES	81
APÊNDICE A - DESVIO PADRÃO DO NÚMERO DE CRUZAMENTOS DE ZERO POR INTERVALO DE TEMPO T DE UM RUÍDO GAUSSIANO BRANCO DE FAIXA LIMITADA EM f_0	84
APÊNDICE B - DIAGRAMAS DETALHADOS DE CIRCUITOS DO DETEC TOR DE VOZ	89
REFERÊNCIAS BIBLIOGRÁFICAS	100

SUMÁRIO

Neste trabalho são apresentadas algumas estratégias para a discriminação do sinal de voz do ruído, aplicáveis aos detectores de voz para sistemas de interpolação digital. Para a definição destas estratégias, são feitos estudos comparativos entre as características dos sinais de voz e do ruído presente nas linhas telefônicas, tais como número médio de cruzamentos de zero, variação de envoltória, etc.

Adicionalmente, é apresentada a implementação de um detector digital de voz para o sistema de interpolação de voz por ora denominado "MCP-60", que está sendo desenvolvido nos laboratórios de Eletrônica & Comunicações da Faculdade de Engenharia da Universidade Estadual de Campinas, bem como os resultados dos testes de desempenho realizados.

CAPÍTULO I

INTRODUÇÃO

I.1 - CONSIDERAÇÕES SOBRE A ATIVIDADE DOS CANAIS TELEFÔNICOS

O sinal de voz pode ser caracterizado por uma portadora na frequência de áudio (200 a 8000 Hz), tendo uma envoltória de baixa frequência (0 a 50 Hz). Como a sua forma de onda apresenta amplitudes praticamente nulas durante certos intervalos de tempo (separação entre sílabas e palavras), e amplitudes elevadas em outros intervalos, pode-se também descrever o sinal de voz como uma série de surtos de energia separados por intervalos de tempo de silêncio. A proporção de tempo ocupada por tais surtos de energia é denominada atividade média α . Mais precisamente, como os intervalos de tempo de silêncio entre os surtos podem ter uma duração muito curta, a atividade média α é definida considerando-se como silêncio (ausência de voz), somente aqueles períodos de no mínimo 350 ms em que a potência média a curto prazo* permanece mais do que 15 dB abaixo da potência média a longo prazo** [1]. Esta definição está razoavelmente de acordo com as experiências subjetivas, resultando, então, que silêncios de duração muito curta entre sílabas, ou mesmo entre palavras, sejam incluídos no tempo considerado como ativo. De acordo com esta definição, a atividade média de uma fala contínua α_c (por exemplo, leitura de um texto) foi determinada entre 65 e 75% [2].

Em uma conversação telefônica típica, contudo, cada usuário gasta, em média, tempos iguais para falar e ouvir. Este fato reduz a atividade média, numa dada direção de transmissão, de um fator $\alpha_s = 0,5$. Isto é, pode-se esperar que a atividade média de um canal telefônico ocupado é, aproximadamente, a metade da atividade de uma fala contínua. Além disto, deve ser considerado que nem todos os canais do sistema telefônico podem estar ocupados simultaneamente, mesmo na hora de maior movimento (HMM), e que cada nova chamada possui um período de baixa atividade em seu início, enquanto a ligação está sendo estabelecida. Estes efeitos contribuem adicionalmente para baixar a atividade média α , e são levados em conta através da eficiência de tronco do sistema, η . Este fator é da

* Potência média a curto prazo, significa potência média sobre um intervalo de tempo de 10 ms a 20 ms [1].

** Ver definição na pág. 13.

ordem de 0,7 na HMM para circuitos nacionais, e de aproximadamente 0,9 para circuitos internacionais [2].

Levando-se em conta os fatores acima mencionados, a atividade média de um canal telefônico, na HMM, num dado sentido de transmissão, é dada por:

$$\alpha = \alpha_c \cdot \alpha_s \cdot \eta \quad (I.1)$$

Portanto, tomando-se $\alpha_c = 0,725$ e $\alpha_s = 0,5$, obtêm-se:

- para circuitos nacionais: $\eta = 0,7 \Rightarrow \alpha = 0,25$
- para circuitos internacionais: $\eta = 0,9 \Rightarrow \alpha = 0,34$

Isto significa que, mesmo na HMM, um canal está ativo, num dado sentido de transmissão, em apenas 25% ou 34% do tempo, segundo se trate de circuitos nacionais ou internacionais, respectivamente.

Além dos parâmetros α_c , α_s e η , outros fatores adicionais também podem ser considerados no cálculo da atividade média α . Numa conversação normal, por exemplo, ocorrem períodos em que nenhuma das partes está falando, quando se aguarda alguém ser chamado ao telefone, quando se está pensando no que dizer, etc. Tais pausas duplas reduzem a atividade média α . Com menor frequência, podem ocorrer também períodos de fala dupla (double talking), quando ambas as partes falam simultaneamente, contribuindo para um aumento na atividade. Levando-se em conta estes e outros fatores, a atividade média pode atingir valores de aproximadamente:

- $\alpha = 0,18$ em circuitos nacionais
- $\alpha = 0,25$ em circuitos internacionais

I.2 - PRINCÍPIOS DE INTERPOLAÇÃO DE VOZ

Considere-se um sistema multiplex de m canais de voz. No caso de a multiplexagem ser em frequência (FDM), são reservadas m faixas disjuntas de frequência para a transmissão de m sinais de voz e, no caso da multiplexagem ser temporal (TDM), são reserva

das m janelas temporais disjuntas para a transmissão de m sinais de voz. De acordo, entretanto, com a seção anterior, na HMM um canal está ativo em média apenas 25% do tempo de transmissão, e no restante do tempo essas faixas de frequências ou janelas temporais ficam ociosas. A interpolação de voz consiste em se aproveitar aquelas janelas temporais ou faixas de frequências, nos intervalos de tempo em que estão ociosas, para a transmissão adicional de mais sinais de voz. Desta maneira, um canal de transmissão (faixa de frequências ou janela temporal) deixa de ficar reservada em tempo integral para a transmissão de uma única conversação de cada vez, e passa a servir, em tempo parcial, a mais de uma conversação.

Se o número de canais de entrada é grande (em torno de 50 canais) a probabilidade de se ter mais do que a metade destes canais simultaneamente ativos (mesmo na HMM) é baixa. Assim, com o auxílio de um circuito comumente denominado "detector de voz", que seja capaz de detectar a presença dos sinais de voz em cada um dos canais de entrada do sistema de transmissão, é possível dobrar a sua capacidade, utilizando-se o princípio da interpolação de voz.

Uma atividade média de 25% é muito difícil de ser obtida na prática, pois os canais telefônicos são afetados por vários tipos de perturbações, como ruído, diafonia, eco, etc., tornando difícil a detecção correta do sinal de voz. Assim, os melhores detectores de voz alcançam, geralmente, uma atividade média da ordem de 40%, restando 60% de tempo ocioso, o qual pode ser aproveitado para transmitir informação efetiva de outros canais.

Quando o número de canais de entrada simultaneamente ativos é igual ou menor que o número de canais de transmissão, os sistemas de interpolação de voz operam como um sistema de transmissão normal, sem introduzir nenhuma degradação devido à interpolação. Quando o número de canais de entrada simultaneamente ativos é maior do que o número de canais de transmissão, diferentes procedimentos de interpolação de voz podem ser usados para contornar esta condição de saturação (número de canais de entrada ativos maior do que o número de canais de transmissão). Neste caso, é introduzida alguma degradação no sinal de voz.

Nos sistemas de multiplexação por divisão de frequência

(FDM), é empregada a interpolação de voz do tipo TASI [3,4] em que pode haver perda parcial de informação (freeze-out) na condição de saturação, até que um canal de transmissão se torne disponível. Por outro lado, nos sistemas de multiplexagem por divisão de tempo (TDM) do tipo MCP (modulação por codificação de pulsos) é possível usar dois procedimentos na condição de saturação: redução do número de bits usados para codificar as amostras PAM, ou redução da frequência de amostragem. O número de janelas temporais é aumentado até o necessário através destes dois procedimentos, às custas de uma pequena degradação na qualidade do sinal, evitando-se, contudo, o problema do "freeze-out".

I.3 - DESCRIÇÃO GERAL DO SISTEMA MCP-60

O Sistema MCP-60 [5,6], ilustrado na Fig. I.1, consiste numa interface digital entre dois sistemas MCP-30 e a linha de transmissão. Trata-se de um sistema de interpolação digital de voz que tem como objetivo transmitir 60 canais telefônicos com a mesma taxa de bits (2,048 Mbit/s) de um Sistema MCP-30. Neste sistema, o problema do "freeze-out" é eliminado através da redução do número de bits por amostra nos instantes de saturação.

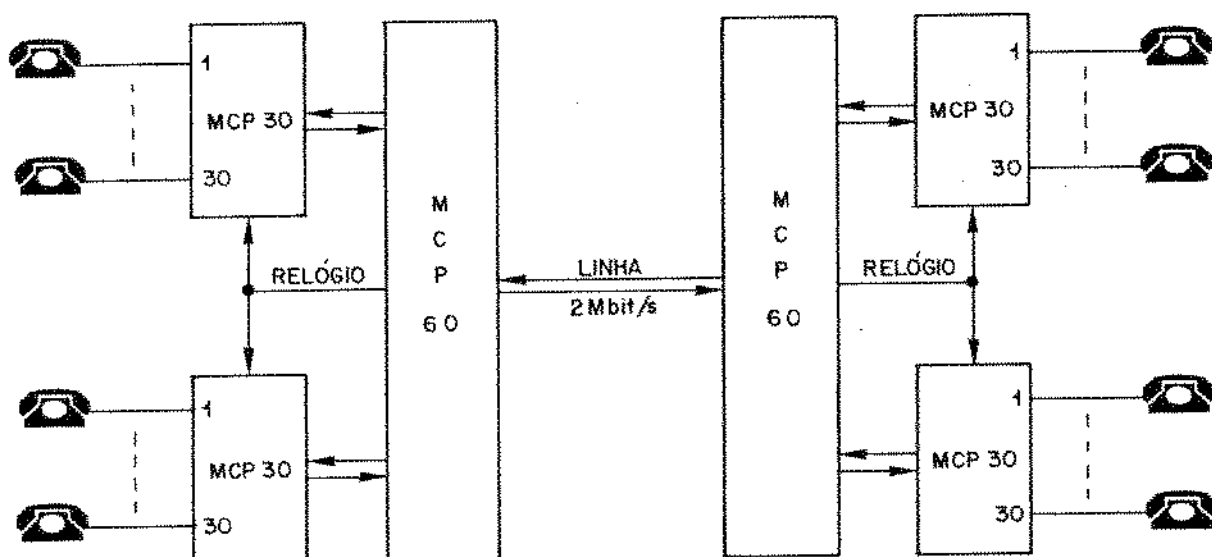


Fig. I.1 - Sistema MCP-60

Os dois Sistemas MCP-30 tributários são síncronos a nível de relógio, o qual é gerado pelo Sistema MCP-60. Mas, como não são necessariamente síncronos em quadro e nem em super-quadro, o Sistema MCP-60 dispõe de uma lógica automática que sincroniza os dois sistemas MCP-30 tributários em quadro e super-quadro.

O quadro do Sistema MCP-60 encontra-se ilustrado na Fig. I.2, e o seu comprimento é o mesmo do quadro do Sistema MCP-30, isto é, 256 bits. Estes estão distribuídos em:

- 4 bits de prefixo para a transmissão de informação de atividade;
- 4 bits de paridade para a correção de erros nos bits de prefixo;
- 8 bits para a transmissão alternada ora da supervisão conjunta dos dois Sistemas MCP-30 tributários, ora do sincronismo de quadro do Sistema MCP-60;
- 8 bits para a transmissão, a cada 15 quadros, da sinalização E&M conjunta dos dois Sistemas MCP-30 tributários e, no 16º quadro, do sincronismo de super-quadro do Sistema MCP-60;
- 232 bits para a transmissão das amostras PAM dos sinais de voz.

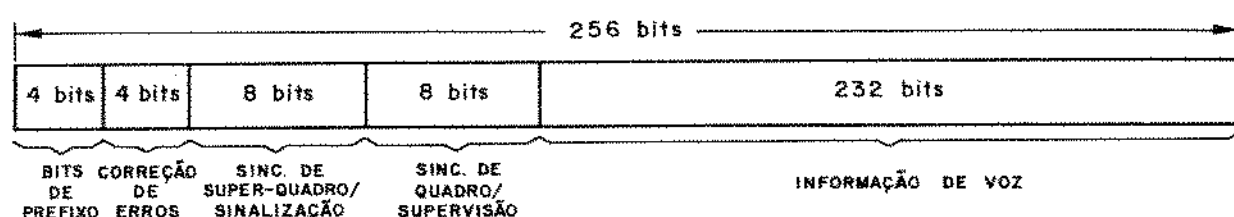


Fig. I.2 - Formatação de quadro do Sistema MCP-60

A informação sobre a atividade dos 60 canais de voz é transmitida através dos 4 bits de prefixo. Cada bit de prefixo indica o estado de atividade de somente um canal, ou seja, é utilizado

zado um bit de prefixo por canal. Como a palavra de prefixo é composta de 4 bits, são necessários 15 quadros para se transmitir o estado de atividade dos 60 canais de voz. Este intervalo de tempo é igual àquele necessário para se transmitir a sinalização E&M dos 60 canais de voz, pois, no Sistema MCP-60, é enviada a informação de sinalização de 4 canais por quadro. Desta maneira, tanto a sincronização da transmissão dos bits de prefixo como a sincronização da sinalização é feita pela palavra de sincronismo de super-quadro, que é transmitida no 16º quadro.

Observe-se que, como no quadro do Sistema MCP-60 são transmitidos somente os canais considerados ativos, é possível, na recepção, baseado nos bits de prefixo, identificar em qual janela temporal está sendo transmitida a amostra do sinal de voz correspondente a um determinado canal de entrada e, além disto, com quantos bits.

Os $L = 232$ bits restantes são os bits que realmente carregam a informação das amostras dos canais de voz. Desta maneira, podem ser transmitidos no máximo 29 canais com 8 bits por amostra. Se o número de canais ativos for maior do que 29, torna-se necessária a redução do número de bits por amostra. A probabilidade de os canais ativos serem codificados com $b=i$ bits por amostra é dada por [5]:

$$P[b=i] = \sum_{N_A=p}^q \binom{60}{N_A} \alpha^{N_A} (1-\alpha)^{60-N_A}, \quad (I.2)$$

onde "p" é o menor número de canais ativos por quadro que provoca congestionamento se forem utilizados $(i+1)$ bits por amostra. Por outro lado, "q" é o número máximo de canais ativos que não provoca congestionamento se forem utilizados "i" bits por amostra. Os valores de p e q satisfazem as relações:

$$p = \left[\frac{L}{i+1} \right] + 1 \quad (I.3)$$

$$e \quad q = \left[\frac{L}{i} \right], \quad (I.4)$$

onde $[x]$ significa o maior inteiro contido em x, e L é o número

total de bits (232 bits) disponível para informação de voz no quadro do sistema MCP-60.

Utilizando-se as expressões acima, obtêm-se a Tabela I.1, que fornece as probabilidades $P[b=i]$ para diferentes valores de atividade média α , supondo-se $\eta = 100\%$.

Conhecendo-se os valores de $P[b=i]$ pode-se calcular a degradação D na relação Sinal/Ruído, dada por [5]:

$$D(\text{dB}) = 10 \log \sum_{i=1}^8 4^{8-i} \cdot P[b=i] \quad (\text{I.5})$$

Assim, com os dados da Tabela I.1 e a eq.(I.5), obtêm-se a Tabela I.2, que fornece a degradação na relação Sinal/Ruído para diferentes valores de atividade média α , supondo-se $\eta = 100\%$.

i	P[b=i]		
	$\alpha=0,43$	$\alpha=0,40$	$\alpha=0,37$
8	0,853	0,925	0,973
7	0,144	0,068	0,026
6	0,023	0,007	0,001

Tabela I.1 - Probabilidade de ocorrência de palavras com comprimento $b=i$ bits ($P[b=i]$) em função da atividade média α , supondo que todos os canais estão ocupados ($\eta=100\%$) [5]

α	Degrad. (dB)
0,43	2,5
0,40	1,2
0,37	0,4

Tabela I.2 - Degradação da relação Sinal/Ruído em função da atividade média α , supondo que todos os canais estão ocupados ($\eta=100\%$) [5]

Deve ser ressaltado que as probabilidades da Tabela I.1 e as degradações da Tabela I.2, foram calculadas considerando - se que as palavras código possuem o mesmo comprimento dentro de cada quadro, de modo que nem todos os bits disponíveis em um quadro são eficientemente utilizados. Na realidade, entretanto, a composição de quadro do Sistema MCP-60 na sua versão mais recente, é tal que algumas palavras códigos possuem um número de bits maior que os demais, de modo a aproveitar todos os bits disponíveis do quadro. Assim, o Sistema MCP-60 apresenta um desempenho ligeiramente superior àquele apresentado na Tabela I.2.

I.4 - OBJETIVO DO TRABALHO

O sucesso dos sistemas de interpolação de voz depende, fundamentalmente, do bom desempenho do detector de voz. Um detector de voz de baixa sensibilidade provoca uma redução da atividade média α , com uma evidente vantagem em termos de capacidade de transmissão; contudo, ele causa uma degradação na qualidade do sinal de voz devido aos cortes no início ou no meio das palavras. Por outro lado, um detector de voz muito sensível provoca um aumento considerável na atividade média α devido à detecção de ruído. Neste caso, ocorre uma degradação na qualidade da transmissão devido a maior quantidade de amostras codificadas com um número de bits menor do que 8.

Para contornar estes problemas, é necessário alcançar um compromisso, de modo que não haja uma degradação notável na qualidade do sinal de voz nem devido aos cortes no início ou no meio das palavras, e nem devido a maior quantidade de amostras com um número reduzido de bits.

Os dados de desempenho apresentados na seção anterior sugerem que, para alcançar tal compromisso, o valor da atividade média α não deve ser superior a 40%. Todavia, devido à presença na linha de ruídos, eco, etc., mesmo este valor de atividade média é difícil de ser alcançado.

O presente trabalho tem como objetivo determinar uma estratégia que discrimine o sinal de voz do ruído, e implementar um

detector digital de voz usando esta estratégia. A discriminação entre o sinal de voz e o ruído deve ser tal que o detector de voz satisfaça o compromisso citado acima. A implementação será feita visando a sua utilização no Sistema MCP-60. Para ser usado por um outro sistema qualquer de interpolação digital de voz, bastará que sejam feitas as adaptações necessárias sem, contudo, modificar a estratégia original para detecção e rejeição de ruído. Adicionalmente, por motivos de compactação e de economia, o detector de voz deverá consistir de um único circuito funcionando em compartilhamento de tempo (time-sharing) para os 60 canais do Sistema MCP-60.

CAPÍTULO II

CARACTERÍSTICAS DO SINAL DE VOZ E DO RUÍDO

II.1 - INTRODUÇÃO

A finalidade deste capítulo é analisar algumas características do sinal de voz e do ruído presentes nos circuitos telefônicos, a fim de se determinar uma estratégia que permita ao detector de voz discriminar estes dois sinais. Com este objetivo, são feitas análises comparativas entre os mesmos, sob o ponto de vista de potência média, função densidade de probabilidade e número médio de cruzamentos de zero.

Com relação ao ruído, é necessário escolher um modelo que represente adequadamente o seu efeito no desempenho do detector de voz. Como o ruído presente num circuito telefônico é a soma de vários ruídos provenientes de várias fontes (ruído térmico, ruído de intermodulação, etc.), é de se esperar que o ruído total satisfaça as condições de uma distribuição gaussiana. Além disso, considerando-se que grande parte dos ruídos que o compõem possuem um espectro muito largo, costuma-se adotar como modelo de ruído nos projetos do detector de voz o ruído gaussiano branco na faixa do canal telefônico [7,8,9]. Portanto, neste capítulo, as análises são feitas com relação ao ruído gaussiano branco de faixa limitada em 3,4 KHz. Adicionalmente, como em alguns casos o valor da atividade média α pode ser aumentada devido ao ruído monótono de alto nível introduzido pelo ambiente que rodeia o interlocutor, algumas análises são feitas, também, com relação a este tipo de ruído.

II.2 - POTÊNCIA MÉDIA DO SINAL DE VOZ TELEFÔNICO E DO RUÍDO DE LINHA

A forma de onda associada ao sinal de voz não é fácil de ser caracterizada. Além das frequências de áudio serem moduladas em amplitude a uma taxa silábica, as pausas entre as palavras e as frases fazem com que a energia da voz seja concentrada em surtos de duração média de 1s, espaçadas por intervalos de tempo de duração média também em torno de 1s. Assim, o sinal de voz consiste de surtos de energia espaçados no tempo aleatoriamente, e de duração também aleatória. Em consequência disto, torna-se muito difícil fazer uma medida precisa do sinal de voz. Contudo, indepen

dentemente dos problemas acima, é necessário que a magnitude do sinal de voz seja medida e caracterizada de alguma maneira útil ao projeto e operação de sistemas envolvendo equipamentos eletrônicos e meios de transmissão de vários tipos. No caso específico do presente trabalho, interessa saber qual é o limiar inferior do nível de voz a ser considerado no projeto do detector de voz. O medidor denominado VU foi desenvolvido para atender a estas necessidades. Este medidor foi projetado para medir a potência média aproximada da voz durante um intervalo silábico (da ordem de 300 ms) [2,10]. Como consequência, o medidor fornece somente quão alto é a voz do usuário quando ele fala, não levando em conta a atividade. Assim, um usuário que faz muitas pausas pode fornecer uma medida em VU maior que um usuário que fala continuamente, mesmo que este tenha uma potência média maior que aquele.

Embora seja um caso irreal, uma voz de volume constante é aquela que, apesar de apresentar flutuações em amplitude a uma taxa silábica, produz uma leitura no medidor VU que é constante com o tempo. Isto é, o ponteiro do medidor VU, embora apresente pequenas flutuações, atinge sempre um mesmo valor máximo a cada sílaba que é emitida. A relação entre VU e potência média de uma fala contínua é dada pela relação:

$$P = (V_0 - 1,4)\text{dBm} \quad (\text{II.1})$$

onde: - V_0 é o volume constante de uma fala contínua. Esta fala contínua inclui somente as pausas naturais entre palavras ou sílabas, não incluindo as pausas para pensar ou para esperar resposta.

- O fator de conversão -1,4 dB é o resultado de testes empíricos com uma grande variedade de pessoas lendo um texto (fala contínua).

A potência média definida pela eq.(II.1) é normalmente denominada potência média a longo prazo (long-term average power). O termo longo prazo significa que a medida foi efetuada com uma constante de tempo superior a 10s.

Como foi visto na seção I.1, a atividade média de uma fala contínua está em torno de 75%. Tomando-se para α_c um valor de

0,725, a eq.(II.1) pode ser reescrita como:

$$P = (V_0 + 10 \log \alpha_c) \text{ dBm} \quad (\text{II.2})$$

Portanto, o volume V_0 pode ser interpretado como sendo a potência média em dBm excluindo as pausas naturais entre sílabas e palavras.

A potência média a longo prazo da voz de um usuário telefônico que fala de maneira não contínua, isto é, com pausas para reflexão e escuta, é relacionada àquela de uma fala contínua introduzindo-se o fator de atividade de carga, α_L :

$$P_M = (V_0 - 1,4 + 10 \log \alpha_L) \text{ dBm} \quad (\text{II.3})$$

Considerando-se apenas os parâmetros α_s e η , o fator de atividade de carga é dado por: $\alpha_L = \alpha_s \eta$. Portanto, para ligações nacionais obtém-se $\alpha_L = 0,35$, e para ligações internacionais $\alpha_L = 0,45$. Levando-se em conta outros fatores de atividade (pausas duplas, por exemplo) estes valores podem cair para 0,25 e 0,35, respectivamente, sendo que o valor $\alpha_L = 0,25$ é recomendado pelo CCITT para o cálculo de carga em sistemas FDM [11].

Note-se que α_L é diferente do fator de atividade, α , dado por $\alpha_c \cdot \alpha_s \cdot \eta$. No fator de atividade de carga, α_L , não entra o parâmetro α_c , pois o mesmo já foi considerado pela constante $10 \log \alpha_c = -1,4$ para $\alpha_c = 0,725$.

Até agora, a caracterização da potência média a longo prazo foi feita, idealmente, somente para um usuário que possui uma voz de volume constante V_0 . Isto é, a eq.(II.3) caracteriza a potência média a longo prazo de um dado canal de voz que carrega a conversação de um usuário que fala com volume constante V_0 . Num caso real, entretanto, através deste canal podem circular conversas de diferentes usuários com diferentes volumes de voz. Assim, o que se tem na realidade, é uma distribuição de volume que, observada no ponto de nível relativo de transmissão de 0 dB (0 dB transmission level point), é uma função de vários parâmetros, dentre os quais:

1. Sexo do assinante

2. Hábitos na maneira de falar do assinante
3. Tipo de chamada
4. Região geográfica
5. Distância da ligação
6. Desempenho do telefone
7. Perdas nas linhas

Observa-se que os homens falam mais alto do que as mulheres, que as chamadas comerciais possuem um volume maior do que as chamadas sociais; as pessoas das grandes cidades falam mais alto do que as pessoas das pequenas comunidades, e também falam mais alto nas ligações de longa distância do que nas ligações locais. Naquelas ligações, tem-se comprovado um aumento de 1 dB no volume da voz para cada 1600 km de distância, apesar da baixa perda e qualidade normal em tais ligações. Tem-se também constatado que as variações no volume devem-se mais ao assinante do que à rede telefônica [2]. Assim, a medida de volume de um grande número de usuários telefônicos, possui uma distribuição gaussiana devido aos efeitos combinados discutidos anteriormente. Assume-se, então, que em termos de volume, o sinal de voz telefônico possui uma distribuição gaussiana com uma média V_0 VU e um desvio padrão de σ dB. Neste caso, a potência média a longo prazo por canal não é mais dada pela eq. (II.3), a não ser que a voz de todos os usuários do sistema esteja a um volume constante V_0 . Para obter a potência média por canal, é preciso converter a distribuição gaussiana, em termos de VU ou dBm, para uma distribuição log-normal em termos de mW. A função log-normal resultante é assimétrica e seu valor médio é maior que o valor médio da distribuição gaussiana em termos de dBm, como mostra o exemplo da Fig. II.1 (os valores desta figura não são típicos, servindo apenas como ilustração). Assim, demonstra-se que o volume V_{0C} correspondente à potência média de voz da distribuição log-normal é dado por [2]:

$$V_{0C} = V_0 + 0,115\sigma^2 \text{ VU} \quad (\text{II.4})$$

Portanto, a potência média a longo prazo por canal, no caso de se considerar vários usuários, é igual a:

$$P_{MC} = (V_0 + 0,115\sigma^2 - 1,4 + 10 \log \alpha_L) \text{ dBm} \quad (\text{II.5})$$

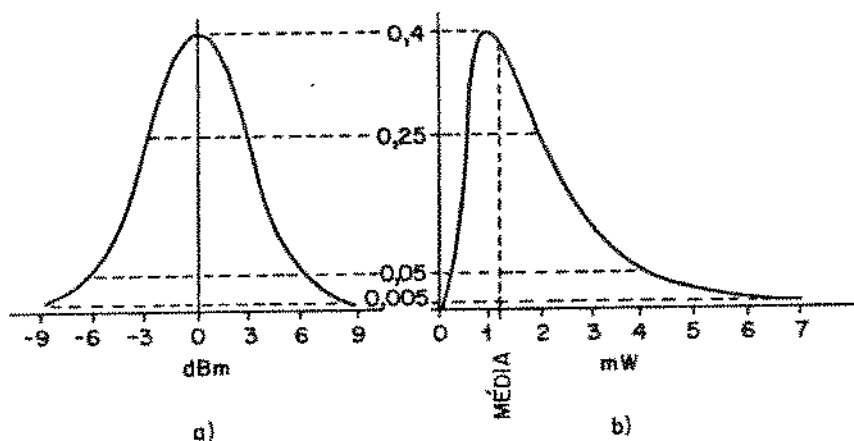


Fig. II.1 - a) Distribuição gaussiana em termos de dBm
 b) Distribuição log-normal em termos de mW

Levantamentos efetuados por diversas administrações e fabricantes de equipamentos de transmissão (BPO, AT&T, SIEMENS), mostram uma média V_0 , no ponto de nível relativo de transmissão de 0 dB, entre -10 e -25 VUO, e um desvio padrão σ entre 4 e 8 dB, sendo que o valor de 5,8 dB é o mais utilizado atualmente.

A potência média dada pela eq.(II.5) é usada para determinar a capacidade de carga de um canal multiplex para transmissão de voz. Sob o ponto de vista do detector de voz, contudo, interessa saber o valor de pior caso da potência média a longo prazo para $\tau_L = 100\%$, isto é, o limiar inferior do nível de voz para o qual o detector de voz deverá, ainda, apresentar uma boa sensibilidade. Este valor de pior caso pode ser calculado diretamente em função da distribuição de volume. Assim, tomando-se como distribuição de pior caso $V_0 = -25$ VUO e $\sigma = 5,8$ dB, e considerando-se três desvios padrão, obtém-se como valor de pior caso (encontrado com uma probabilidade da ordem de 10^{-3}) um volume de -42,4 VUO ou, aplicando-se a eq.(II.3) para $\sigma_L = 100\%$, uma potência média de -43,8 dBm0.

O valor de potência média por canal, P_{MC} , adotado pelo CCITT como objetivo do projeto, é de -15 dBm0. O CCITT adota também um desvio padrão σ de 5,8 dB, e um fator de carga α_L de 0,25. Sob estas condições, aplicando-se a eq.(II.5), obtém-se um volume médio V_0 de -11,5 VUO. Considerando-se três desvios padrão de 5,8 dB em relação a este valor de V_0 , resulta que o valor de pior caso da potência média é da ordem de -30 dBm0. Portanto, o valor de

pior caso da potência média adotado neste trabalho, está com uma margem de segurança de aproximadamente 14 dB em relação ao valor adotado pelo CCITT.

Até este ponto, foi feito um estudo a fim de determinar a potência média P_M de pior caso para $\alpha_L = 100\%$, a ser considerada no projeto do detector de voz. Com relação ao ruído, tal estudo torna-se mais simples, pois os valores máximos de potência de ruído que podem estar presentes em circuitos telefônicos estão especificados pelo CCITT [12]. Estes valores estão apresentados na Tabela II.1, e correspondem aos valores máximos da potência de ruído admissível em circuitos telefônicos com vistas à manutenção, isto é, tomada de ação a fim de descobrir alguma falha.

Distância em km da rota	Ruído medido em dBm0
< 320	-52,5
321 a 640	-50,5
641 a 1600	-48,5
1601 a 2500	-46,5
2501 a 5000	-43,5
5001 a 10000	-40,5
10001 a 20000	-37,5

Tab. II.1 - Valores máximos de potência de ruído permitido em circuitos telefônicos

Pela Tab. II.1 observa-se que o maior nível de ruído é de -37,5 dBm0. Considerando-se, contudo, que é possível que alguns circuitos telefônicos tenham um nível de ruído maior que o especificado pelo CCITT e que, além disto, os valores da tabela acima não incluem o ruído da linha de assinante, o projeto do detector de voz será feito supondo-se que o valor máximo da potência média de ruído é de -35 dBm0.

II.3 - FUNÇÃO DENSIDADE DE PROBABILIDADE DO SINAL DE VOZ

Na Fig. II.2 são apresentadas curvas experimentais da função densidade de probabilidade do sinal de voz, passando por diferentes tipos de microfone, juntamente com a função densidade de probabilidade do ruído gaussiano para efeito de comparação [1]. Entre as várias curvas teóricas que têm sido propostas para a representação da função densidade de probabilidade do sinal de voz, as mais usuais são a exponencial e a distribuição gama, sendo que a distribuição gama é a que mais se aproxima das curvas experimentais.

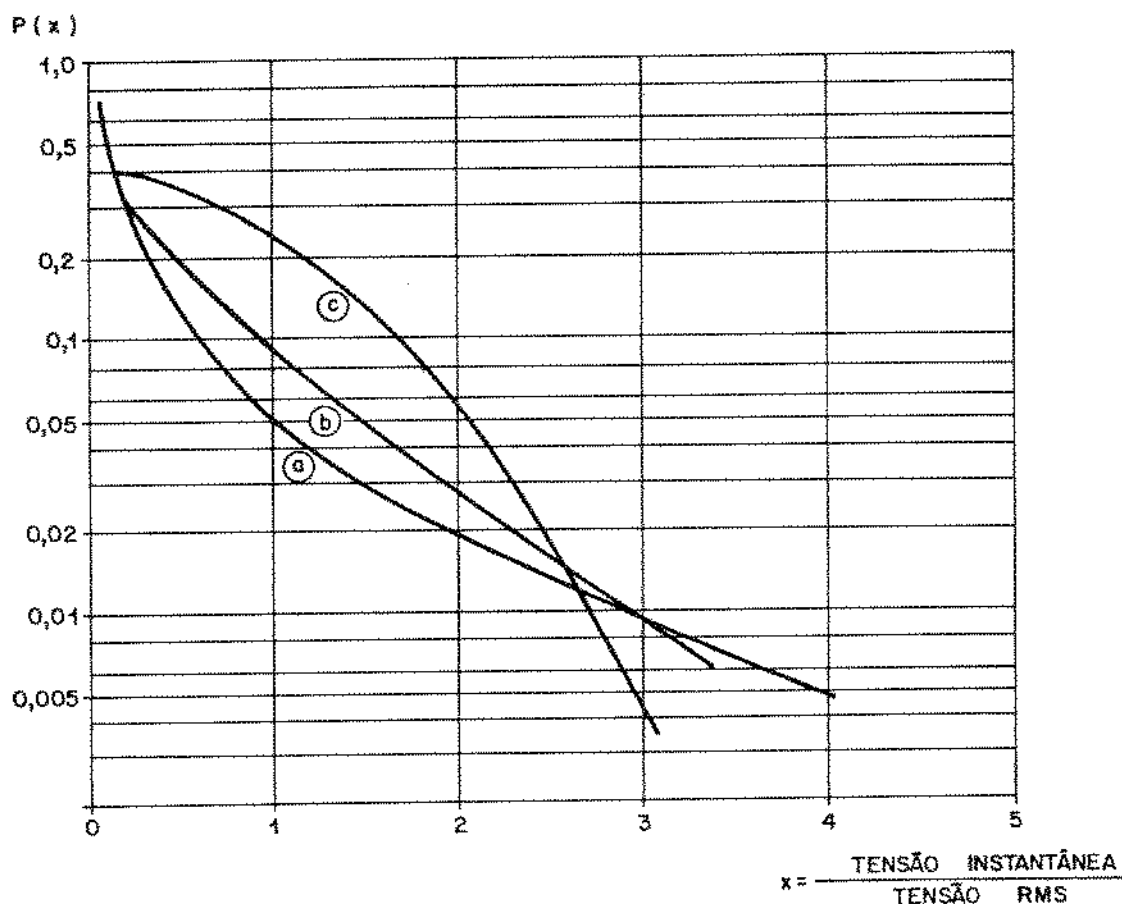


Fig. II.2 - Função densidade de probabilidade para o sinal de voz e ruído gaussiano [1]

- a) Curva experimental usando telefone BPO Nº 13
- b) Curva experimental usando microfone de alta qualidade
- c) Função densidade de probabilidade do ruído gaussiano

A distribuição gama é assumida ser simétrica em relação ao nível 0 (zero), e sua parte positiva é dada por uma expressão da forma:

$$p(x) = \frac{k}{2\Gamma(\ell)} (kx)^{\ell-1} e^{-kx}, \quad (\text{II.6})$$

onde:

ℓ é um parâmetro

$$k = \sqrt{\ell(\ell+1)}$$

$\Gamma(\ell)$ é a função gama

x = tensão instantânea/tensão RMS

O melhor ajuste da distribuição gama à curva b (microfone de alta qualidade) é obtido com $\alpha = 0,5$, e o melhor ajuste à curva a (microfone BPO Nº 13) é obtido com $\alpha = 0,2$. Para os telefones mais modernos é razoável assumir uma distribuição próxima à curva b, e portanto uma distribuição gama com $\alpha = 0,5$. Comparando-se assim a curva b com a curva c associada ao ruído gaussiano, observa-se que o sinal de voz apresenta uma probabilidade de a tensão instantânea ultrapassar a tensão RMS maior do que para o ruído. Em outros termos, a voz possui um fator de pico maior que o do ruído.

Uma análise probabilística mais precisa deste fato pode ser feita comparando-se a distribuição acumulada complementar da voz e do ruído. Para tanto, considere-se inicialmente, para facilitar a análise, que o sinal de voz apresenta uma função densidade de probabilidade $p(x)$ próxima à exponencial ($\ell=1$), isto é:

$$p(x) = \frac{\sqrt{2}}{2} e^{-\sqrt{2}x} \quad (\text{II.7})$$

A função de distribuição acumulada complementar para um sinal de voz retificado torna-se, portanto:

$$F_{|x|}(x) = \begin{cases} e^{-\sqrt{2}x} & , \quad x > 0 \\ 0 & , \quad x < 0 \end{cases} \quad (\text{II.8})$$

A função de distribuição acumulada complementar para um ruído gaussiano de média zero retificado é dada por:

$$F_{|x|}(x) = \frac{2}{\sqrt{2\pi}} \int_x^{\infty} e^{-\lambda^2/2} d\lambda, \quad (II.9)$$

As funções (II.8) e (II.9) encontram-se representadas na Fig.II.3. Observe-se que para x acima de 4,5 dB, a probabilidade do valor instantâneo da voz ultrapassar o seu valor RMS torna-se rapidamente bem maior do que para o ruído. Por exemplo, a probabilidade do valor instantâneo da voz estar mais do que 12 dB acima do seu valor RMS, está em torno de 0,35%, enquanto que para o ruído está em torno de 0,01%. Observe-se pela Fig. II.2 que quanto menor o valor de λ , maior será esta probabilidade. Portanto, para um sinal de voz telefônico ($0,2 < \lambda < 0,5$), a probabilidade no exemplo acima é maior do que 0,35%.

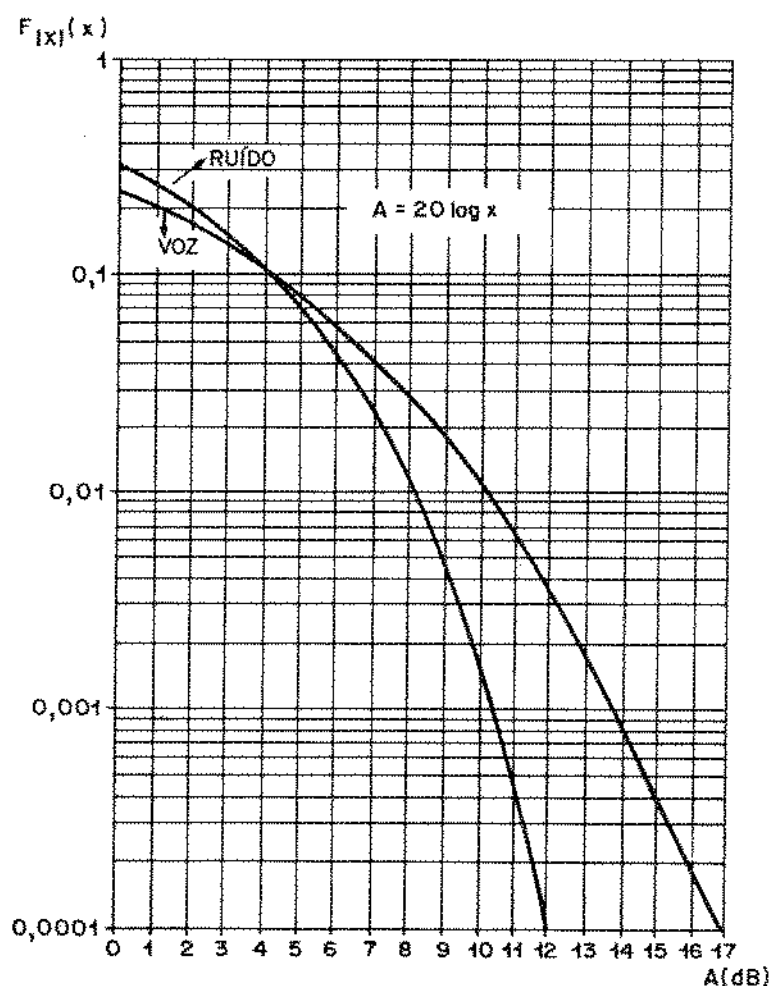


Fig. II.3 - Função distribuição acumulada complementar para voz e ruído

II.4 - O MECANISMO DA PRODUÇÃO DA VOZ

A Fig. II.4 procura ilustrar as características do trato vocal humano. O mesmo é constituído por um tubo com aproximadamente 17 cm de comprimento, em conjunto com várias estruturas conjugadas, que influenciam de diferentes maneiras o comportamento do fluxo de ar em seu interior. A região superior é subdividida nos tratos nasal e oral, entre os quais o véu palatino atua como uma válvula flexível de controle das ressonâncias nasais em determinadas emissões.

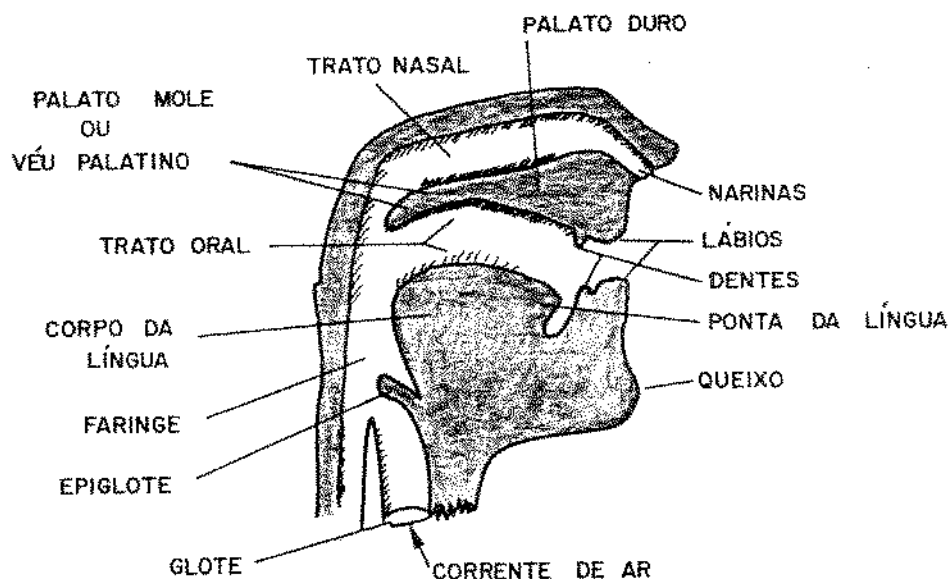


Fig. II.4 - Trato vocal humano

O tubo começa na glote, que é uma passagem estreita da laringe onde se encontram pares de pequenos músculos chamados cordas vocais. Estas cordas vocais são fixas na base e mõveis no topo. Durante a simples respiração, as cordas vocais ficam afastadas entre si, isto é, a glote permanece aberta e o ar passa livremente através dela. No momento em que se começa a falar, a glote se estreita, as cordas vocais se aproximam, e o fluxo de ar impellido

pelo pulmão ao chegar à glote choca-se com as cordas vocais . A pressão exercida pelo fluxo de ar aumenta até que as cordas vocais se desloquem. Com o deslocamento, as cordas, pela própria tensão, tendem a voltar à posição inicial, a pressão de ar as separa, a tensão torna a uní-las, e assim sucessivamente. Desta maneira, a corrente de ar vinda dos pulmões é transformada, pela vibração das cordas vocais, numa sequência de pulsações bruscas de ar. A frequência de repetição dos pulsos de ar varia de 90 Hz, para o baixo, até cerca de 300 Hz para o soprano e pessoas de voz estridente [13]. Cada vez que as cordas vocais voltam à posição inicial , cortando o pulso de ar excitador, o ar na parte superior do tubo vibra ou ressoa durante alguns milésimos de segundo. As cordas vocais, então, se separam e o fluxo de ar recomeça, criando condições para o próximo ciclo.

O comprimento da coluna de ar vibrante é igual à distância da glote até os lábios, onde se dá o acoplamento entre as vibrações do ar interior e o ar exterior circundante. Esta coluna de ar vibra de vários modos ou frequências de ressonância, correspondentes a diferentes múltiplos do quarto de comprimento de onda acústica. Existe uma acentuada ressonância ou pico de energia numa frequência tal que o comprimento do tubo é igual a um quarto do comprimento de onda; um outro pico de energia localiza-se numa frequência em que o comprimento do tubo é igual a três quartos do comprimento de onda; e assim por diante, em todos os múltiplos ímpares do quarto do comprimento de onda. Se um tubo de 17,4 cm de comprimento tivesse um diâmetro constante de um extremo a outro , esses picos de energia estariam localizados nas frequências de ressonância de 500 Hz, 1.500 Hz, 2.500 Hz, 3.500 Hz, e assim por diante. Tais frequências de ressonância são denominadas frequências formantes da fala. O trato vocal humano, porém, não é um tubo de diâmetro constante de um extremo a outro, e além do mais seu formato varia de um fonema para outro, de modo que as frequências de ressonância não são fixadas a intervalos regulares como no exemplo acima. A Tabela II.2 fornece, como exemplo, os valores médios das três primeiras formantes de diversas vogais do idioma inglês [14].

	F ₁	F ₂	F ₃
beet	270	2290	3010
bit	390	1990	2550
bet	530	1840	2480
bat	660	1720	2410
but	520	1190	2390
hot	730	1090	2440
bought	570	840	2410
foot	440	1020	2240
boot	300	870	2240
bird	490	1350	1690

Tab. II.2 - Formantes das vogais do idioma inglês [14]

Os sons das vogais são formados mediante o ajuste das frequências formantes, no trato vocal, segundo os valores apresentados na Tab. II.2. Estes valores são valores médios, pois, na realidade, variam de uma pessoa para outra. Além disto, a presença de consoantes adjacentes pode, ocasionalmente, modificar os valores fornecidos.

Para gerar outros sons, como as consoantes, entram em ação os diversos componentes do trato vocal mostrados na Fig. II.4. Os sons das consoantes consistem principalmente de vários estalos, sibilos, chiados e interrupções impostas à coluna de ar vibrante pelo fechamento dos lábios, pelo contato da ponta da língua com o palato duro, etc. De acordo com a ação exercida pelos vários componentes do trato vocal, as consoantes são divididas em diversas classes. Assim, em função da intensidade de obstrução imposta ao fluxo de ar, tem-se as consoantes oclusivas e fricativas. Em função da ação das cordas vocais, tem-se as consoantes sonoras e surdas. Adicionalmente, há uma série de outras classes, mas estas são as que mais dizem respeito ao projeto do detector de voz.

As consoantes oclusivas são caracterizadas pela aproximação completa de dois órgãos da boca, o que determina uma interrupção momentânea do fluxo de ar vindo dos pulmões. Ao cessar este obstáculo em virtude do afastamento rápido de tais órgãos, o ar acumulado atrás deles sai repentina e violentamente, ocasionando

um ruído seco, comparável a uma pequena explosão. Ex.: palácio, bala, tão.

As consoantes fricativas resultam da aproximação incompleta de dois órgãos da boca, o que obriga o fluxo de ar comprimir-se a fim de passar pela fenda estreita que assim se forma. O ar, ao escoar ininterruptamente por esta fenda, produz um ruído comparável a uma fricção. Ex.: sela, zabumba, xã.

As consoantes sonoras são as que, ao serem produzidas, põem as cordas vocais em vibração, pois, o fluxo de ar encontra a glote fechada. Ex.: bala. gato.

As consoantes surdas, ao contrário, são produzidas sem vibração das cordas vocais, pois o fluxo de ar encontra a glote aberta. Ex.: pato, caro.

II.5 - NÚMERO MÉDIO DE CRUZAMENTOS DE ZERO

Para um dado sinal, a razão na qual ocorrem os cruzamentos de zero depende de seu conteúdo espectral. Em outras palavras, se o seu espectro estiver centrado em torno de frequências baixas, a taxa de cruzamentos de zero será baixa; se, ao contrário, o seu espectro estiver centrado em torno de frequências altas, a taxa de cruzamentos de zero será elevada.

No caso do espectro de potência do sinal de voz, observa-se um pico em torno de 250 a 500 Hz, e acima desta frequência o espectro cai de 8 a 10 dB/oitava (vide Fig. II.4). Isto significa que muitos fonemas apresentam uma taxa de cruzamentos de zero relativamente baixa. Este fato pode ser melhor verificado fazendo-se uma análise das formas de onda associadas aos diversos fonemas [13;14]. A Fig. II.5 mostra, por exemplo, a forma de onda associada à vogal 'i', como em "beet". A forma de onda consiste de uma oscilação de baixa frequência e de grande amplitude, à qual se superpõe uma oscilação de alta frequência e de pequena amplitude. Isto está de acordo com a Tabela II.2, onde se tem que a primeira formante é de baixa frequência, e a segunda formante é de frequên

cia bem superior. Neste caso tem-se, pois, uma concentração de energia em baixas frequências e, portanto, a vogal 'i' apresenta uma baixa taxa de cruzamentos de zero.

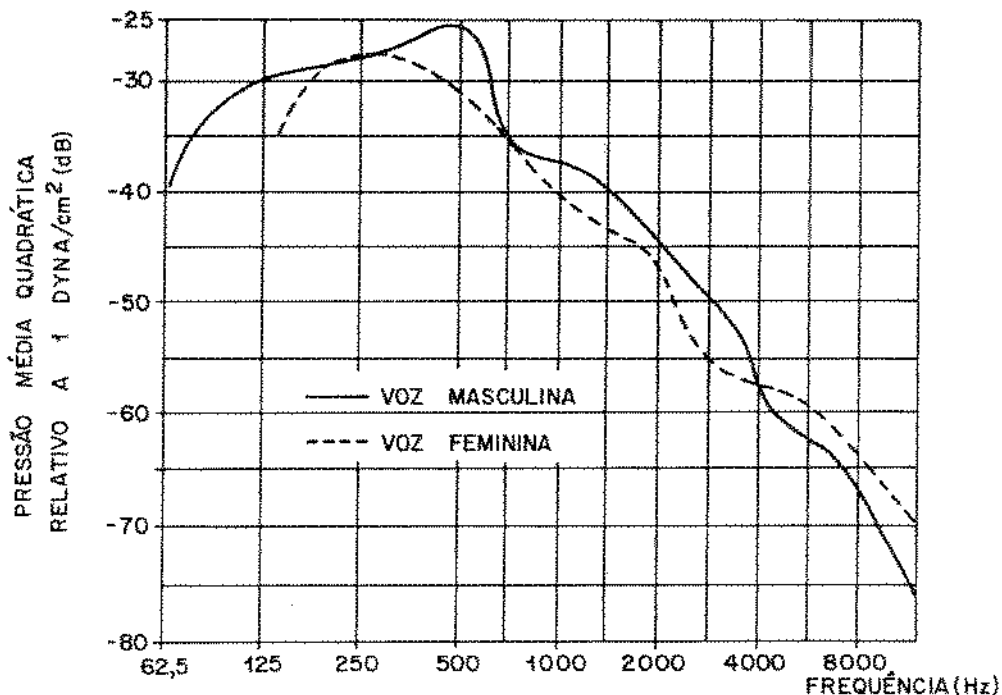


Fig. II.4 - Densidade espectral de potência para o sinal de voz [14]

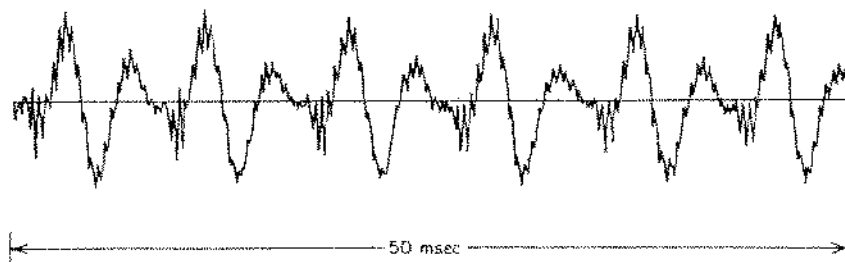


Fig. II.5 - Forma de onda para a vogal 'i' [14]

Fazendo-se uma análise das formas de onda de outros fonemas, é possível classificar os diversos sons correspondentes em termos de amplitude e concentração de energia com a frequência, em

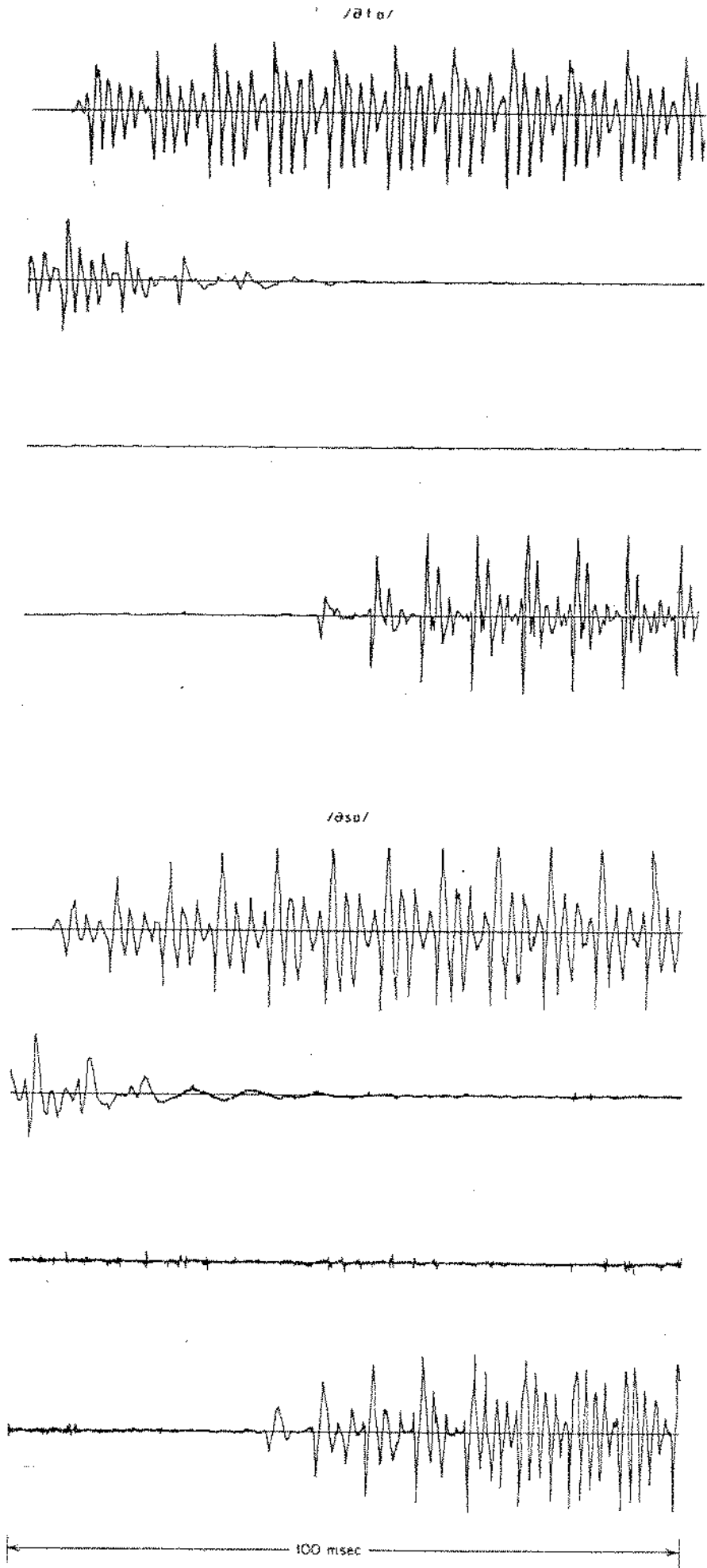


Fig. II.6 - Forma de onda para "F" e "S" [14]

basicamente três categorias:

1. Sons de alta amplitude, com concentração de energia principalmente em baixas frequências (vogais, semi-vogais e as consoantes nasais. Ex.: a ; i ; m).
2. Sons de pequena amplitude com concentração de energia principalmente em baixas frequências (consoantes fricativas e oclusivas sonoras. Ex.: v ; b).
3. Sons de pequena amplitude, com concentração de energia principalmente em altas frequências (consoantes fricativas surdas. Ex.: s ; f ; x).

Desta classificação, conclue-se que todos os fonemas a apresentam baixa taxa de cruzamentos de zero, com exceção das consoantes fricativas surdas. Estas consoantes, além de possuírem uma taxa de cruzamentos de zero relativamente alta, apresentam uma amplitude muito reduzida, da ordem de 20 dB, em comparação com os fonemas sonoras (vide Fig. II.6). Contudo, se as consoantes fricativas surdas passarem por um filtro passa-baixas com uma frequência de corte em torno de 3.400 Hz, a concentração de energia em altas frequências será sensivelmente atenuada. Este fato pode ser melhor visualizado através da Fig. II.7., na qual é apresentada a densidade espectral de energia para o caso típico do "s", quando se pronuncia a palavra "sect".

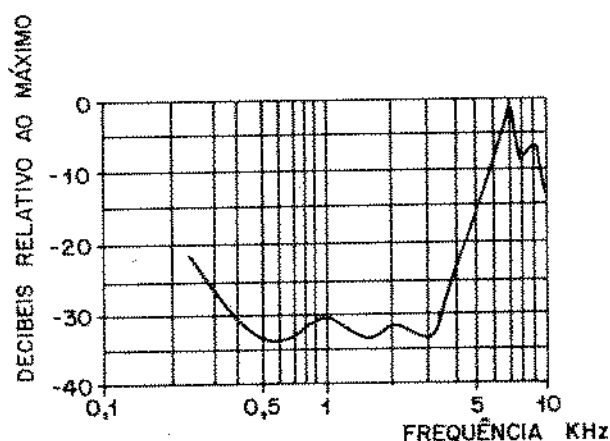


Fig. II.7 - Densidade espectral de energia do "s" como em "sect" [15]

Observe-se que, após uma filtragem passa-baixas, a concentração de energia em baixas frequências será mais acentuada que em altas frequências. Isto é, a sua forma de onda correspondente será um sinal de baixa frequência, ao qual se superpõe um sinal de alta frequência de amplitude menor. Portanto, a passagem por um filtro passa-baixas causará uma diminuição na taxa de cruzamentos de zeros. Entretanto, como a diferença de amplitude entre os sinais de baixa e de alta frequência não é muito acentuada, como ilustra a Fig. II.8, podem ocorrer períodos onde o número de cruzamentos de zero é baixo, como em T_a , bem como períodos onde o número de cruzamentos de zero é elevado, como em T_b .

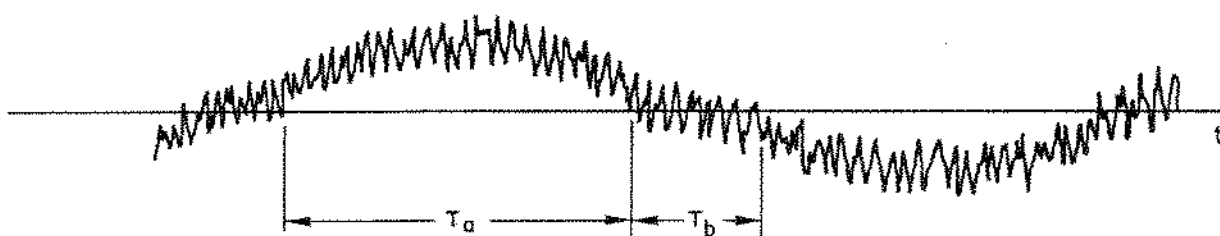


Fig. II.8 - Representação aproximada da forma de onda para o "s", depois de passar por um filtro passa-baixas

Isto ocorre no caso de uma transmissão telefônica, pois o sinal de voz sofre, além da filtragem devido aos cabos e equipamentos multiplex, a filtragem do microfone do aparelho telefônico, o qual apresenta a resposta em frequência mostrada na Fig. II.9-[1].

Algumas observações devem ser feitas também com relação às consoantes oclusivas surdas (Ex.: p ; t). Estas consoantes apresentam três intervalos de tempo distintos: um intervalo de tempo de silêncio, durante o qual ocorre a interrupção do fluxo de ar que vem dos pulmões sem vibrar as cordas vocais. Em seguida, há um breve intervalo de fricção (altas frequências), devido ao escape repentino do ar armazenado no intervalo de tempo anterior. Logo em seguida, ocorre um intervalo de tempo onde predomina a comp

nente de baixa frequência. Contudo, se estas consoantes passarem por um filtro passa-baixas, apresentarão predominantemente baixas frequências, pois os intervalos de fricção serão sensivelmente atenuados.

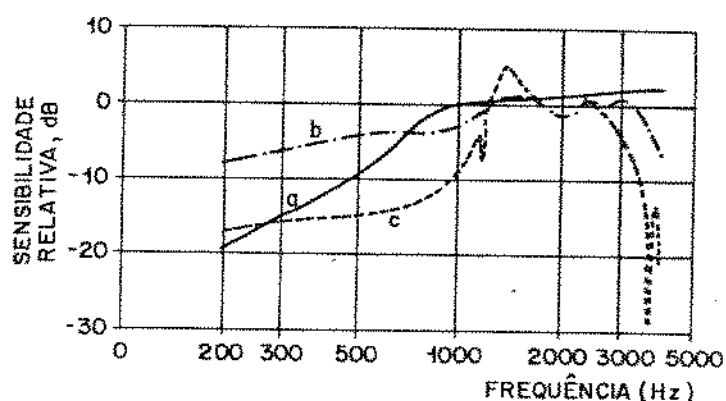


Fig. II.9 - Resposta em frequência do microfone telefônico
a) Objetivo de projeto
b) Microfone de um telefone moderno
c) Microfone de um telefone antigo

Na Fig. II.10 é mostrado um histograma da taxa de cruzamentos de zero para os diversos sons de voz [14].

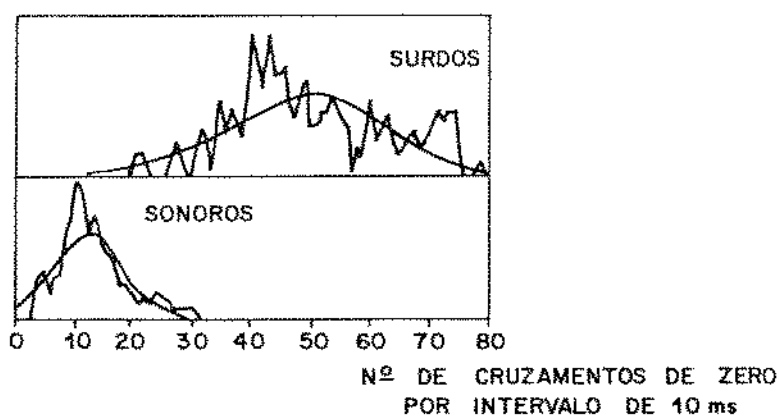


Fig. II.10 - Distribuição do número de cruzamentos de zero para fonemas sonoros e surdos [14]

Note-se que na média tem-se em torno de 1000 cruzamentos de zero/s para os sons da categoria 1 e 2 (sons sonoros), e em torno de 4.500 cruzamentos de zero/s para os sons da categoria 3 (sons surdos).

Considere-se, agora, para efeito de comparação, um ruído gaussiano branco passando através de um sistema linear de função de transferência $C(\omega)$. Demonstra-se [16] que o número médio de cruzamentos de zero por unidade de tempo, apresentado pelo ruído gaussiano na saída de $C(\omega)$ é dado por:

$$\lambda(0) = \frac{1}{\pi} \left(\frac{\int_0^{\infty} \omega^2 C^2(\omega) d\omega}{\int_0^{\infty} C^2(\omega) d\omega} \right)^{1/2} \quad (II.10)$$

Considerando-se $C(\omega)$ um filtro passa-baixas ideal com frequência de corte 3,4 KHz, obtêm-se:

$$\lambda(0) \approx 3.925 \text{ cruzamentos de zero/s}$$

Portanto, a taxa de cruzamentos de zero dos fonemas sonoros é bem menor comparada com a do ruído gaussiano branco de faixa limitada em 3,4 KHz. Esta mesma diferença não é observada quando se trata de fonemas surdos. Entretanto, neste caso, o valor de 4.500 cruzamentos de zero/s corresponde ao valor médio da taxa de cruzamentos de zero de vários fonemas fricativos e oclusivos surdos, e não somente de sons fricativos. O valor médio da taxa de cruzamentos de zero devido somente aos fonemas fricativos surdos (e aos períodos de fricção dos fonemas oclusivos surdos) está situado no extremo direito do histograma, sendo, portanto, maior que o valor médio da taxa de cruzamentos de zero do ruído. O valor médio da taxa de cruzamentos de zero dos períodos de baixa frequência de fonemas oclusivos surdos está situado no extremo esquerdo do histograma, o qual chega a se sobrepor com o histograma de sons sonoros e, portanto, é menor que o valor médio da taxa de cruzamentos de zero do ruído.

Ressalta-se que se os sons fricativos passarem por um

filtro passa-baixas de frequência de corte 3,4 KHz, a taxa de cruzamentos de zero no período T_b (vide Fig. II.8) embora seja alterada, continuará sendo superior à do ruído, pois ainda ficará restando um máximo de concentração de energia em torno de 3,4 KHz.

Um outro parâmetro importante a ser comparado entre o sinal de voz e o ruído é o desvio padrão do número de cruzamentos de zero. No Apêndice A é mostrado que o desvio padrão σ_r do número de cruzamentos de zero, sobre um intervalo de tempo T , do ruído gaussiano branco de faixa limitada em f_0 , é dado por:

$$\sigma_r = \left(\frac{\omega_1 T}{\pi} - \frac{\omega_1^2 T^2}{\pi^2} + \frac{T}{\pi^2} \int_0^T \left(1 - \frac{\tau}{T}\right) \frac{D_{24}}{(1-R^2)^{3/2}} \left[-\arccos \frac{D_{24}}{D_{22}} + \frac{\sqrt{D_{22}^2 - D_{24}^2}}{D_{24}} \right] d\tau \right)^{1/2} \quad (II.11)$$

onde:

T é o intervalo de tempo durante o qual foi medido o número de cruzamentos de zero;

$$D_{22} = \omega_1^2 (1-R^2) - R'^2$$

$$D_{24} = R''(1-R^2) - RR'^2$$

$$\omega_1 = \omega_0/\sqrt{3} = 2\pi f_0/\sqrt{3}$$

$R = \text{sen}(\omega_0 \tau)/\omega_0 \tau$ (coeficiente de correlação do ruído branco de faixa limitada em f_0 .)

$$R' = \frac{dR}{d\tau}$$

$$R'' = \frac{d^2 R}{d\tau^2}$$

Usando-se estas expressões, foi traçada a curva da Fig. II.11, que mostra o desvio padrão σ_r do número de cruzamentos de zero do ruído em função de T , para $f_0=4$ KHz. Tomando - se,

agora, a distribuição do número de cruzamentos de zero do sinal de voz (Fig. II.10), e fazendo-se uma aproximação dos histogramas por uma curva gaussiana, é possível verificar que o desvio padrão do número de cruzamentos de zero do sinal de voz σ_v , é bem maior que o desvio padrão do número de cruzamentos de zero do ruído, σ_r , para um mesmo intervalo de tempo T (Ex.: para T=10 ms, $\sigma_v \gg \sigma_r \approx 3,3$).

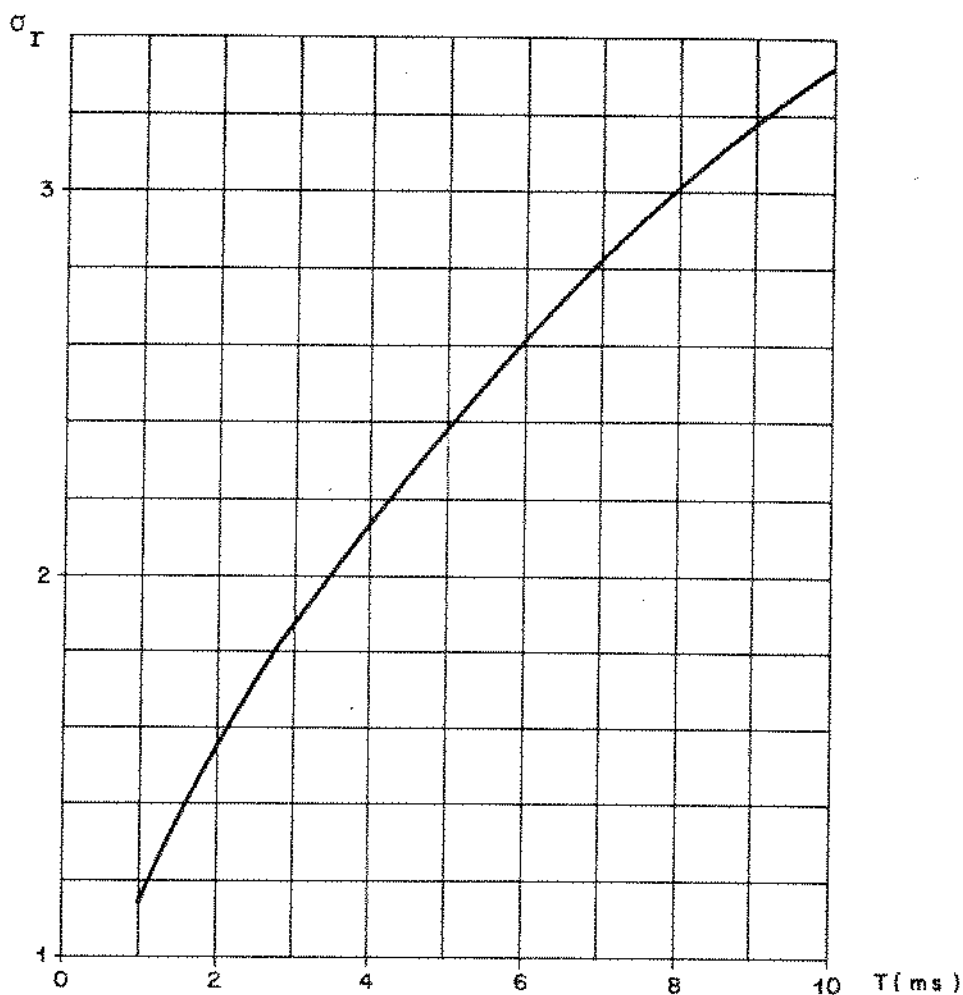


Fig. II.11 - Desvio padrão σ_r do número de cruzamentos de zero sobre um intervalo de tempo T para um ruído gaussiano branco de faixa limitada em $f_0 = 4$ KHz.

II.6 - PERIODICIDADE E VARIACÃO DE AMPLITUDE DA ENVOLTÓRIA

O sinal de voz e o ruído apresentam características bastante diferentes em termos de periodicidade e variação de amplitude da envoltória. Observa-se que as vogais, semi-vogais e as consoantes fricativas sonoras apresentam formas de onda aproximadamente periódicas, cujos picos ocorrem em grupos (período), como ilustra a Fig. II.12.

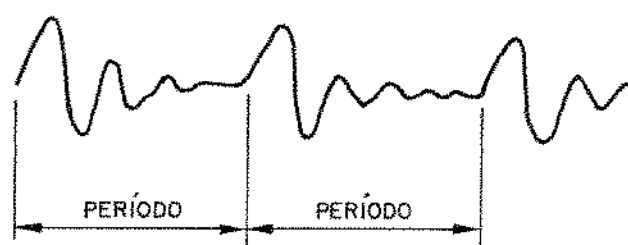


Fig. II.12 - Forma de onda típica de uma vogal

Dentro daqueles grupos, os picos são uniformemente espaçados no tempo, e suas amplitudes decrescem continuamente (e não abruptamente) até o final de cada período do grupo. Além disto, os picos pertencentes a cada grupo são espaçados no tempo de acordo com o seu conteúdo espectral, e sua duração é superior a 1 ms. Por outro lado, as consoantes oclusivas sonoras não são agrupadas como descrito anteriormente. Contudo, mesmo assim, cada pico tem duração de vários milisegundos.

Já o ruído apresenta picos distribuídos de uma maneira totalmente aleatória, e não tem as características observadas para a forma de onda do sinal de voz.

Em termos de variação de amplitude da envoltória, observa-se que a envoltória do sinal de voz apresenta variações maiores que a envoltória do ruído. Isto ocorre porque, como foi dito, numa mesma palavra ou sílaba os diferentes fonemas apresentam amplitudes acentuadamente diferentes entre si. Assim, podem ocorrer grandes variações na amplitude da envoltória entre uma pausa e início de uma palavra, entre vogais e consoantes, entre fonemas sonoros e surdos.

II.7 - RUÍDO AMBIENTAL

Em uma sala qualquer, o ruído ambiente pode ser considerado, de uma maneira geral, como sendo composto de duas partes:

1. ruído produzido por fontes internas e
2. ruído produzido por fontes externas, entrando através de portas ou janelas.

O ruído interno pode ser produzido por pessoas conversando e se movimentando, rádios, máquinas de escrever, ventiladores, etc. O ruído externo é, geralmente, devido ao tráfego de veículos na rua. Estes ruídos podem, ainda, ser classificados em ruídos intermitentes e ruídos contínuos. Pode-se entender como ruído intermitente aquele ruído de pequena duração e que provoca grande variação de nível (o ruído provocado por algum objeto ao cair no chão, bater de porta, etc.). Como ruído contínuo pode ser considerado, por exemplo, o ruído de um ventilador, ruído de tráfego de veículos na rua ou superposição de vários ruídos gerados por diversas fontes ao mesmo tempo, os quais apresentam pequenas variações de nível.

Sob o ponto de vista do detector de voz, os ruídos intermitentes não são muito críticos, pois tais ruídos não causam um aumento apreciável na atividade média. O ruído ambiental contínuo, contudo, já pode provocar um aumento elevado na atividade média, de modo que o detector de voz deve apresentar uma alta imunidade a este tipo de ruído.

Na Fig. II.13 é apresentada uma curva típica da densidade espectral de potência de ruído ambiental monótono [12]. Inúmeras medidas do espectro do ruído ambiental efetuadas em diversos tipos de localidades mostram, para cada um destes tipos de localidades, uma forma do espectro de potência aproximadamente igual ao da Fig. II.13 [17]. Portanto, é razoável assumir que, embora o espectro do ruído ambiental em um determinado lugar possa ser grandemente afetado pela presença de alguma fonte particular de ruído, na maioria dos casos o espectro de potência do ruído ambiental monótono será semelhante àquele apresentado na Fig. II.13.

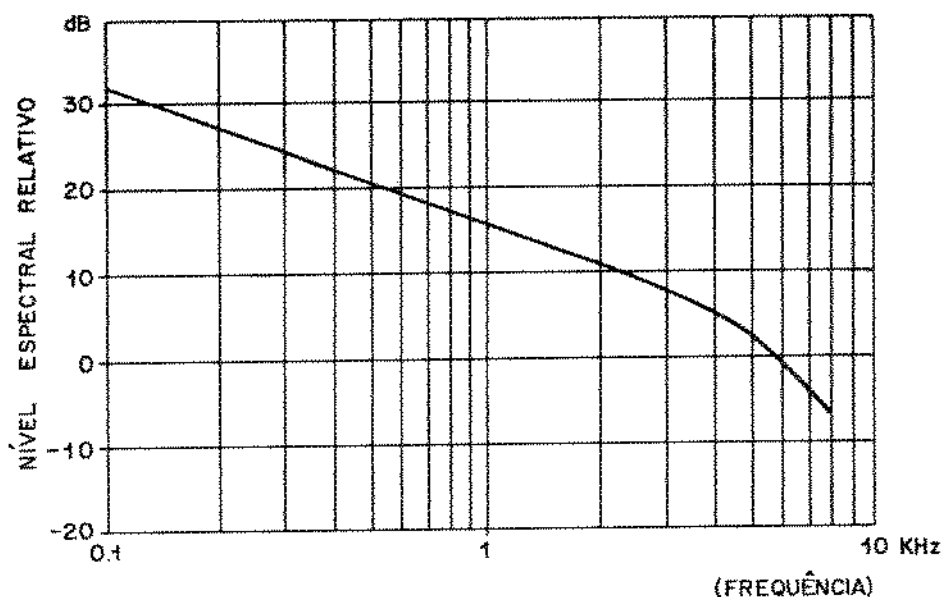


Fig. II.13 - Densidade espectral de potência do ruído ambiental [12]

Uma característica notável que pode ser observada da Fig. II.13, é que o espectro do ruído ambiental está concentrado principalmente em baixas frequências. E uma diferença importante que se observa quando se compara o espectro do ruído ambiental com o do sinal de voz, é que o ruído ambiental apresenta uma concentração de energia em frequências mais baixas ainda do que o sinal de voz e, portanto, apresenta uma taxa de cruzamentos de zero também mais baixa. Entretanto, esta diferença não é muito acentuada quando se compara o ruído ambiental com os sons do tipo sonoro.

Um outro fator importante a ser levado em conta no projeto do detector de voz, é a potência média do ruído ambiental. Na Tabela II.3 são apresentadas algumas medidas do nível de ruído ambiental realizadas por Seacord [18]. Estas medidas cobrem uma grande variedade de tipos de lugar, desde residências muito silenciosas até fábricas bastante ruidosas, e diferentes áreas, desde áreas rurais até áreas urbanas de grandes cidades, como Nova York, Chicago, etc. Na primeira coluna desta tabela são apresentados os valores da potência acústica média, em dB, relativos a 10^{-16} watt/cm², para os diversos tipos de localidade. Na coluna seguinte, são apresentados os valores do desvio padrão da distribuição de ruído entre localidades do mesmo tipo. Na terceira coluna é apresentada

uma estimativa da potência elétrica média do ruído em dBm0. Esta estimativa foi feita levando-se em conta que uma potência acústica de voz de 95 dB (relativa a 10^{-16} watt/cm²) incidente sobre o microfone do aparelho telefônico, produz uma potência elétrica da ordem de -20 dBm0 [1,19]. Finalmente, na quarta coluna é apresentada, para cada tipo de localidade, a faixa de potência na qual está contido o nível médio do ruído ambiental em 98% dos casos.

TIPO DE LOCALIDADE	P.MÉDIA (dB)	DESVIO PADRÃO	P.MÉDIA (dBm0)	P.MÉDIA P/98% DOS CASOS (dBm0)	
Residência	s/ rádio ligado	43	5,5	-72	< -61
	c/ rádio ligado	50	8,0	-65	< -49
Loja pequena (< 6 funcionários)	53,5	7,5	-61,5	< -46,5	
Loja grande (> 5 funcionários)	61	6,0	-54	< -42	
Escritório pequeno (< 3 escrivainhas)	53,5	6,5	-61,5	< -46,5	
Escritório médio (3 a 10 escrivainhas)	58	6,5	-57	< -44	
Escritório grande (> 10 escrivainhas)	64,5	4,5	-50,5	< -41,5	
Escritório de fábrica	61	9,5	-53,5	< -34,5	
Lugares comerciais diversos	56	7,5	-59	< -44	
Fábrica	77	12,0	-38	< -14	

Tab. II.3-Valores médio de ruído ambiental

A partir da Tab. II.3 é possível se observar que o nível do ruído ambiental é bastante elevado em alguns casos, podendo comprometer a atividade média α . Entretanto, considerando-se como caso atípico o ruído ambiental de uma fábrica, e considerando - se também a baixa sensibilidade do microfone do aparelho telefônico (vide Fig. II.9), é razoável impor como objetivo de projeto que o detector de voz seja imune ao ruído ambiental de potência média até -44 dBm0.

CAPÍTULO III

ESTRATÉGIAS PARA DETECÇÃO DO SINAL DE VOZ E
REJEIÇÃO DE RUÍDO

III.1 - INTRODUÇÃO

Em sistemas de interpolação de voz, a parte mais crítica é o detector de voz. De fato, o sucesso destes sistemas depende do bom desempenho do detector de voz, cuja implementação torna-se complexa devido à presença inevitável de ruído nos canais telefônicos. Estes ruídos, em alguns casos, possuem a mesma ordem de magnitude que os sons fricativos como o "s" e "f", os quais possuem um nível extremamente baixo, e uma duração em torno de 200ms.

Devido ao baixo nível dos sons fricativos, um detector de voz convencional [21], com um limiar de -40 dBm0, pode causar cortes no início e mesmo no meio das palavras. Assim, por exemplo, uma palavra como "seca" é ouvida como "eca". Além disto, no caso de um detector de voz convencional, como a potência média de ruído numa linha telefônica real pode se aproximar do limiar de -40 dBm0 em alguns casos, tem-se como consequência um aumento significativo da atividade média α devido à detecção deste ruído.

O problema dos cortes iniciais de sons fricativos pode ser solucionado atrasando-se o sinal de voz em torno de 200ms. Contudo, esta alternativa pode não ser viável na prática devido ao aumento do tempo de propagação, que pode criar novos problemas tanto nas ligações de curta como nas de longa distância. Além disto, mesmo que seja dado um atraso, não se evitará os cortes iniciais nos sinais de voz de nível muito baixo, ou um aumento na atividade devido à detecção de ruídos de nível muito elevado.

Neste capítulo é feita uma descrição geral do detector de voz convencional, e são propostos alguns tipos de detector de voz não-convencional que podem resolver os problemas descritos acima. Em particular, é feita uma descrição detalhada do detector de voz não-convencional que foi desenvolvido para ser utilizado no sistema MCP-60.

III.2 - OBJETIVOS DE PROJETO PARA O DETECTOR DE VOZ

Baseado nos estudos realizados no Capítulo II, as seguintes características de desempenho serão consideradas, no presente trabalho, como objetivos básicos de projeto para o detector de voz:

1. Alta sensibilidade ao sinal de voz, mesmo que seja de potência média* tão baixa quanto -44 dBm0.
2. Rejeição ao ruído gaussiano branco de faixa limitada em $3,4$ KHz e potência média inferior a -35 dBm0.
3. Rejeição ao ruído impulsivo.
4. Rejeição ao ruído ambiental monótono de potência média inferior a -44 dBm0.

Entende-se como detector de voz de alta sensibilidade a aquele que, embora cause cortes iniciais nas palavras, estes cortes são de duração muito curta, não causando efeito perceptível ao ouvido humano. Tendo em vista a constante de tempo do ouvido humano, estes cortes, para não serem notados, devem ter uma duração de no máximo 40 ms [20]. Portanto, o detector de voz deve apresentar um desempenho tal que não somente seja conservada a intelibilidade, mas também não seja notada, subjetivamente, nenhuma diferença entre o sinal de voz detectado e o sinal de voz original.

III.3 - DETECTOR DE VOZ CONVENCIONAL

Os detectores de voz convencionais têm sido utilizados nos sistemas de interpolação de voz do tipo TASI [3,4,21]. Este tipo de detector de voz, ilustrado na Fig. III.5, extrai a envoltória do sinal de entrada e compara a amplitude da envoltória com um limiar fixo L_C . O valor utilizado para este limiar está entre -40 dBm0 e -30 dBm0, tendo em vista os níveis de ruído especificados pelo CCITT (vide Tabela II.1).

O detector de envoltória consiste de um retificador de onda completa seguido de um filtro passa-baixas RC com uma constante de tempo entre 10 e 20 ms, e fornece pois uma indicação da potência média a curto prazo do sinal de voz. Um canal é considerado ativo toda vez que a amplitude da envoltória for maior que o limiar de comparação L_C . Um limiar L_C de -40 dBm0 consiste de um

(*) Deste ponto em diante, quando for dito potência média, ficará implícito potência média a longo prazo para $\alpha_L = 100\%$.

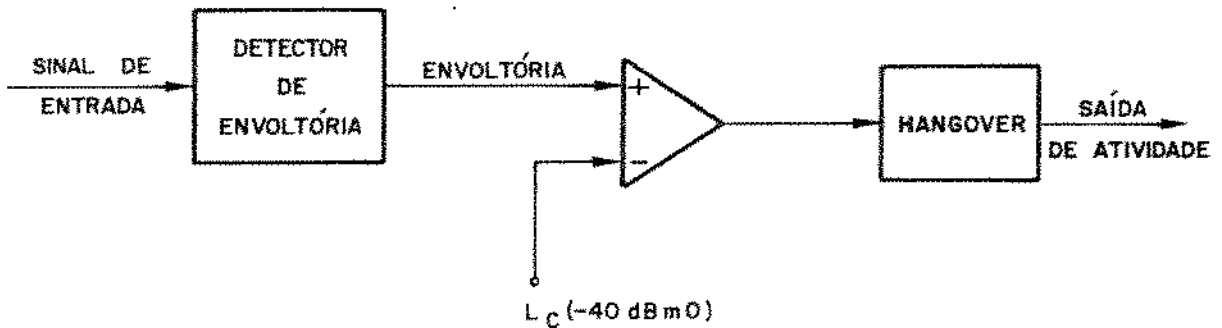


Fig. III.1 - Detector de voz convencional

nível DC correspondente à aplicação de uma senóide de -40 dBm0 ao detector de envoltória.

O comparador é seguido de um circuito de "hangover". Este circuito continua mantendo a indicação de que o canal está ativo durante um certo intervalo de tempo após o nível da envoltória ter caído abaixo de L_C . Este intervalo de tempo, normalmente denominado tempo de "hangover", é dimensionado de modo a evitar cortes abruptos no meio ou no final das palavras, quando a amplitude da envoltória do sinal de voz cai abaixo do limiar L_C . O tempo de "hangover" não deve ser de duração muito longa, pois pode comprometer o valor da atividade média α . Para este tipo de detector de voz, tem sido utilizado um tempo de "hangover" de aproximadamente 240ms.

A sensibilidade do detector de voz convencional torna-se crítica quando o nível do sinal de voz é muito baixo, ou quando se pronuncia, após uma pausa maior que o tempo de "hangover", alguma palavra que se inicia com sons fricativos. Além disto, se o limiar de comparação for fixado em -40 dBm0 , a imunidade ao ruído também torna-se crítica, pois em alguns casos o ruído de linha pode se aproximar deste valor. Para resolver tais problemas, torna-se necessário estabelecer outros critérios, além da potência média a curto prazo do sinal de entrada, para decidir se o canal está ativo ou não. O detector de voz que se utiliza destes critérios adicionais, costuma ser denominado detector de voz não - convencional.

III.4 - DETECTOR DE VOZ NÃO-CONVENCIONAL

Um detector de voz não-convencional utiliza, além do critério da potência do sinal de entrada, outros critérios ou estratégias para decidir se o canal está ativo ou não. Tais estratégias são baseadas nas características do sinal de voz e do ruído presentes num circuito telefônico. De acordo com os estudos apresentados no Capítulo II, estas características podem ser resumidas conforme a Tabela III.1.

PARÂMETROS	RUÍDO		VOZ	
	AMBIENTAL	DE LINHA	SONORO	SURDO
Fator de pico	*	PEQUENO	GRANDE	PEQUENO
Nº médio de cruz. de zero	M.BAIXO	MÉDIO	BAIXO	ALTO/BAIXO
Desvio padrão do nº de cruz. de zero	*	PEQUENO	GRANDE	GRANDE
Variação na amplitude da envoltória	*	PEQUENO	GRANDE	

* PARÂMETROS NÃO ESTUDADOS NO CASO DO RUÍDO AMBIENTAL

Tab. III.1 - Características do sinal de voz e do ruído

Baseado nestas características, é possível conceber várias estratégias para detecção de voz e rejeição de ruído. Algumas destas estratégias são descritas a seguir:

1. DETECTOR DE VOZ COM ANALISADOR DE VARIACÃO DE ENVOLTÓRIA

Na Fig. III.2 é apresentado o diagrama de blocos de um detector de voz que utiliza, além da informação da potência média a curto prazo, a informação da variação de envoltória do sinal de entrada [8]. Este detector de voz foi projetado para ser usado em

sistemas MCP, e portanto processa diretamente as amostras do sinal de voz codificadas sob forma digital.

A variação de envoltória do sinal de entrada é dada por:

$$V_e(nT) = \left| \frac{s(nT)}{s(nT-\Delta t)} \right| \quad (\text{III.1})$$

onde Δt é um atraso apropriado dado à envoltória do sinal.

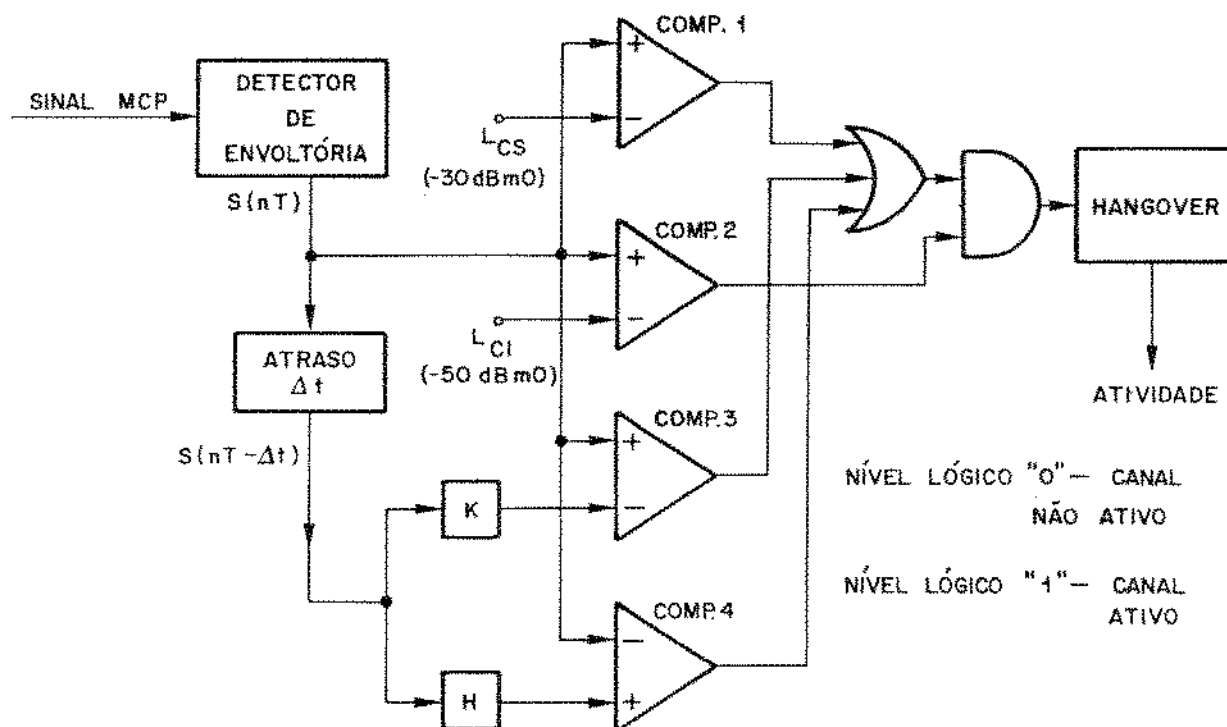


Fig. III.2 - Detector de voz com detector de variação de envoltória [8]

O canal é considerado ativo somente se uma das seguintes desigualdades for satisfeita:

$$V_e(nT) > K \quad , \quad K > 1 \quad (\text{III.2})$$

$$V_e(nT) < H \quad , \quad H < 1 \quad (\text{III.3})$$

Estas desigualdades foram estabelecidas baseadas no fato de que o ruído é um processo aleatório com pequenas variações de envoltória ($H < V_e(nT) < K$), enquanto que o sinal de voz, ao contrário, apresenta grandes variações de envoltória ($V_e(nT) > K$ ou $V_e(nT) < H$). Os valores ótimos de K , H , da constante de tempo RC do detector de envoltória e o tempo de "hangover", podem ser determinados através de testes subjetivos.

Por outro lado, independente da variação de envoltória, o canal será considerado ativo se a sua potência média a curto prazo for maior que o limiar de -30 dBm0, e não-ativo se for menor que o limiar de -50 dBm0

2. DETECTOR DE VOZ COM ANALISADOR DE VARIANÇA DO NÚMERO DE CRUZAMENTOS DE ZERO

Este detector de voz, ilustrado na Fig. III.3 [9], baseia-se no fato de que o ruído apresenta uma pequena variação no número de cruzamentos de zero, durante um dado intervalo de tempo T .

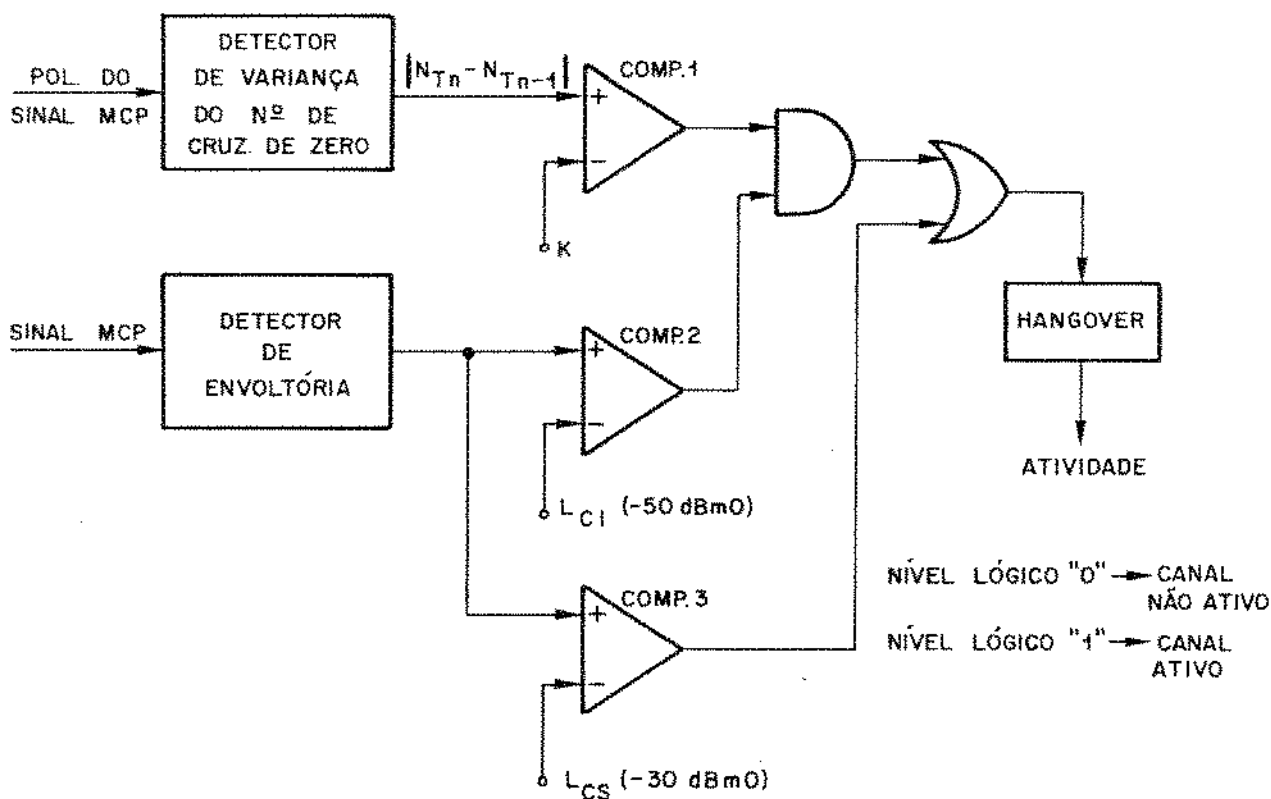


Fig. III.3 - Detector de voz com analisador de variância do nº de cruzamentos de zero

A ocorrência de um cruzamento de zero no sinal digital MCP é facilmente detectada pela ocorrência de inversão do bit de polaridade. Um canal é considerado ativo se satisfizer a seguinte desigualdade:

$$V = |N_{T_n} - N_{T_{n-1}}| > K \quad (\text{III.4})$$

onde: N_{T_n} é o número total de cruzamentos de zero ocorrido durante o intervalo de tempo T_n .

$N_{T_{n-1}}$ é o número total de cruzamentos de zero ocorrido durante o intervalo de tempo anterior T_{n-1} .

Os valores ótimos de K e T_n podem ser determinados através de testes subjetivos.

Independente, contudo, da variação do número de cruzamentos de zero ocorrida entre os intervalos de tempo T_n e T_{n-1} , um canal é considerado ativo se a potência média a curto prazo for maior que o limiar de -30 dBm0, e não-ativo se for menor que o limiar de -50 dBm0.

3. DETECTOR DE VOZ COM ANALISADOR DE NÚMERO MÉDIO DE CRUZAMENTOS DE ZERO

Um diagrama de blocos de um detector de voz que se baseia na análise do número médio de cruzamentos de zero é apresentado na Fig. III.4 [7].

O detector de número de cruzamentos de zero verifica, continuamente, o número N_A de amostras ocorridas a cada K cruzamentos de zero consecutivos. Na Fig. III.5, por exemplo, o número N_A de amostras ocorridas a cada $K=3$ cruzamentos de zero consecutivos é igual a 6,4,8,10, etc.

De acordo com esta estratégia, um determinado canal é considerado ativo somente se uma das seguintes desigualdades for satisfeita:

$$N_{A2} < N_A \leq N_{A3} \quad (\text{III.5})$$

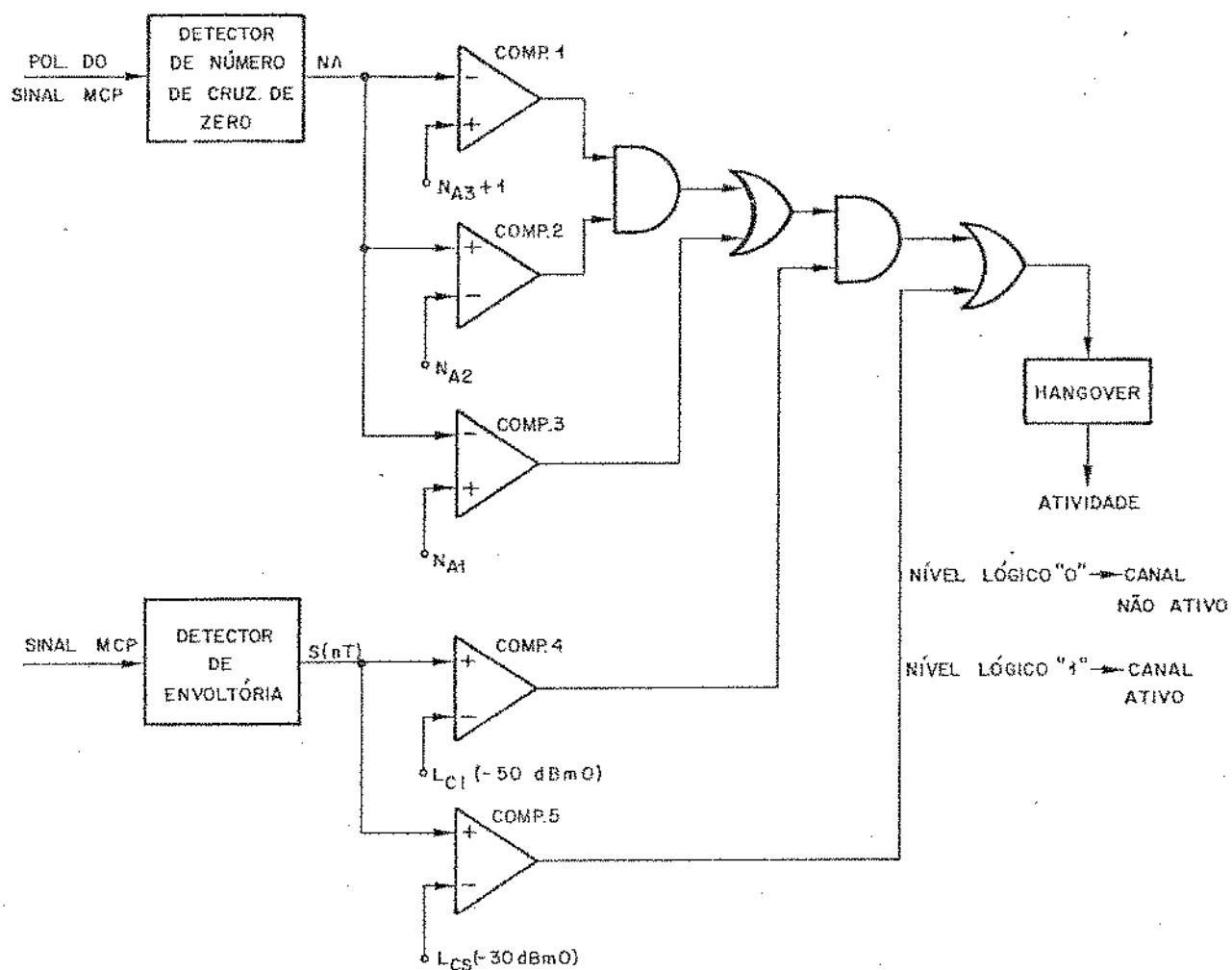


Fig. III.4 - Detector de voz com analisador de número médio de cruzamentos de zero

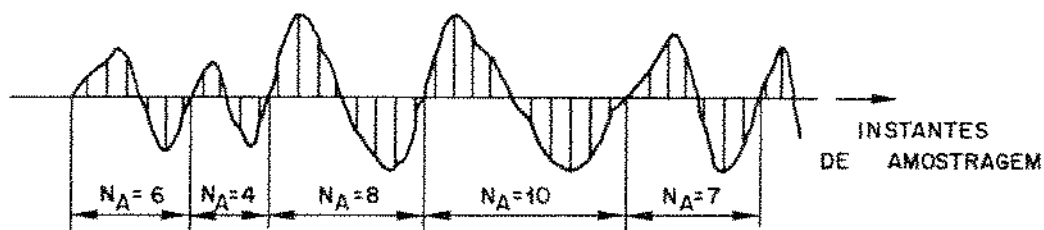


Fig. III.5 - Número de amostras N_A ocorridas a cada $K=3$ cruzamentos de zero consecutivos

ou $N_A < N_{A1}$ (III.6)

onde: N_A é o número de amostras ocorridas a cada K cruzamentos de zero consecutivos

$$N_{A3} > N_{A2} > N_{A1}$$

Estas desigualdades são baseadas no fato de que o sinal de voz consiste de sons sonoros que são de baixa frequência, e sons fricativos surdos que são de alta frequência. Os sons sonoros apresentam um número N_A de amostras relativamente elevado (entre N_{A2} e N_{A3}) a cada K cruzamentos de zero, enquanto que para os sons fricativos surdos este número é relativamente baixo (menor que N_{A1}). Para o ruído, por outro lado, o número N_A de amostras a cada K cruzamentos de zero é intermediário entre N_{A1} e N_{A2} . Além disto, se o número N_A de amostras for extremamente elevado, maior do que N_{A3} , o sinal de entrada será considerado como ruído, pois neste caso pode se tratar de algum tipo particular de ruído de frequência muito baixa, como indução de 60 Hz da rede de energia elétrica.

Os valores ótimos de K , N_{A1} , N_{A2} e N_{A3} podem ser determinados através de testes subjetivos.

Analogamente aos casos anteriores, independente do valor de N_A , um canal é considerado ativo se a sua potência média a curto prazo exceder o limiar de -30 dBm0, e não ativo se for inferior ao limiar de -50 dBm0.

III.5 - DETECTOR DE VOZ Nº 1

Além das estratégias descritas na seção anterior, várias outras podem ser concebidas baseadas nas características do sinal de voz apresentadas na Tab. III.1. Entretanto, neste trabalho, tendo em vista que o detector de voz deve funcionar em compartilhamento de tempo para vários canais, optou-se pela terceira estratégia descrita acima que, em princípio, é a que apresenta menor complexidade de implementação. Utilizando-se, então, esta estratégia, foi implementado e testado o detector de voz Nº 1, o qual se encontra ilus

trado na Fig. III.6. O mesmo é totalmente digital e processa diretamente o sinal MCP de 8 bits/amostra.

O diagrama de blocos do detector de voz Nº 1 é idêntico ao da Fig. III.4, com exceção do detector de envoltória que foi substituído por um detector de pico a fim de simplificar a implementação, eliminando-se a utilização de um filtro digital.

O detector de pico consiste de dois comparadores e de um contador. O primeiro comparador possui um limiar de comparação L_{CS} , colocado numa região de alta probabilidade de ser ultrapassado pelo sinal de voz e de baixa probabilidade de ser ultrapassado pelo ruído. O contador conta o número p de amostras de um dado canal que ultrapassam o limiar de comparação L_{CS} entre dois cruzamentos de zero consecutivos, e fornece este número ao segundo comparador. Observe-se que, como o sinal de voz possui um fator de pico maior que o do ruído, para uma mesma potência média o detector de pico apresenta uma maior sensibilidade ao sinal de voz do que ao ruído. Além disto, para aumentar a imunidade ao ruído sem diminuir, porém, a sensibilidade ao sinal de voz, o detector de pico possui uma estratégia segundo a qual um canal é considerado ativo somente se o número p for maior do que um valor p_1 pré-fixado (vide exemplo da Fig. III.5). Esta estratégia baseia-se no fato de que os picos do sinal de voz têm uma duração de várias amostras MCP, e que o mesmo não ocorre para o ruído. Testes subjetivos indicaram que ao se passar de $p_1=0$ para $p_1=3$, a sensibilidade do detector de pico ao sinal de voz praticamente não se altera, enquanto que a potência média do ruído necessária para engatilhar o detector de pico deve ser acrescida de aproximadamente 10 dB.

O limiar do detector de pico, escolhido de modo a alcançar um bom compromisso entre a sensibilidade ao sinal de voz e a imunidade ao ruído, é igual ao valor de pico de uma senóide de -32 dBm0 de potência média. O código correspondente a este limiar é X0100010 para sistemas MCP com codificação segundo a Lei A de compressão.

Sob estas condições (limiar de -32 dBm0, e $p=4$ amostras acima do limiar entre dois cruzamentos de zero sucessivos para o canal ser considerado ativo), testes subjetivos mostraram que o detector de pico é totalmente imune ao ruído gaussiano branco de

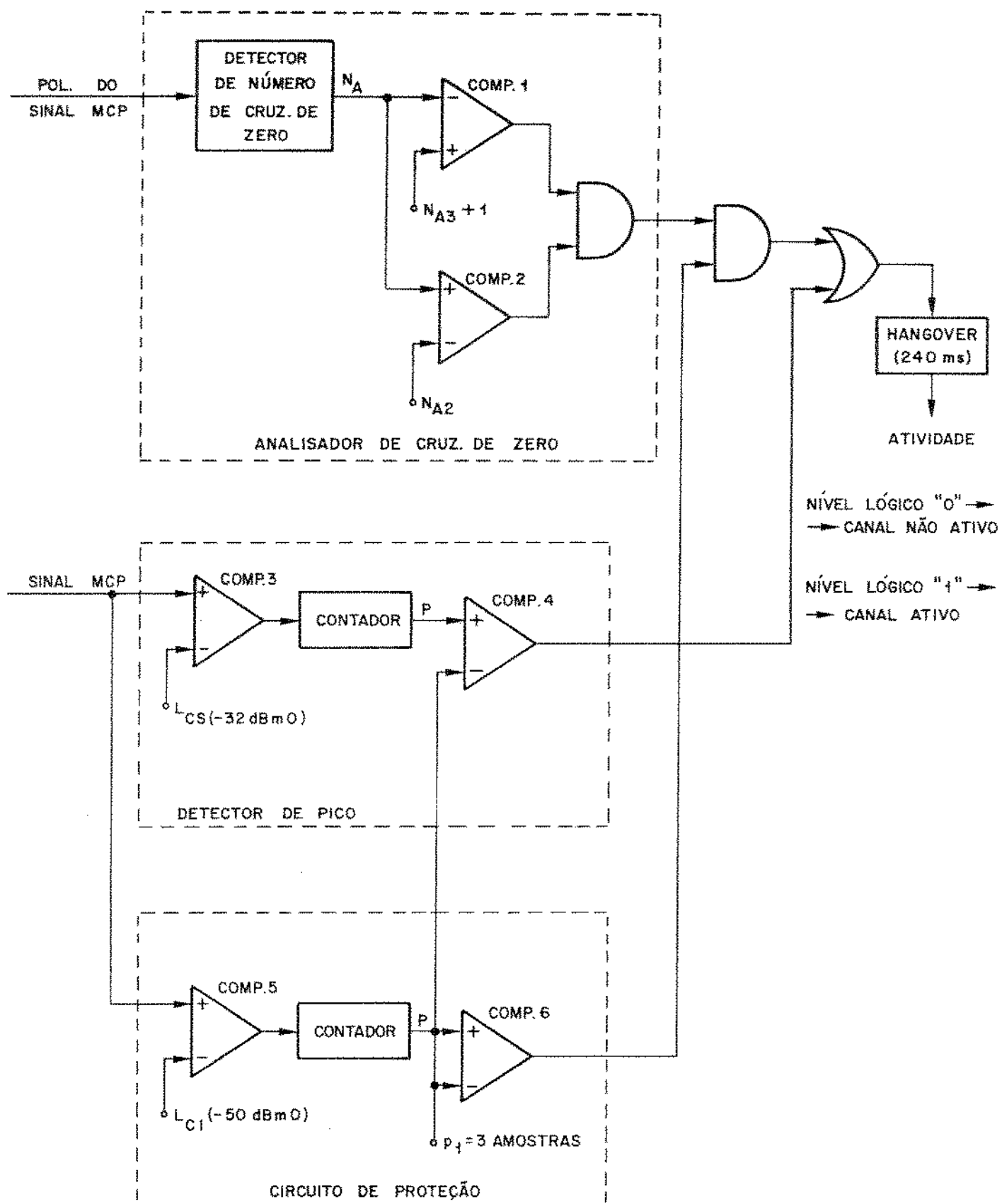


Fig. III.6 - Detector de voz Nº 1

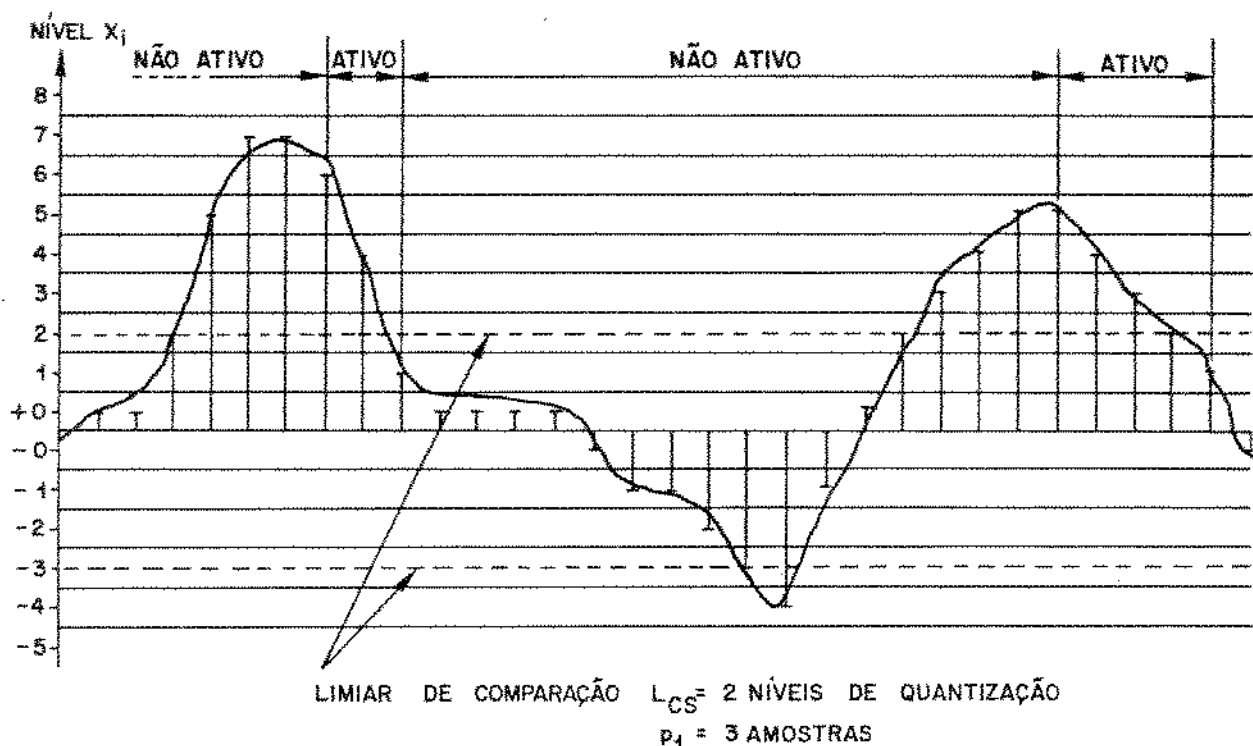


Fig. III.7 - Estratégia do detector de pico

faixa limitada em 3,4 KHz, e de potência média até -34 dBm0, enquanto que a sua sensibilidade ao sinal de voz é considerável. Contudo, como acontece no caso do detector de envoltória, o detector de pico também não apresenta uma boa sensibilidade quando o sinal de voz possui uma potência média muito baixa (como -40 dBm0), ou quando ocorrem sons fricativos. Para solucionar estes problemas, o detector de voz Nº 1 é dotado de um analisador de cruzamentos de zero idêntico ao da Fig. III.4.

Os valores ótimos de K , N_{A1} , N_{A2} e N_{A3} foram determinados através de vários testes subjetivos como sendo:

- $K = 3$ cruzamentos de zero
- $N_{A1} =$ inexistente
- $N_{A2} = 22$ amostras
- $N_{A3} = 33$ amostras

Portanto, de acordo com a estratégia do analisador de cruzamentos de zero, descrito na seção anterior, um determinado canal é consi

derado ativo somente se:

$$22 \text{ amostras} < N_A \leq 33 \text{ amostras}$$

Estes valores ótimos estão de acordo com o esperado teoricamente, pois a maioria dos fonemas possui uma taxa de cruzamentos de zero baixa (em torno de 1.000 cruz. de zero/s) comparada com a do ruído (≈ 4000 cruz. de zero/s). Este fato é melhor verificado comparando-se o número médio de cruzamentos de zero para o sinal de voz e para o ruído, no intervalo de tempo $N_{A2} \cdot T < \tau \leq N_{A3} \cdot T$, onde T é o intervalo entre as amostras MCP e é igual a $125 \mu s$. Como $N_{A2} = 22$ amostras e $N_{A3} = 33$ amostras, resulta que $2,75 \text{ ms} < \tau \leq 4,125 \text{ ms}$, e o número médio m_V de cruzamentos de zero para o sinal de voz neste intervalo de tempo é de:

$$2,75 < m_V \leq 4,125$$

O número médio m_r de cruzamentos de zero para o ruído gaussiano branco de faixa limitada em $3,4 \text{ KHz}$, no mesmo intervalo de tempo acima, é de:

$$11,00 < m_r \leq 16,50$$

Portanto, como os valores de m_V estão muito mais próximos de $K=3$ do que m_r , a estratégia implementada com os parâmetros ajustados segundo os valores acima, é muito mais sensível ao sinal de voz do que ao ruído.

Com relação ao valor ótimo de N_{A1} , os testes subjetivos indicaram que não existe tal valor. Isto já era esperado, pois, como foi concluído no Capítulo II, os sons fricativos, ao passarem por um filtro passa-baixas de frequência de corte em torno de $3,4 \text{ KHz}$, tem a sua taxa de cruzamentos de zero bastante reduzida. Assim, os sons fricativos são detectados pelo analisador de cruzamentos de zero quando ocorrem os períodos em que o nº de cruzamentos de zero é baixo (período T_a na Fig. II.8).

A Fig. III.8 ilustra a estratégia do analisador de cruzamentos de zero. Segundo a estratégia, o canal é considerado ativo quando:

1. Ocorrer um e somente um cruzamento de zero no intervalo de 0 a 22 amostras e um ou mais cruzamentos de zero no intervalo de 22 a 33 amostras (sinal (A) na Fig. III.8).
2. Não ocorrer nenhum cruzamento de zero no intervalo de 0 a 22 amostras, mas ocorrer dois ou mais cruzamentos de zero no intervalo de 22 a 33 amostras (sinal (B) na Fig. III.8).

Para todas as outras possibilidades, o canal é considerado não-ativo. Observe-se que o detector de voz considera como sinal de voz não somente os sinais que apresentam um baixo número de cruzamentos de zero, como é o caso do sinal (A), mas também aqueles sinais que apresentam uma grande probabilidade de ter um baixo número de cruzamentos de zero no primeiro intervalo de tempo de 0 a 22 amostras, e um alto número de cruzamentos de zero no segundo intervalo de tempo de 22 a 33 amostras, como é o caso do sinal (B). Um caso típico do sinal (B) é o sinal correspondente ao "s", pois o mesmo apresenta (vide Fig. II.8) um período em que o nº de cruzamentos de zero é baixo (período T_a), seguido por um período em que o nº de cruzamentos de zero é elevado (período T_b).

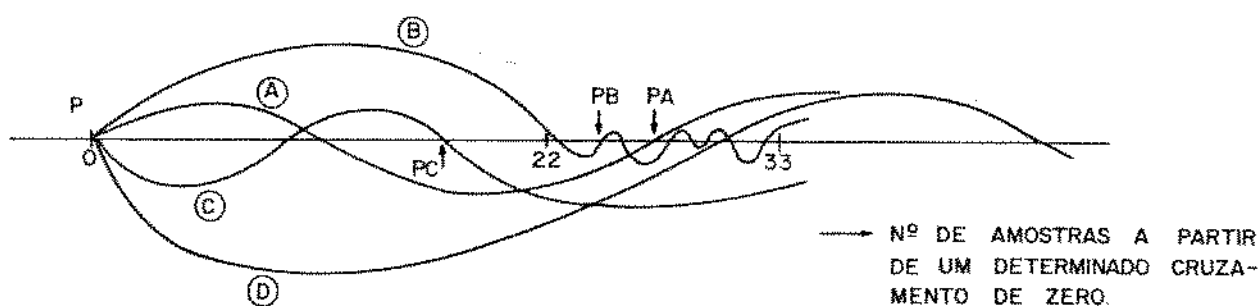


Fig. III.8 - Estratégia para detecção de voz e rejeição de ruído em função do nº de cruzamentos de zero

A verificação da estratégia é sempre iniciada a partir

da ocorrência de um cruzamento de zero, e terminada somente depois da ocorrência de mais dois cruzamentos de zero consecutivos. No instante em que ocorrer o terceiro cruzamento de zero, o analisador de cruzamentos de zero decide se o sinal presente é voz ou ruído e, ao mesmo tempo, inicia uma nova verificação. No caso do sinal (B) da Fig. III.6, por exemplo, a verificação é iniciada no ponto PO e terminada no ponto PB. No ponto PB, contudo, ao mesmo tempo que é decidido se o sinal presente é voz ou ruído, inicia-se uma nova verificação, e assim sucessivamente.

Observe-se que o sinal (C) é considerado ruído, pois N_A , neste caso, é menor que $N_{A2} = 22$ amostras. Analogamente, o sinal (D) também é considerado ruído, pois N_A é maior que $N_{A3} = 33$ amostras.

Note-se que dificilmente ter-se-á um sinal de voz que não apresente pelo menos três cruzamentos de zeros no intervalo de 0 a 33 amostras. Portanto, no pior caso serão necessários 4,125ms para se fazer uma verificação. Tendo em vista que a constante de tempo do ouvido humano é da ordem de 40 ms [20], aproximadamente 10 tentativas podem ser feitas para a detecção do sinal de voz sem causar uma degradação sensível subjetivamente na sua qualidade.

Para a estratégia em termos de cruzamentos de zero, uma senóide de frequência f é considerada como sinal de voz somente se:

$$f_1 \leq f < f_2 ,$$

onde:

$$f_1 = \frac{1}{N_{A3} \cdot T}$$

$$f_2 = \frac{1}{N_{A2} \cdot T} \quad \text{e} \quad T = 125\mu\text{s}.$$

Para $N_{A2} = 22$ amostras e $N_{A3} = 33$ amostras, obtém-se 242 Hz $\leq f < 364$ Hz. Isto significa que a faixa de detecção do sinal de voz (faixa de frequências onde deve ocorrer uma concentração de energia no espectro do sinal de entrada para que este seja considerado como voz pelo Analisador de Cruz. de Zero) é de 242 Hz a

364 Hz. Esta faixa de detecção do sinal de voz está de acordo com o esperado teoricamente, pois corresponde aproximadamente à faixa do espectro do sinal de voz onde ocorre o pico de concentração de energia.

A atividade final é dada, independentemente, tanto pelo detector de pico como pelo analisador de cruzamentos de zero, de modo a aumentar a probabilidade de detectar o sinal de voz.

Usando-se um tempo de "hangover" de aproximadamente 240 ms, constatou-se através de testes subjetivos, que a estratégia descrita é extremamente sensível ao sinal de voz mesmo de potência extremamente baixa, e totalmente imune ao ruído gaussiano branco de faixa limitada em 3,4 Hz. Contudo, esta alta sensibilidade ao sinal de voz pode criar novos problemas, como detecção de eco e de ruído ambiental, pois, o eco nada mais é que um sinal de voz de potência muito baixa, e o ruído ambiental apresenta um espectro concentrado em baixas frequências (vide seção II.7). Para minimizar estes problemas, o detector de voz consta, ainda, de um circuito de proteção cuja saída somente permite que o analisador de cruzamentos de zero seja engatilhado se o sinal de entrada exceder um determinado limiar L_{C1} mínimo, segundo uma estratégia idêntica ao detector de pico descrito anteriormente. Como a probabilidade de ocorrer um sinal de voz de nível menor que -44 dBm0 é muito pequena, da ordem de 10^{-3} , escolheu-se um limiar L_{C1} de proteção igual ao valor de pico de uma senóide de potência média igual a -50 dBm0.

Vários testes subjetivos indicaram que este detector de voz é extremamente sensível ao sinal de voz mesmo de potência média tão baixa quanto -45 dBm0, excetuando-se os casos de palavras que se iniciam com os sons fricativos surdos "f", "s" e "x", como em "fila", "silvar" e "xícara". Dos três casos, o mais crítico é o "x", o qual na maioria das vezes sofre um corte total, pois não é detectado pelo analisador de cruzamentos de zero pelo fato de seu espectro estar centrado na parte superior da faixa do canal telefônico. Para que esta consoante não seja cortada, é necessário que a sua amplitude ultrapasse o limiar de comparação L_{CS} do detector de pico pelo menos 4 vezes entre dois cruzamentos de zero consecutivos, o que exige que a palavra que o contém atinja uma

potência média de aproximadamente -20 dBm0. Por outro lado, as consoantes fricativas surdas "f" e "s" já não são tão críticas. Os testes subjetivos mostraram que estas consoantes começam a sofrer cortes parciais quando a potência média da palavra que os contém cai abaixo de aproximadamente -36 dBm0, e que os cortes totais ocorrem em apenas em torno de 10% dos casos. Tem-se observado, também, que estes cortes devem-se muitas vezes ao circuito de proteção inibindo a saída do analisador de cruzamentos de zero, pois estas consoantes, tal como a consoante "x", apresentam uma amplitude muito reduzida comparativamente aos outros fonemas (vide Capítulo II). Além disto, verificou-se que o desempenho deste detector de voz é muito sensível à variação de largura de faixa do canal telefônico. Isto ocorre porque a faixa de detecção do sinal de voz está em torno da frequência de corte inferior do canal telefônico, de modo que qualquer variação nesta frequência de corte pode influenciar no desempenho do analisador de cruzamentos de zero.

Com relação à imunidade ao ruído, o detector apresenta uma rejeição total ao ruído gaussiano branco de faixa limitada em 3,4 KHz, e de potência média de até -34 dBm0. Entretanto, esta mesma imunidade não é apresentada quando o sinal de entrada é um ruído ambiental de alto nível, devido ao baixo valor do limiar do circuito de proteção (em torno de -50 dBm0).

Esta estratégia usada no detector de voz Nº 1, apesar de apresentar os problemas descritos acima, constitui-se na estratégia básica do detector de voz Nº 2, descrito a seguir, e que foi desenvolvido objetivando contornar todos os problemas acima citados.

III.6 - DETECTOR DE VOZ Nº 2

A estratégia do detector de voz Nº 2 tem como base a estratégia do detector de voz Nº 1. Para contornar os problemas apresentados pelo detector de voz Nº 1, foram introduzidas algumas estratégias adicionais no analisador de cruzamentos de zero e no circuito de proteção.

A estratégia adicional introduzida no analisador de cruzamentos de zero, tem por finalidade tornar o desempenho do detector de voz menos dependente da variação da faixa do canal telefônico, e melhorar a detecção dos sons fricativos. O problema da influência da variação da faixa do canal telefônico pode ser minimizado se a faixa de detecção do sinal de voz for estendida para frequências mais altas. Para conseguir isto, considere-se inicialmente o analisador de cruzamentos de zero descrito na seção anterior. Como $m_r > m_v$, se os valores de N_{A2} e N_{A3} forem redimensionados de modo a tornar o analisador sensível não mais do sinal de voz, mas sim ao ruído, obter-se-á uma faixa de detecção de ruído situada em frequências bem superiores do que a faixa de detecção do sinal de voz. Além disto, como σ_r é muito pequeno, será possível usar f_1 e f_2 relativamente próximos (faixa de detecção de ruído estreita) e um tempo de "hangover" pequeno sem que o ruído sofra algum corte. De fato, através de testes experimentais, determinou-se que para $f_1 = 888$ Hz e $f_2 = 2666$ Hz (isto é, $N_{A2} = 3$ e $N_{A3} = 9$), e um tempo de "hangover" de 9 ms, o ruído gaussiano de faixa limitada em 3,4 KHz (ou mesmo em 2,5 KHz) é detectado continuamente sem sofrer nenhum corte, enquanto que a sensibilidade ao sinal de voz é extremamente baixa. Portanto, se a saída do circuito de "hangover" for invertida, tem-se como resultado um analisador de cruzamentos de zero bastante imune ao ruído e altamente sensível ao sinal de voz. Este analisador considera como voz todo sinal que apresentar uma concentração de energia na faixa de frequência de 0 a 888 Hz, e acima de 2666 Hz. Entretanto, devido ao valor elevado de σ_v , podem ocorrer cortes abruptos no meio e no final das palavras. Para evitar estes cortes, torna-se necessário utilizar um outro circuito de "hangover" como ilustra a Fig. III.9. O circuito de "hangover" com a saída invertida impede, durante um determinado intervalo de tempo, que o ruído ative o circuito de "hangover" que lhe segue, razão pela qual será denominado de agora em diante como Inibidor de Ruído. Isto é, toda vez que aparecer um pulso positivo na sua entrada, indicando que o sinal presente é ruído, o inibidor de ruído impede que o circuito de "hangover" se ja ativado durante um intervalo de tempo (tempo de inibição de ruído) de 9 ms. Desta maneira, o circuito de "hangover" será ativado somente se a entrada do inibidor de ruído permanecer no nível lô

gico "0" (presença de sinal de voz) durante um intervalo de tempo maior do que o tempo de inibição de ruído de 9 ms.

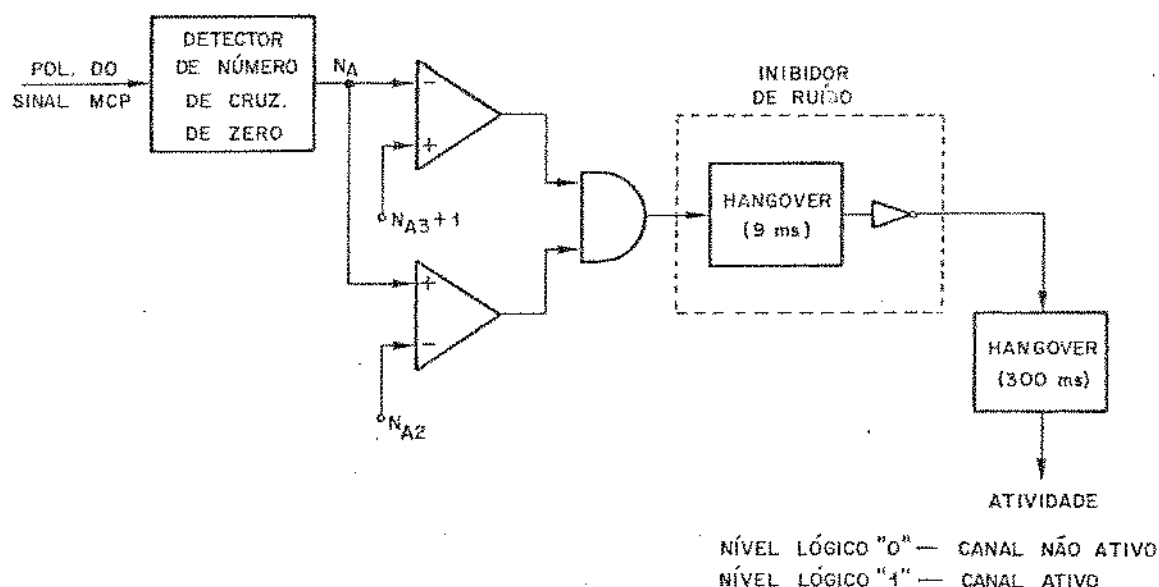


Fig. III.9 - Analisador de cruzamento de zero com inibidor de ruído

Quanto maior for o tempo de inibição de ruído, mais imune ao ruído será o analisador de cruzamentos de zero. Observe-se, entretanto, que o circuito inibidor de ruído causa um corte inicial no sinal de voz durante um intervalo de tempo igual ao tempo de inibição de ruído. Assim, o tempo de inibição de ruído não deve ser de duração muito longa. Observe-se, ainda, que a sensibilidade e a imunidade do analisador de cruzamentos de zero dependem, simultaneamente, do tempo de inibição de ruído e das frequências f_1 e f_2 . Quanto mais próximas forem f_1 e f_2 , maior deverá ser o tempo de inibição para manter a imunidade ao ruído; e quanto menor for o tempo de inibição de ruído, mais afastados devem estar f_1 e f_2 . Portanto, as frequências $f_1 = 888$ Hz e $f_2 = 2666$ Hz, e o tempo de inibição de 9 ms foram determinados, subjetivamente, de modo a alcançar um compromisso entre a imunidade ao ruído e a sensibilidade ao sinal de voz. Para estes valores, a imunidade ao ruído é bastante elevada, e o corte inicial de 9 ms nas palavras causado pelo circuito inibidor de ruído é totalmente imperceptível.

O tempo ótimo de "hangover" foi determinado através de

testes subjetivos, como sendo de aproximadamente 300 ms. Observe-se que, agora, a faixa de detecção do sinal de voz foi deslocada de 346 Hz para 888 Hz. Com isto, os testes experimentais mostram que o desempenho do analisador de cruzamentos de zero com inibidor de ruído é totalmente independente de variações na faixa do canal telefônico de até em torno de 200 Hz. Além disto, note-se que este analisador de cruzamentos de zero considera adicionalmente como sinal de voz, todo sinal que apresentar uma concentração de energia na faixa de frequências acima de 2666 Hz. Isto tem melhorado consideravelmente a detecção dos sons fricativos, pois os mesmos podem apresentar picos de concentração de energia tanto em baixas quanto em altas frequências (como o "s" e "f"), ou somente em altas frequências (como o "x").

O analisador de cruzamentos de zero com inibidor de ruído exige, entretanto, um tempo de "hangover" demasiadamente longo. Para poder operar com um tempo de "hangover" menor, o detector de voz N° 2 utiliza dois analisadores de cruzamentos de zero em conjunto, com estratégias ligeiramente diferentes entre si, de modo a aumentar a probabilidade de detectar o sinal de voz, conforme ilustra a Fig. III.10. Enquanto que o primeiro analisador de cruzamentos de zero examina o número N_A de amostras, a cada $K=3$ cruzamentos de zero, o segundo analisador examina o número N'_A de amostras a cada $K=4$ amostras. Os valores ótimos de N'_{A2} e N'_{A3} foram determinados como sendo:

$$N'_{A2} = 5$$

$$N'_{A3} = 14$$

Para o segundo analisador de cruzamentos de zero, tem-se que:

$$f'_1 = \frac{1,5}{N'_{A3} \cdot T}$$

$$f'_2 = \frac{1,5}{N'_{A2} \cdot T} \quad , \text{ onde } T = 125\mu\text{s}.$$

Portanto, as duas faixas de detecção do sinal de voz do segundo

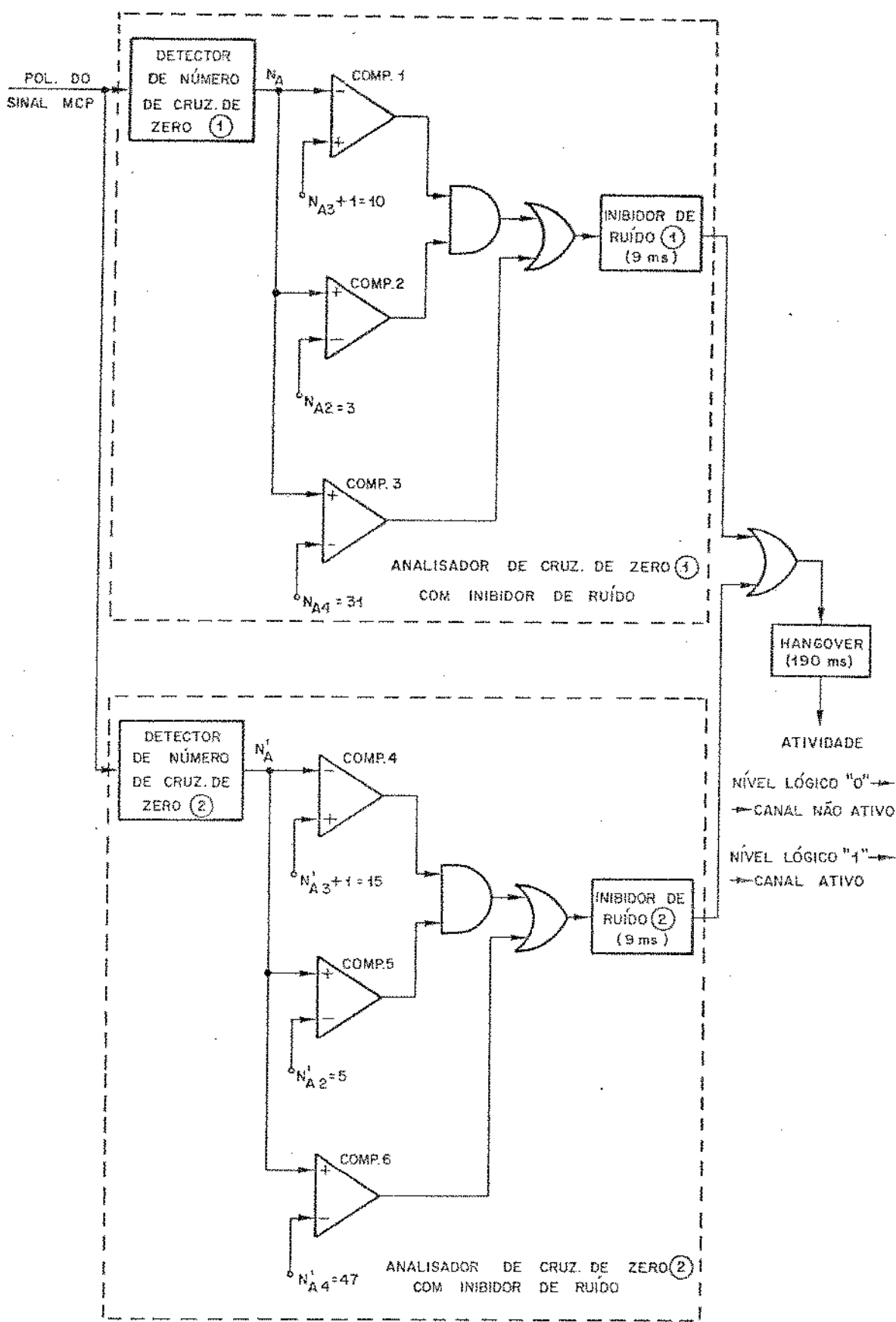


Fig. III.10 - Dois analisadores de cruz. de zero com inibidor de ruído operando em conjunto para reduzir o tempo de "hangover"

analisador de cruzamentos de zero são: $0 < f' < 857 \text{ Hz}$ e $f' > 2400 \text{ Hz}$. Como pode ser observado, as faixas de detecção do sinal de voz do primeiro e do segundo analisador de cruzamentos de zero são praticamente iguais e, além disto, ambos os analisadores apresentam também o mesmo desempenho. Contudo, a probabilidade de ocorrer um corte no sinal de voz simultaneamente nos dois analisadores é bem menor do que a probabilidade de ocorrer um corte no sinal de voz em um só dos analisadores. Desta maneira, com o uso de dois analisadores, conseguiu-se reduzir o tempo de "hangover" de 300 ms para 190 ms.

Tendo em vista que o espectro do sinal de voz cai rapidamente em frequências muito baixas, foi utilizada uma estratégia adicional que considera como sendo ruído todo sinal que apresentar números de amostras N_A e N'_A muito elevados, maiores do que N_{A4} e N'_{A4} , respectivamente. Em outros termos, o primeiro e o segundo analisador de cruzamentos de zero consideram também como ruído todo sinal que apresentar uma concentração de energia em frequências inferiores respectivamente a:

$$f_3 = \frac{1}{N_{A4} \cdot T}$$

$$f'_3 = \frac{1,5}{N'_{A4} \cdot T}, \quad \text{onde } T = 125\mu\text{s}.$$

Considerando-se agora os dois analisadores de cruzamentos de zero, e os circuitos de inibição de ruído associado a cada um deles, um determinado sinal é considerado como voz se satisfizer pelo menos uma das desigualdades abaixo durante um intervalo de tempo superior a 9 ms:

$$N_{A3} < N_A \leq N_{A4}$$

$$N_A \leq N_{A2}$$

$$N'_{A3} < N'_A \leq N'_{A4}$$

$$N'_A \leq N'_{A2},$$

onde: $N_{A2} < N_{A3} < N_{A4}$

$N'_{A2} < N'_{A3} < N'_{A4}$

Para satisfazer estas condições, é necessário que o sinal de entrada apresente uma concentração de energia em pelo menos uma das faixas de frequência abaixo discriminadas:

- | | |
|--|---------------------------------|
| a) $250 \text{ Hz} \leq f < 888 \text{ Hz}$ | } Analisador de cruz. de zero 1 |
| b) $f \geq 2666 \text{ Hz}$ | |
| c) $252 \text{ Hz} \leq f' < 857 \text{ Hz}$ | } Analisador de cruz. de zero 2 |
| d) $f' \geq 2400 \text{ Hz}$ | |

Para melhorar a imunidade do detector de voz ao ruído ambiental monótono de alto nível, pode-se usar o fato de que o mesmo apresenta uma taxa de cruzamentos de zero mais baixa ainda do que o sinal de voz. Entretanto, tal característica não pode ser usada da mesma maneira como foi usada para rejeitar o ruído de línha, pois a diferença na taxa de cruzamentos de zero não é muito acentuada quando se compara o ruído ambiental com as vogais e semi-vogais. Contudo, como estes fonemas apresentam uma amplitude elevada, eles são detectados eficientemente por um detector de pico, mesmo com um limiar de comparação relativamente alto. Através de testes subjetivos, constatou-se que um detector de pico comum, com um limiar de comparação de 20 níveis de quantização, é bastante sensível a estes fonemas, mesmo que a potência média do sinal de voz seja tão baixa quanto -44 dBm0, enquanto que é poucas vezes ativado por um ruído ambiental monótono de potência média inferior a -44 dBm0. Por outro lado, para que este mesmo detector de pico se torne sensível aos fonemas fricativos surdos, quando a potência média do sinal de voz é da ordem de -40 dBm0, o limiar de comparação deve ser reduzido para cerca de 4 níveis de quantização, pois os fonemas fricativos surdos apresentam amplitudes bem menores que os outros fonemas.

Diante destes resultados, e levando-se em conta que a discriminação entre o ruído ambiental e os fonemas fricativos sur

dos pode ser facilmente efetuada devido à grande diferença na taxa de cruzamentos de zero entre os mesmos, o aumento de imunidade ao ruído ambiental é feito através de um circuito de proteção que contém um limiar de comparação variável entre dois níveis prefixados. Em outras palavras, caso seja detectado que o sinal de entrada é um fonema fricativo, o limiar de comparação deve assumir um valor mínimo, $V_m = 4$ níveis de quantização. Em caso contrário, deve assumir um valor máximo, $V_M = 20$ níveis de quantização.

A estratégia para rejeição de ruído ambiental está esquematizado na Fig. III.11. Segundo esta estratégia, um determi

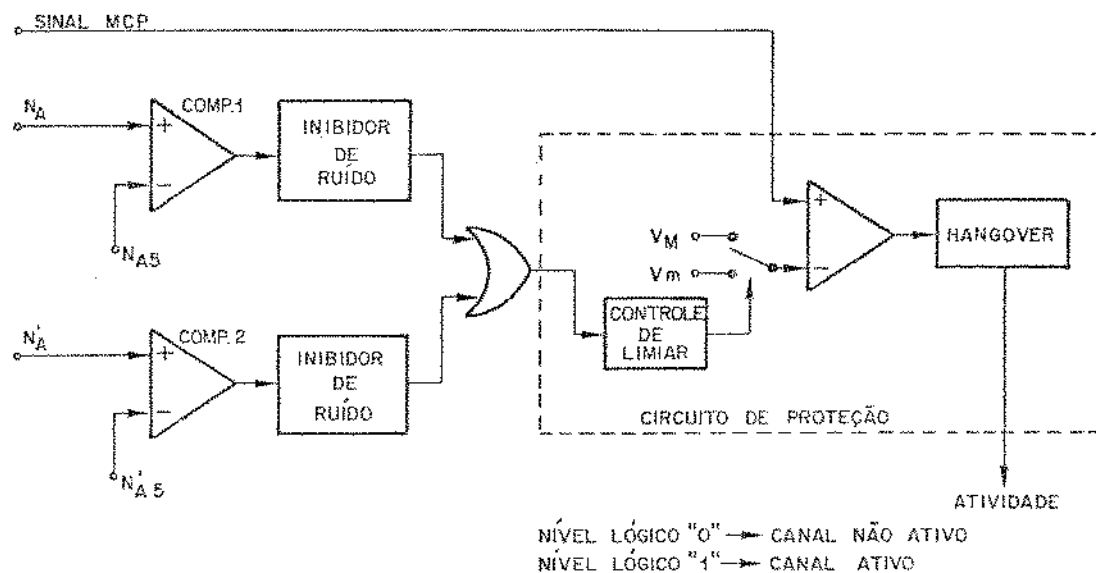


Fig. III.11 - Estratégia para rejeição de ruído ambiental

nado sinal, é considerado fonema fricativo somente se satisfizer uma das seguintes desigualdades abaixo durante um intervalo de tempo superior ao tempo de inibição de ruído:

$$N_A \leq N_{A5}$$

$$N'_A \leq N'_{A5}$$

Uma vez satisfeita esta condição, a saída do circuito de controle de limiar faz com que o limiar de comparação assumo o valor mínimo V_m , e permaneça neste valor durante um determinado intervalo de

tempo denominado tempo de controle de limiar. Adicionalmente, toda vez que a amostra do sinal de entrada for maior do que o limiar de comparação V_m ou V_M , é acionado um circuito de "hangover". Através de vários testes subjetivos foram encontrados os seguintes valores ótimos:

$$N_{A5} = 16 \text{ amostras}$$

$$N'_{A5} = 21 \text{ amostras}$$

$$\text{Tempo de inibição de ruído} = 9 \text{ ms}$$

$$\text{Tempo de controle de limiar} \approx 80 \text{ a } 96 \text{ ms}$$

$$\text{Tempo de "hangover"} \approx 128 \text{ a } 256 \text{ ms}$$

Portanto, as faixas de detecção de sons fricativos do primeiro e segundo analisador de cruzamentos são, respectivamente:

$$f \geq 500 \text{ Hz}$$

e $f' \geq 571 \text{ Hz}$

Na Fig. III.12 está ilustrado o diagrama de blocos completo do detector de voz Nº 2. O mesmo consiste de dois analisadores de cruzamentos de zero, cada um com dois inibidores de ruído (um para inibir o ruído de linha e o outro para inibir o ruído ambiental), de um detector de pico, de um circuito de proteção com limiar de comparação variável, e de um detector de sinalização MFC (comp. 7).

Observe-se que um determinado canal será considerado ativo somente se tanto a saída do circuito de "hangover" 1, como a saída do circuito de "hangover" 2 estiverem indicando, simultaneamente, que o canal está ativo (nível lógico "1").

Devido ao detector de pico com um limiar de comparação L_{CS} de 34 níveis de quantização, se a amplitude do sinal de entrada ultrapassar este limiar mais do que $p_1 = 3$ vezes entre dois cruzamentos de zero consecutivos, o canal será considerado ativo, independentemente dos valores de N_A e N'_A .

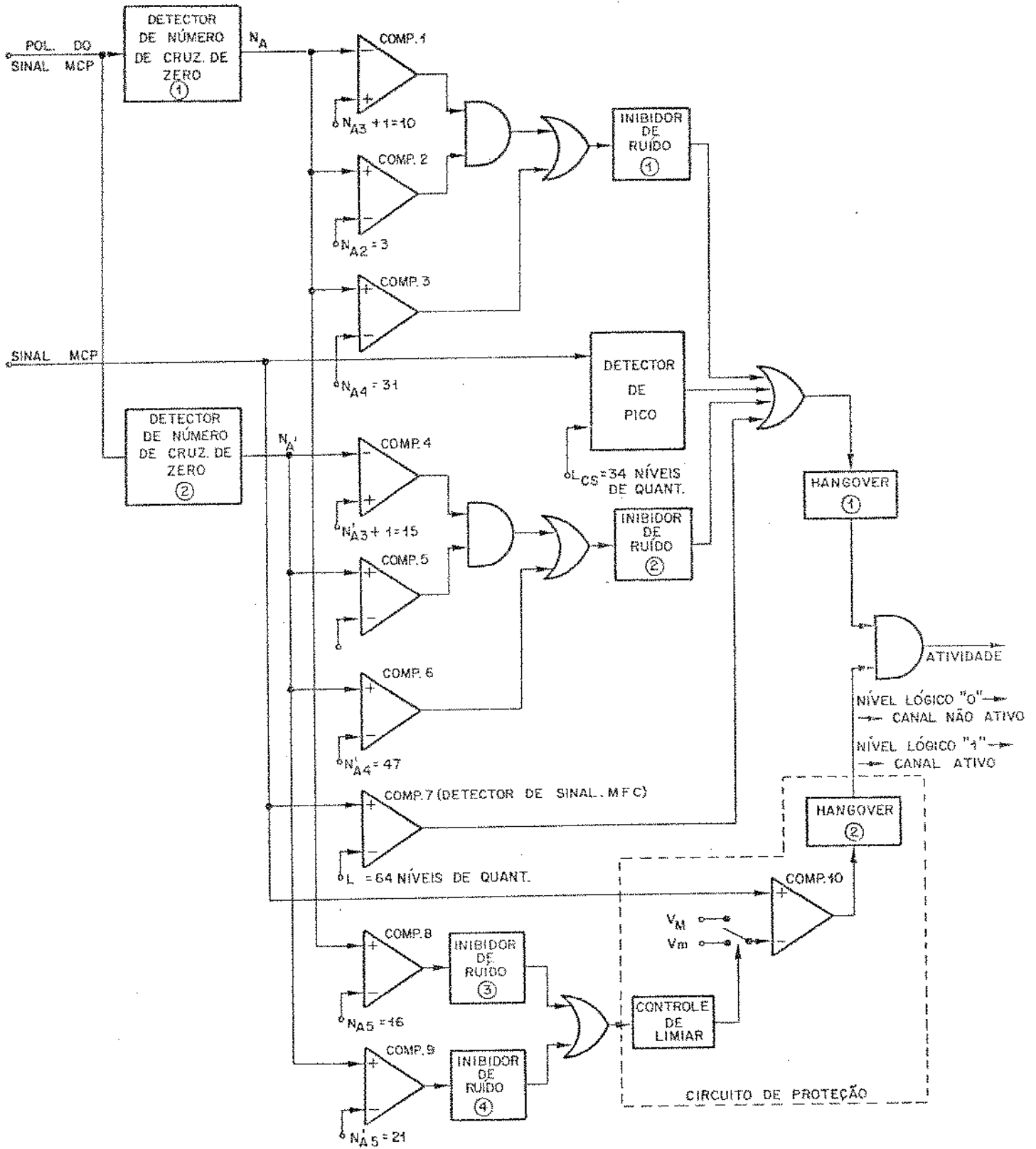


Fig. III.12 - Detector de voz Nº 2

O detector de sinalização MFC consiste de um simples comparador de nível com um limiar de 64 níveis de quantização, sendo pois ativado por qualquer sinal senoidal de potência média superior a -20 dBm0.

III.7 - DETECTOR DE OFF-SET

Um fato importante é que o sinal MCP não deve conter tensão de off-set, pois a mesma pode causar erros na determinação do número de cruzamentos de zero e também no valor da amplitude do sinal de voz. Portanto, o sinal MCP antes de ser recebido pelo detector de voz, deve passar por um detector de tensão de off-set, o qual determina o valor da tensão de off-set e o retira do sinal MCP, corrigindo a sua amplitude e a sua polaridade. Para isto, o detector de off-set utilizado pelo Sistema MCP-60 examina, durante um intervalo de tempo de 512 ms (4096 amostras) o número de amostras negativas N_{An} , e o número de amostras positivas, N_{Ap} . Caso o Sistema MCP-30 tributário não apresente tensão de off-set, como o sinal de voz e o ruído não possuem nível DC, os valores de N_{Ap} e N_{An} devem ser aproximadamente iguais. No caso em que $N_{Ap} > N_{An} \pm 32$, o detector de off-set inicia a correção do off-set subtraindo das amostras MCP 1 (um) nível de quantização no primeiro ciclo de 512 ms, 2 níveis no segundo ciclo, e assim sucessivamente, até que se atinja a condição $N_{Ap} = N_{An} \pm 32$, quando então a correção entra em regime estacionário. No caso em que $N_{Ap} < N_{An} \pm 32$, é efetuado o procedimento inverso (adição de níveis de quantização).

Observe-se que, como a tensão de off-set é compensada de 1 nível de quantização a cada 512 ms, se o seu valor for muito elevado, pode ser gasto um tempo relativamente longo até o detector determinar o valor correto da tensão de off-set e entrar em regime estacionário. Por exemplo, se a tensão de off-set for de 10 níveis de quantização, pode ser gasto um tempo de transição de 5 s. Entretanto, como a tensão de off-set para um determinado canal é praticamente constante, este tempo de transição não é crítico, uma vez que só ocorrerá no início de operação do Sistema MCP-60.

CAPÍTULO IV

IMPLEMENTAÇÃO

IV.1 - INTRODUÇÃO

Neste capítulo é feita uma descrição do detector de voz Nº 2 a nível de diagrama de circuitos.

Para facilitar o entendimento do diagrama de circuitos do detector de voz que deve atuar em compartilhamento de tempo para os 60 canais de voz do Sistema MCP-60, é feita, primeiramente, uma descrição do mesmo funcionando somente para um canal.

No Apêndice B encontram-se os diagramas detalhados de circuitos de cada um dos blocos que constituem o detector de voz.

IV.2 - DIAGRAMA DE CIRCUITOS DO DETECTOR DE VOZ PARA UM CANAL

O diagrama de circuitos do detector de voz para um canal, ilustrado na Fig. IV.1, é constituído basicamente de:

- detector de pico
- dois analisadores de cruzamentos de zero
- quatro inibidores de ruído
- dois circuitos de "hangover"
- circuito de proteção com controle de limiar

DETETOR DE PICO:

O detector de pico consta de um comparador de 8 bits e de um contador de 4 bits. O limiar de comparação L_{CS} do comparador está fixo em um valor igual ao valor de pico de uma senóide com -32 dBm0 de potência média. Somente é permitido ao contador contar quando a entrada "ENABLE" estiver no nível lógico "1", o que ocorre quando a amplitude da amostra do sinal MCP for maior do que o limiar de comparação L_{CS} . Como a frequência de relógio 2 (CLOCK 2) é igual a 8 KHz (vide Fig. IV.2), e as amostras do sinal MCP chegam à entrada do comparador com esta frequência e, além disso, o contador é zerado pelo sinal CLEAR 1 a cada ocorrência de cruzamento de zero, o contador conta exatamente o número de amostras que ultrapassam o limiar de comparação L_{CS} entre dois cruza

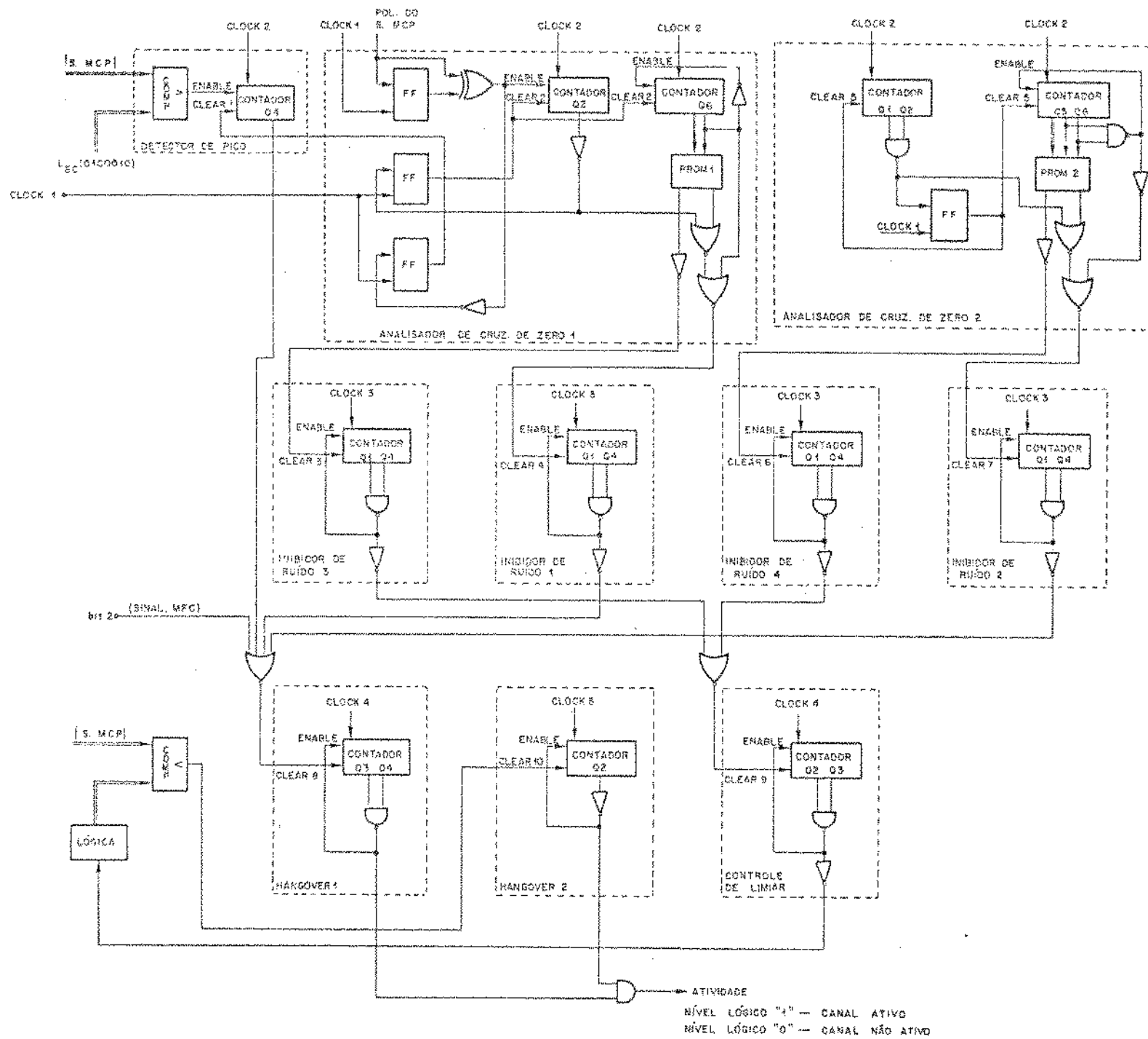


Fig. IV.1 - Diagrama de circuitos do detector de voz para um canal

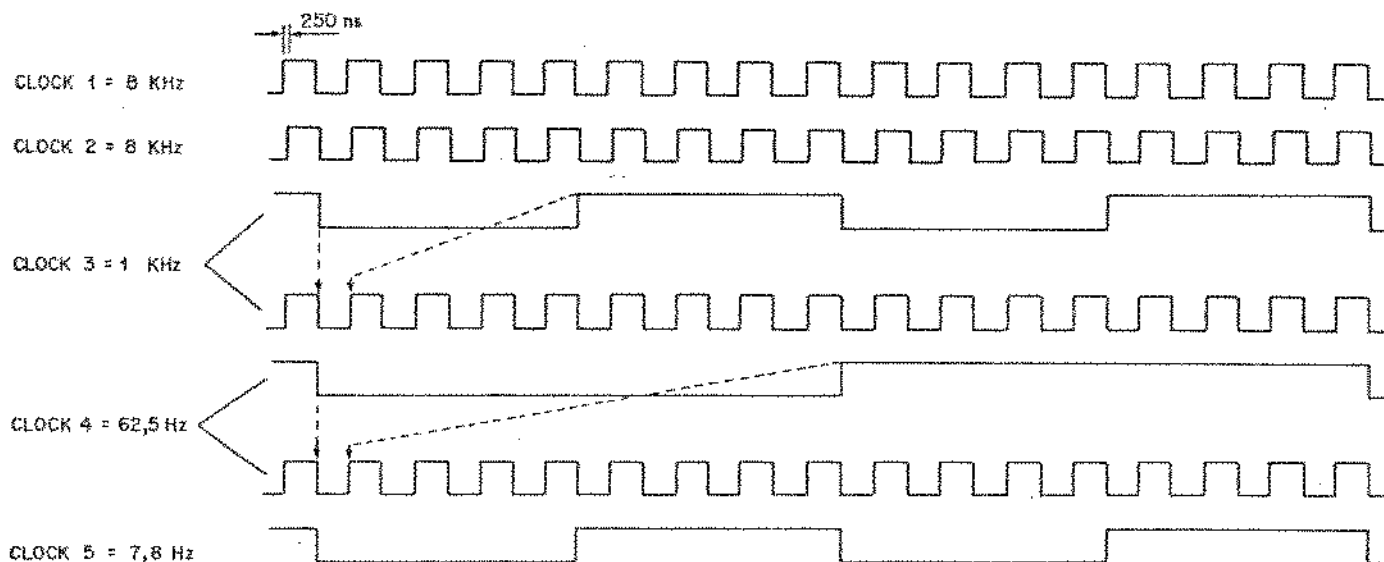


Fig. IV.2 - Ondas de controle do detector de voz para um canal

mentos de zero consecutivos. Se dentro deste intervalo de tempo ocorrerem 4 ou mais amostras acima do limiar de comparação L_{CS} , o quarto bit (Q4) do contador assume o nível lógico "1", indicando que o sinal presente é voz, e o circuito de "hangover" 1 é acionado. Em caso contrário, a saída Q4 do contador permanece no nível lógico "0".

ANALISADOR DE CRUZAMENTOS DE ZERO 1

O analisador de cruzamentos de zero 1 consta de três flip-flops tipo D, dois contadores, sendo um de 4 bits e outro de 8 bits, uma memória programável de leitura (PROM), e algumas portas (1 Exclusivo-OR, 1 Inversor e 2 NOR). O contador de 4 bits conta o número de cruzamentos de zero, e o contador de 8 bits conta o número de amostras N_A que ocorrem a cada $K=3$ cruzamentos de zero.

A detecção de um cruzamento de zero é feita comparando-se a polaridade da amostra anterior com a da amostra presente, o que é realizado usando-se um flip-flop tipo D e uma porta Exclusivo-OR. A saída desta porta é conectada à entrada ENABLE do contador de 4 bits, de modo que o mesmo somente poderá contar quando a

polaridade da amostra anterior for diferente da amostra atual, isto é, quando ocorrer um cruzamento de zero. O contador de 4 bits é zerado pelo sinal CLEAR 2 a cada dois cruzamentos de zero. Isto é, a partir do instante da ocorrência de um determinado cruzamento de zero, quando é zerado pelo sinal CLEAR 2, o contador de 4 bits conta até 2 cruzamentos de zero, sendo, então, novamente zerado pelo mesmo sinal para iniciar uma nova contagem e, assim sucessivamente, de modo que o tempo de observação da ocorrência de N_A amostras é definido para $K=3$ cruzamentos de zero consecutivos.

Como o contador de 8 bits possui o mesmo relógio (CLOCK 2) e o mesmo sinal de zeramento (CLEAR 2) que o contador de 4 bits, no mesmo instante em que este contador iniciar a contagem do número de cruzamentos de zero, o contador de 8 bits também inicia a contagem do número de amostras N_A , o qual é fornecido à memória PROM. A memória PROM é programada de modo que a sua saída para a lógica combinacional constituída de duas portas NOR, assumo o nível lógico "0" quando o número de amostras fornecido pelo contador for maior do que 3 e menor do que 10. Assim, a saída (CLEAR 4) da lógica combinacional assume o nível lógico "0" e aciona o inibidor de ruído 1, se $3 < N_A \leq 9$ ou $N_A > 31$. Em caso contrário, a saída da lógica assume o nível lógico "1" e não aciona o inibidor de ruído. Além disto, a memória PROM é programada de modo que a sua saída para o inibidor de ruído 3 (CLEAR 3) assumo o nível lógico "1" se o número N_A de amostras MCP fornecido pelo contador for maior do que 16, quando então é acionado o inibidor de ruído 3. Na Tabela IV.1 é apresentada a programação da PROM.

Observe-se que o instante de zeramento dos contadores é atrasado, através de um flip-flop tipo D, de quase um período de relógio em relação ao instante em que é verificada a ocorrência de cruzamentos de zero (vide ondas de controle da Fig. IV.2). Este atraso é necessário porque a memória PROM mais a lógica combinacional levam um certo tempo para determinar o estado de atividade do canal e ativar o circuito inibidor de ruído, caso se verifique que o canal está não-ativo.

ENTRADAS					PROM 1				PROM 2			
E	D	C	B	A	SAÍDA PARA A LÓGICA COMBINACIONAL	SAÍDA PARA O INIBIDOR DE RUÍDO 3	SAÍDA PARA A LÓGICA COMBINACIONAL	SAÍDA PARA O INIBIDOR DE RUÍDO	SAÍDA PARA A LÓGICA COMBINACIONAL	SAÍDA PARA O INIBIDOR DE RUÍDO		
0	0	0	0	0	1	0	1	0	1	0		
0	0	0	0	1	1	0	1	0	1	0		
0	0	0	1	0	1	0	1	0	1	0		
0	0	0	1	1	0	0	1	0	1	0		
0	0	1	0	0	0	0	1	0	1	0		
0	0	1	0	1	0	0	0	0	0	0		
0	0	1	1	0	0	0	0	0	0	0		
0	0	1	1	1	0	0	0	0	0	0		
0	1	0	0	0	1	0	1	0	1	0		
0	1	0	0	1	1	0	1	0	1	0		
0	1	0	1	0	1	1	1	0	1	0		
0	1	0	1	1	0	1	1	0	1	0		
1	0	0	0	0	1	1	1	0	1	0		
1	0	0	0	1	1	1	1	0	1	0		
1	0	0	1	0	1	1	1	0	1	0		
1	0	0	1	1	0	1	1	0	1	0		
1	0	1	0	0	1	1	1	0	1	0		
1	0	1	0	1	1	1	1	0	1	0		
1	0	1	1	0	1	1	1	0	1	0		
1	0	1	1	1	1	1	1	0	1	0		
:	:	:	:	:	:	:	:	:	:	:		
1	1	1	1	1	1	1	1	1	1	1		

Tab. IV.1 - Tabela de Programação da PROM

ANALISADOR DE CRUZAMENTOS DE ZERO 2

O analisador de cruzamentos de zero 2 é idêntico ao analisador de cruzamentos de zero 1, com a diferença de que os contadores são zerados a cada três cruzamentos de zero, isto é, o tempo de observação da ocorrência de N'_A amostras é definido para $K = 4$ cruzamentos de zero consecutivos. Além disto, agora a PROM é programada de modo que a sua saída para a lógica combinacional constituída de duas portas NOR, assuma o nível lógico "0" quando número de amostras fornecido pelo contador for maior do que 5 e menor do que 15. Adicionalmente, a saída para o inibidor de ruído 4 é programada de modo a assumir o nível lógico "1" quando $N'_A > 21$.

INIBIDOR DE RUÍDO 1

O inibidor de ruído 1 consiste de um contador de 4 bits e de uma lógica combinacional constituída por uma porta NAND e um inversor. O relógio (CLOCK 3) do contador é de 1 KHz, e é zerado pelo analisador de cruzamentos de zero 1 através do sinal CLEAR 4. Toda vez que o contador é zerado, a saída do inibidor de ruído assume o nível lógico "0" e permanece neste nível durante 9 ciclos de relógio, isto é, durante 9 ms. Assim, a saída do inibidor de ruído somente assumirá o nível lógico "1", ativando o circuito de "hangover", se o contador deixar de ser zerado pelo analisador de cruzamentos de zero 1 durante um intervalo de tempo maior do que 9 ms.

INIBIDORES DE RUÍDO 2, 3 E 4

Idêntico ao inibidor de ruído 1.

CIRCUITO DE PROTEÇÃO COM CONTROLE DE LIMIAR

O circuito de proteção com controle de limiar consiste de um comparador de 8 bits, de um contador de 4 bits e de duas lógicas combinacionais. O contador possui um sinal de relógio (CLOCK

4) de 62,5 Hz, e pode ser zerado ou pelo inibidor de ruído 3 ou pelo inibidor de ruído 4. Sempre que o contador é zerado, a saída do circuito de controle de limiar assume o nível lógico "0", acionando uma lógica combinacional que, então, abaixa o limiar de comparação do comparador de 20 para 4 níveis de quantização. A lógica combinacional à saída do contador constituída por uma porta NAND e um inversor, faz com que a saída do circuito de controle de limiar permaneça no nível lógico "0" durante 6 ciclos de relógio de 62,5 Hz. Desta maneira, o limiar de comparação do comparador permanece em 4 níveis de quantização durante um intervalo de tempo variável de 80 a 96 ms (tempo de controle de limiar).

CIRCUITO DE HANGOVER 1

O tempo de "hangover" 1 é conseguido através de um contador e de uma lógica combinacional realizada por uma porta NAND de duas entradas. O contador é zerado quando a saída de um dos circuitos abaixo assume o nível lógico "1":

- detector de pico
- inibidor de ruído 1
- inibidor de ruído 2,

ou quando o 2º bit mais significativo da amostra MCP assumir o nível lógico "1" (detecção de sinalização MFC). A saída da porta NAND assume o nível lógico "1" sempre que o contador é zerado, e permanece neste nível até que o contador atinja um valor pré-determinado de acordo com o tempo de "hangover" desejado. Neste caso, a saída da porta NAND permanece no nível lógico "1" durante a ocorrência de 12 ciclos do relógio (CLOCK 4) de 62,5 Hz, isto é, durante um intervalo de tempo variável entre 176 e 192 ms. Uma vez atingido este valor, a saída da porta NAND vai para o nível lógico "0" e desativa o contador através da entrada ENABLE.

CIRCUITO DE HANGOVER 2

O funcionamento deste circuito é idêntico ao do circuito de "hangover" 1, com a diferença de que, sempre que é ativado, a

sua saída permanece no nível lógico "1" durante a ocorrência de dois ciclos de relógio (CLOCK 5) de 7,8 Hz, isto é, durante um intervalo de tempo variável entre 128 a 256 ms.

IV.3 - DIAGRAMA DE CIRCUITOS DO DETECTOR DE VOZ PARA 60 CANAIS

A Fig. IV.3 mostra o mesmo diagrama de circuitos do detector de voz da Fig. IV.1, porém funcionando para todos os 60 canais do Sistema MCP-60 através de compartilhamento de tempo. Neste caso, os contadores da Fig. IV.1 foram substituídos por contadores endereçáveis, os quais foram implementados usando-se uma memória de leitura e escrita aleatória (RAM), um somador e um flip-flop tipo D. Estes contadores, além das entradas de relógio (CLOCK) clear e enable, possuem ainda as entradas de controle de escrita, (\overline{WE}) e leitura (\overline{ME}), e entradas de endereço (vide as ondas de controle da Fig. IV.4). Para cada canal o contador endereçável conta o número de cruzamentos de zero ou o número de amostras do sinal MCP, e armazena o resultado obtido no endereço correspondente a este canal (operação de escrita). Ele possibilita, ainda, zerar o valor armazenado no endereço correspondente a um determinado canal para iniciar uma nova contagem (CLEAR), inibir a contagem (ENABLE), ou ler o valor armazenado num determinado endereço. Esta leitura ocorre quando a onda de controle \overline{ME} está no nível lógico "0", e \overline{WE} está no nível lógico alto. Se, por outro lado, tanto \overline{ME} como \overline{WE} estiverem no nível lógico "0", a operação é de escrita (vide tabela IV.2).

\overline{ME}	\overline{WE}	OPERAÇÃO	SAÍDA
L	L	ESCRITA	TRI-STATE
L	H	LEITURA	DADOS

Tab. IV.2 - Tabela verdade da memória RAM (74C910)

Desta maneira, é possível implementar um detector de voz idêntico ao da Fig. IV.1, mas que funcione para todos os 60 canais em compartilhamento de tempo.

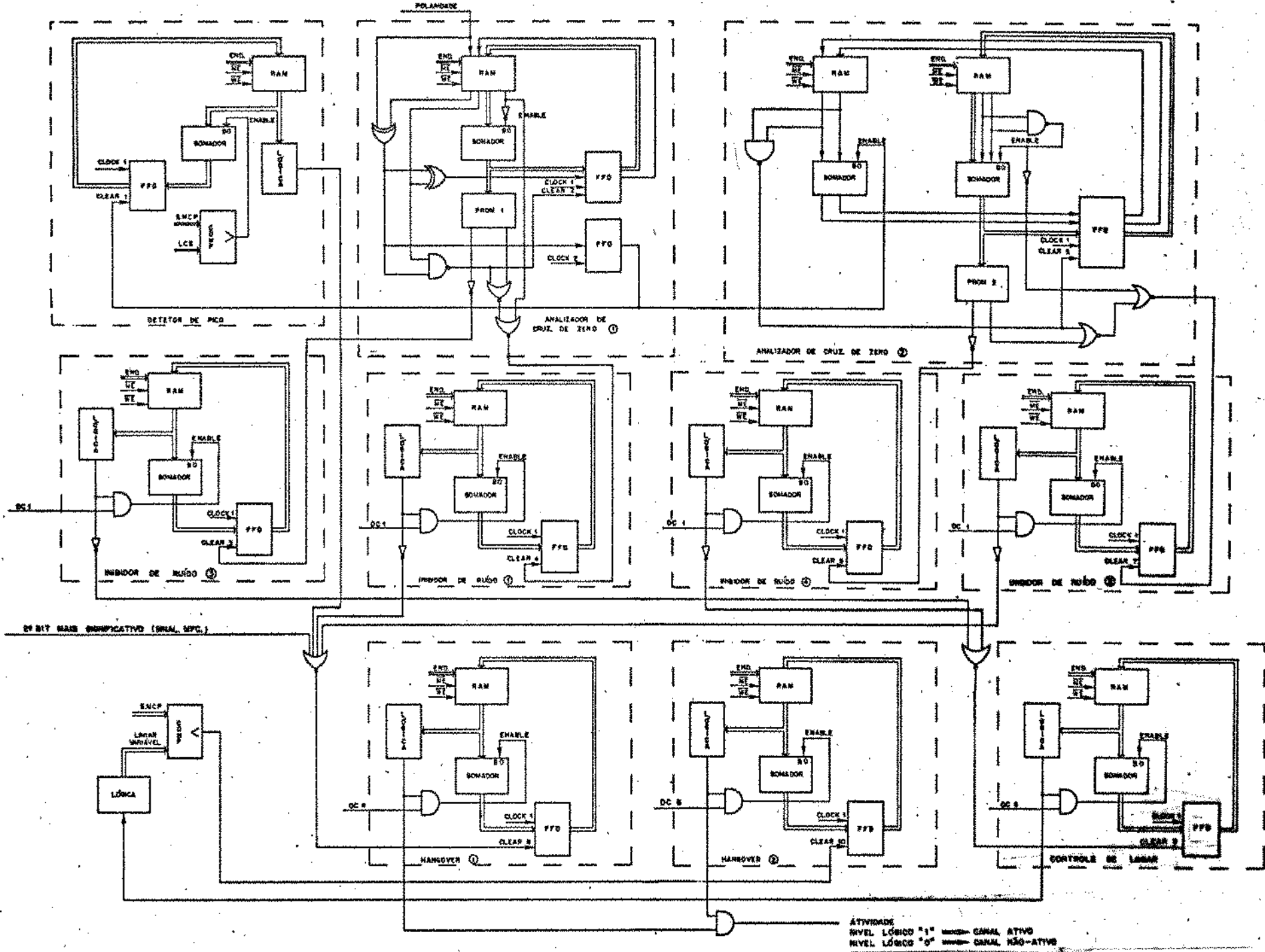


Fig. IV.3 - Diagrama de circuitos do detector de voz para os 60 canais do Sistema MCP-60

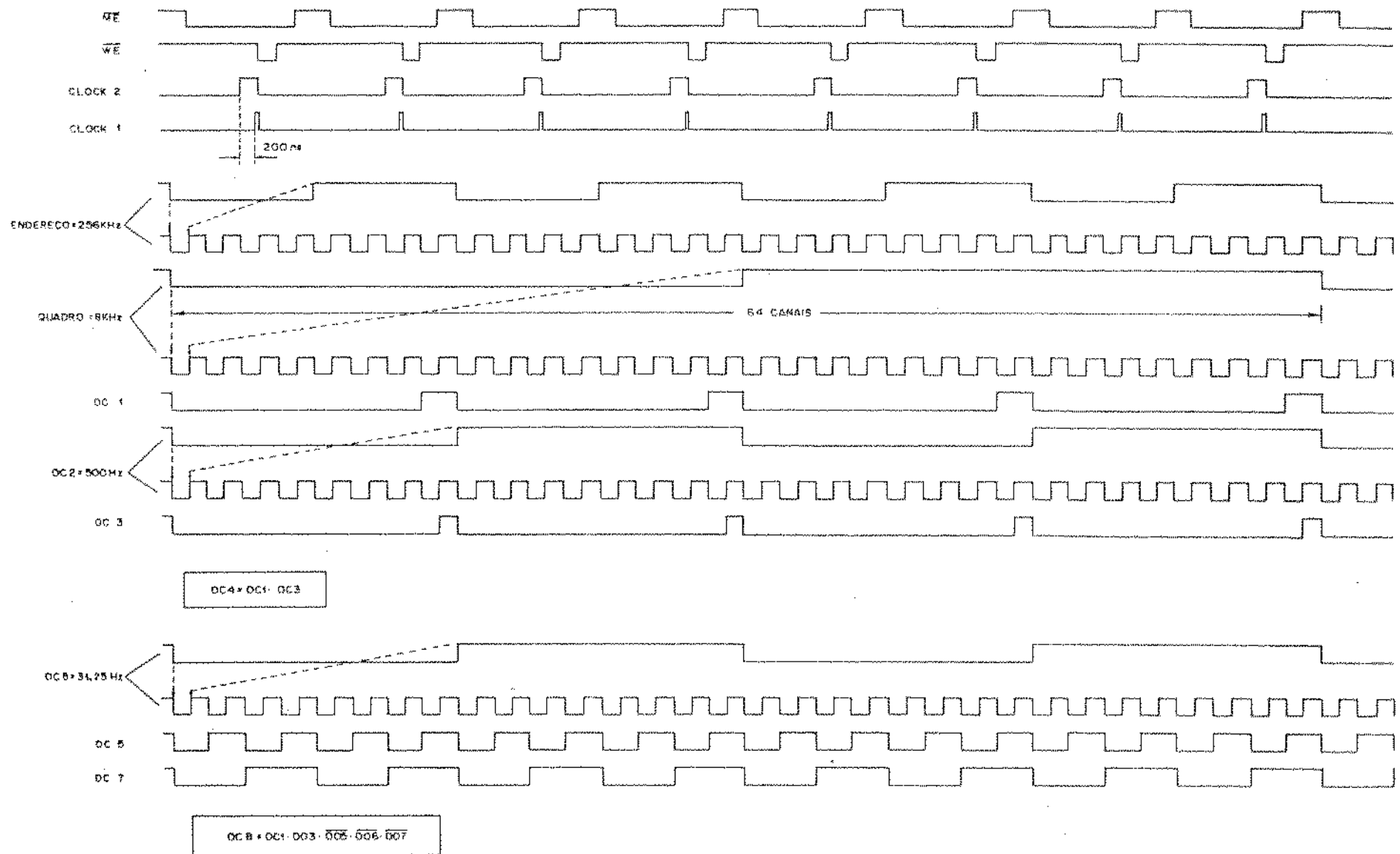


Fig. IV.4 - Ondas de controle do detector de voz funcionando em compartilhamento de tempo

Como as amostras dos 60 canais chegam em série, multiplexadas no tempo sequencialmente, o detector de voz é também compartilhado sequencialmente no tempo pelos 60 canais do Sistema MCP-60. Além disto, tendo em vista que a duração de um quadro é de 125 μ s, o procedimento de verificação de atividade de cada canal é feito dentro de um intervalo de tempo de aproximadamente 2 μ s.

Para a implementação do detector de voz mostrado na Fig. IV.3, foram empregados 75 chips (incluindo também a geração das ondas de controle mostradas na Fig. IV.4). Destes 75 chips, 23 são de integração em pequena escala (SSI), e o restante é de integração em média escala (MSI). Dos chips SSI, 3 são de tecnologia CMOS, e o restante TTL do tipo LS (Low Schottky); dos chips MSI, 12 são de tecnologia CMOS, e o restante TTL do tipo LS, com exceção das memórias PROM's, que são do tipo normal.

CAPÍTULO V

TESTES DE DESEMPENHO E
CONCLUSÕES

V.1 - TESTES DE DESEMPENHO

Para verificar o desempenho do detector de voz implementado, foram realizados os seguintes testes:

a) Detecção de voz: para efetuar este teste foram feitos três tipos de gravação em fitas cassete. O primeiro tipo de gravação consistiu de diversas palavras divididas em vários grupos, de acordo com o fonema inicial. Para esta gravação foi usado um microfone de resposta plana, e as palavras foram pronunciadas de modo a se manter uma pausa entre as mesmas maior do que o tempo de "hangover", a fim de se poder observar se o efeito dos cortes parciais causado pelo detector de voz no início destas palavras é perceptível. O segundo tipo de gravação consistiu das mesmas palavras da primeira gravação, somente que agora estas palavras passaram por uma ligação telefônica real. Finalmente, o terceiro tipo de gravação consistiu de diversas conversações telefônicas reais.

Os testes subjetivos com palavras do primeiro tipo de gravação indicaram que o detector de voz é extremamente sensível às palavras de todos os grupos, mesmo quando o teste foi efetuado com uma potência média tão baixa quanto -44 dBm₀. Entretanto, esta sensibilidade mostrou-se um pouco variável de acordo com os diversos fonemas que compõem a palavra. Verificou-se que o caso mais crítico ocorre quando as palavras se iniciam com os sons fricativos "s" e "f" seguido da vogal "a", com em "sabão", "faculdade", etc. Nestes casos, quando a potência média da palavra cai abaixo de aproximadamente -40 dBm₀, o efeito dos cortes parciais nestes fonemas começa a se tornar perceptível, embora seja mantida a inteligibilidade dos mesmos. Isto ocorre porque estes fonemas apresentam uma amplitude muito reduzida comparativamente aos outros fonemas, como, por exemplo, as vogais. Em outros termos, embora a potência média das palavras que o contêm seja alta, a potência média correspondente ao "s" ou "f" é muito baixa, o que faz com que o circuito de proteção iniba a saída do circuito de "hangover" 1 (vide Fig. III.12). Além disto, o fato destes fonemas apresentarem uma amplitude muito reduzida, permite que qualquer ruído (por exemplo, ruído de fita, ruído do gravador) mascare a sua

característica espectral, fazendo com que os analisadores de cruzamentos de zero os interpretem como ruídos.

Quando o teste é efetuado mantendo-se a potência média das palavras acima de -40 dBm0, o efeito destes cortes parciais torna-se imperceptível. Isto deve-se ao fato de que, neste caso, a amplitude dos sons correspondentes a estes fonemas ultrapassa com maior probabilidade o limiar de comparação do circuito de proteção e, além disto, a característica espectral destes sons torna-se menos mascarável pelo ruído.

Por outro lado, para palavras iniciadas com diversos outros fonemas como "x", "j", "z", etc, o detector de voz mostrou-se extremamente sensível mesmo para potência média bem inferior a -44 dBm0, devido ao espectro de potência destes fonemas ser mais favorável para os analisadores de cruzamentos de zero os considerarem como voz.

Tendo em vista que num caso real o sinal de voz sofre em frequências acima de 3,4 KHz uma filtragem adicional devido ao microfone telefônico, aos cabos e outros equipamentos multiplex, foi realizado um teste colocando-se em série com o filtro passa-baixas do Sistema MCP-30, um outro filtro passa-baixas em 3,4 KHz, de modo a aumentar a atenuação em altas frequências. Obteve-se então uma melhoria sensível na detecção de sons fricativos surdos "s" e "f". Isto é explicado pelo fato de uma filtragem adicional em frequências elevadas proporcionar uma característica espectral mais favorável de ser detectada como sendo de voz pelo analisador de cruzamentos de zero.

Um teste mais real foi realizado usando-se o segundo tipo de gravação, pois neste caso foram levados em conta adicionalmente as filtrações do microfone telefônico, dos cabos e outros equipamentos multiplex, e além disto, o ruído de linha. Verificou-se então, como era esperado, uma melhoria na detecção de palavras iniciadas com os fonemas fricativos surdos "s" e "f", de modo que em nenhum caso foi possível notar-se alguma diferença entre estas palavras antes e depois de sua passagem pelo detector de voz.

O teste com conversações telefônicas reais foi efetuado mantendo-se a potência média do sinal de voz abaixo de -40 dBm0,

de modo que o eco fosse cortado pelo circuito de proteção, uma vez que o Sistema MCP-60 ainda não teve incorporado um circuito supressor de eco. Como resultado, verificou-se também que os cortes parciais provocados pelo detector de voz são imperceptíveis.

A fim de verificar a variação do desempenho do detector de voz com a relação S/R, foram realizados testes adicionando-se ao sinal de voz um ruído gaussiano branco de faixa limitado em 3,4 KHz. Verificou-se que a sensibilidade do detector de voz não é afetada de modo perceptível para uma relação S/R de até 35 dB. Para uma relação S/R abaixo deste valor, o efeito dos cortes parciais sofrido pelos fonemas fricativos surdos "s" e "f", começa a se tornar perceptível. Entretanto, para os outros fonemas, excluindo os fricativos acima mencionados, foi verificada uma boa sensibilidade para uma relação S/R de até 15 dB.

Objetivando-se verificar a influência da variação da frequência de corte inferior do canal telefônico no desempenho do detector de voz, passou-se o sinal de voz por um filtro passa-altas antes de sua aplicação ao detector de voz. Verificou-se que, com exceção das palavras que se iniciam com os sons fricativos surdos "s" e "f", a sensibilidade do detector de voz não é alterada para uma frequência de corte de até 500 Hz. Para os "s" e "f", entretanto, esta frequência de corte deve se manter no máximo em torno de 300 Hz para se evitar cortes perceptíveis na sua detecção.

b) Imunidade ao ruído: os testes efetuados mostraram que o detector de voz apresenta uma imunidade quase total para o ruído gaussiano branco de faixa limitada em 3,4 KHz, e de potência média até -34 dBm0 (foi observado, em média, uma detecção falsa de duração inferior a 190 ms a cada 40 s). Esta imunidade não foi muito afetada mesmo quando se alterou a faixa do ruído para aproximadamente 2,5 KHz.

Para verificar a imunidade ao ruído ambiental, foram feitas gravações de dois tipos de ruído ambiental passando através de uma linha telefônica. A primeira consistiu de ruído de rua (o relhão do centro da cidade de Campinas), e a segunda de ruído ambiental em um estabelecimento comercial. O detector de voz apresentou uma alta imunidade (em torno de 95% do tempo) a estes ruí

dos de potência média até aproximadamente -44 dBm0.

c) medida de atividade média: na Fig. V.1 é mostrada uma uma função de distribuição acumulada de canais ativos. Esta função foi obtida a partir dos resultados de uma medida de atividade média efetuada mantendo-se, através de toca-fitas, todos os 60 canais ocupados, durante 8 minutos, com conversações telefônicas reais. Desta função tem-se que $P(N_A \leq 22) \approx 0,5$. Portanto, a atividade média α , nestes 8 minutos de testes, foi de aproximadamente 37%.

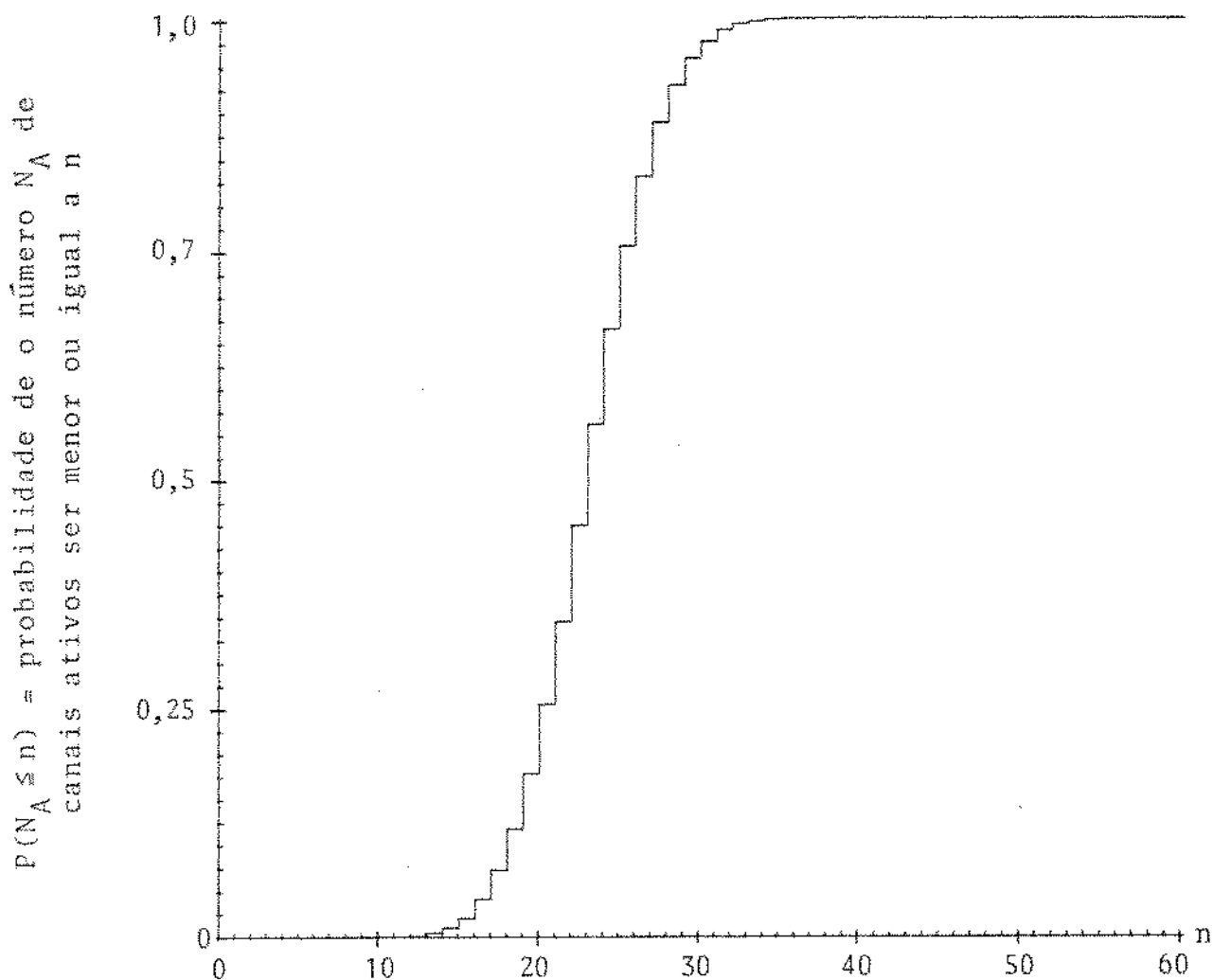


Fig. V.1 - Função de distribuição acumulada de canais ativos

V.2 - CONCLUSÕES

Segundo os testes de desempenho realizados, pode-se dizer que o detector de voz discrimina eficientemente o sinal de voz do ruído. De fato, a potência média de sinal de voz necessária para engatilhar o detector de voz está em torno de 10 a 15 dB abaixo da potência média de ruído gaussiano necessária para engatilhar o mesmo. Embora esta mesma diferença não seja observada quando se trata de ruído ambiental, a rejeição do detector de voz a este tipo de ruído ainda está dentro dos objetivos de projeto estabelecidos no Capítulo III.

Com relação ao ruído impulsivo não foi realizado nenhum teste, pois as medidas de atividade média usando um detector de voz convencional com limiar de -40 dBm0, sem nenhuma estratégia para rejeição de ruído [3], mostram que o aumento na atividade média devido ao ruído impulsivo não é crítico. Além disto, devido a grande variação nas características do ruído impulsivo, tanto em amplitude como em duração, torna-se muito difícil realizar testes de laboratório para avaliar sua influência no desempenho do detector de voz.

Um fator importante a ser lembrado é que a sensibilidade do detector de voz é praticamente independente da potência do sinal de voz, desde que este esteja acima do limiar mínimo do circuito de proteção. Como o eco nada mais é que um sinal de voz de baixa potência, ele pode causar um aumento apreciável na atividade média do canal. Portanto, é imprescindível que o Sistema MCP-60, assim como qualquer outro sistema de interpolação de voz, seja dotado de um supressor ou cancelador de eco.

Devido a inexistência de dados sobre a potência média de sinal de voz em circuitos telefônicos no Brasil, o desempenho do presente detector de voz foi otimizado levando-se em conta os dados de pior caso de circuitos telefônicos de outros países. Assim, caso seja verificado através de testes de campo um valor mais elevado para a potência média de pior caso do sinal de voz, os valores mínimo (V_m) e máximo (V_M) do limiar de comparação do circuito de proteção poderão ser aumentados a fim de se melhorar a imunidade ao ruído ambiental.

APÉNDICES

APÊNDICE A

DESVIO PADRÃO DO NÚMERO DE CRUZAMENTOS DE ZERO,
POR INTERVALO DE TEMPO T,
DE UM RUÍDO GAUSSIANO BRANCO DE FAIXA LIMITADA EM f_0

A eq.(II.11), usada no Capítulo II para o cálculo do desvio padrão σ_T do número de cruzamentos de zero de um ruído gaussiano branco de faixa limitada em f_0 , sobre um intervalo de tempo T, é um caso particular de uma equação mais geral dado por [16]:

$$D_p = 2T\lambda_1(x_0) - 2T^2\lambda_1^2(x_0) + 4T \int_0^T \left(1 - \frac{\tau}{T}\right) v_2(x_0, x_0, \tau) d\tau \quad (A.1)$$

onde:

$\lambda_1(x_0)$ é o número de vezes que o ruído cruza, com derivada positiva, um determinado nível x_0 durante um intervalo de tempo T

$$v_2(x_0, x_0, \tau) = \int_0^\infty \int_0^\infty y_1 y_2 \omega_4(x_0, y_1, x_0, y_2, \tau) dy_1 dy_2 \quad (A.2)$$

$\omega_4(x_0, y_1, x_0, y_2, \tau)$ é a densidade de probabilidade conjunta de um processo aleatório qualquer e de sua derivada em dois instantes diferentes t e t+ τ .

A eq.(A.1) fornece o desvio padrão D_p do número de vezes que um processo aleatório qualquer cruza um determinado nível x_0 , durante um intervalo de tempo T.

No caso particular de um ruído gaussiano, tem-se [16]:

$$\lambda_1(x_0) = \frac{1}{2\pi} \exp(-x_0^2/2\sigma^2) \cdot \omega_1 \quad (A.3)$$

$$\omega_4(x_0, y_1, x_0, y_2, \tau) = \frac{1}{(2\pi)^2 \sigma^4 \sqrt{D}} \exp \left\{ - \frac{1}{2\sigma^2 D} [2(D_{11} + D_{13})x_0^2 + D_{22}(y_1^2 + y_2^2) + 2D_{24}y_1y_2 + 2(D_{12} + D_{14})(y - y_1)x_0] \right\} \quad (A.4)$$

onde:

σ é o desvio padrão da distribuição do ruído gaussiano.

$$D = (\omega_1^2 - R''^2)(1 - R^2) + R'^2(R'^2 - 2RR'' - 2\omega_1^2) = \frac{D_{22}^2 - D_{24}^2}{1 - R^2} \quad (A.5)$$

$$D_{11} = \omega_1^4 - R''^2 - \omega_1^2 R'^2$$

$$D_{22} = \omega_1^2(1 - R^2) - R'^2$$

$$D_{12} = R'R'' + RR'\omega_1^2$$

$$D_{14} = R'^3 - R'\omega_1^2 - RR'R''$$

$$D_{13} = RR''^2 - \omega_1^4 R - R'^2 R''$$

$$D_{24} = R''(1 - R^2) + RR'^2$$

$$\omega_1 = \frac{\int_0^{\infty} \omega^2 F(\omega) d\omega}{\int_0^{\infty} F(\omega) d\omega} \quad (A.6)$$

$F(\omega)$ é a forma do espectro de potência do ruído

R é o coeficiente de correlação do ruído gaussiano

$$R' = \frac{dR}{d\tau}$$

$$R'' = \frac{d^2R}{d\tau^2}$$

Como no caso interessa saber o desvio padrão σ_T do número médio de cruzamentos de zero de um ruído gaussiano branco de faixa limitado em $\omega_0 = 2\pi f_0$, tem-se:

$$x_0 = 0$$

$$F(\omega) = |C(\omega)|^2,$$

onde:

$C(\omega)$ é um filtro passa-baixas ideal de frequência de corte f_0 .

$$R(\tau) = R = (\text{sen}\omega_0\tau)/(\omega_0\tau)$$

Portanto, as eqs.(A.3), (A.4) e (A.6) podem ser escritas, respectivamente, como:

$$\lambda_1(x_0) \Big|_{x_0=0} = \frac{\omega_1}{2\pi} \tag{A.7}$$

$$\omega_4(x_0, y_1, x_0, y_2, \tau) \Big|_{x_0=0} = \frac{1}{(2\pi)^2 \sigma^4 \sqrt{D}} \exp \left\{ - \frac{1}{2\sigma^2 D} (D_{22} y_1^2 + D_{22} y_2^2 + 2D_{24} y_1 y_2) \right\} \tag{A.8}$$

$$\omega_1 = \omega_0/\sqrt{3} = 2\pi f_0/\sqrt{3} \tag{A.9}$$

Substituindo-se a eq.(A.8) na eq.(A.2), tem-se:

$$v_2(x_0, x_0, \tau) \Big|_{x_0=0} = \int_0^\infty \int_0^\infty y_1 y_2 \frac{1}{(2\pi)^2 \sigma^4 \sqrt{D}} \exp \left\{ - \frac{1}{2\sigma^2 D} (D_{22} y_1^2 + D_{22} y_2^2 + 2D_{24} y_1 y_2) \right\} dy_1 dy_2 \tag{A.10}$$

Colocando-se esta equação num sistema de coordenadas cilíndricas circulares, tem-se:

$$v_2(x_0, x_0, \tau) \Big|_{x_0=0} = \frac{1}{(2\pi)^2 \sigma^4 \sqrt{D}} \int_0^{\pi/2} \int_0^{\infty} \frac{\text{sen}2\theta}{2} \exp \left\{ -\frac{1}{2\sigma^2 D} (D_{22} + D_{24} \text{sen}2\theta) r^2 \right\} r^3 dr d\theta$$

Para resolver a integral interna na expressão acima, considere-se:

$$K = \frac{1}{2\sigma^2 D} (D_{22} + D_{24} \text{sen}2\theta) \quad , \text{ de modo que:}$$

$$\begin{aligned} & \int_0^{\infty} \exp \left\{ \frac{-1}{2\sigma^2 D} (D_{22} + D_{24} \text{sen}2\theta) r^2 \right\} r^3 dr = \\ & = \int_0^{\infty} \exp(-Kr^2) r^3 dr = \frac{-1}{2K} \int_0^{\infty} r^2 \exp(-Kr^2) \cdot (-2Kr) dr = \\ & = \frac{-1}{2K} (r^2 \exp(-Kr^2)) \Big|_0^{\infty} - \int_0^{\infty} \exp(-Kr^2) \cdot 2r dr = \frac{1}{2K^2} = \\ & = \frac{2\sigma^4 D^2}{(D_{22} + D_{24} \text{sen}2\theta)^2} \quad , \quad \text{para } K > 0 \end{aligned}$$

Logo,

$$v_2(x_0, x_0, \tau) \Big|_{x_0=0} = \frac{D^{3/2}}{(2\pi)^2} \int_0^{\pi/2} \frac{\text{sen}2\theta}{(D_{22} + D_{24} \text{sen}2\theta)^2} d\theta$$

Fazendo $\alpha = 2\theta$ e $K' = \frac{D_{24}}{D_{22}}$, obtêm-se:

$$\begin{aligned}
 v_2(x_0, x_0, \tau) \Big|_{x_0=0} &= \frac{D^{3/2}}{(2\pi)^2 D_{22}^2} \cdot \frac{1}{2} \cdot \int_0^\pi \frac{\text{sen}\alpha}{(1+K' \text{sen}\alpha)^2} d\alpha = \\
 &= \frac{D^{3/2}}{(2\pi)^2 D_{22}^2} \cdot \frac{1}{2} \left[\frac{\cos\alpha}{(K'^2-1)(1+K' \text{sen}\alpha)} \Big|_0^\pi + \frac{K'}{K'^2-1} \int_0^\pi \frac{d\alpha}{1+K' \text{sen}\alpha} \right] \\
 &= \frac{D^{3/2}}{(2\pi)^2 D_{22}^2} \cdot \left[\frac{1}{2} \frac{\cos\alpha}{(K'^2-1)(1+K' \text{sen}\alpha)} + \frac{K'}{K'^2-1} \cdot \frac{2}{\sqrt{1-K'^2}} \cdot \right. \\
 &\quad \left. \cdot \text{arc tg} \left(\frac{\text{tg } \alpha/2 + K'}{\sqrt{1-K'^2}} \right) \right] \Big|_0^\pi = \\
 &= \frac{D^{3/2}}{(2\pi)^2 D_{22}^2} \cdot \left[\frac{1}{1-K'^2} - \frac{K'}{(1-K')^{3/2}} \cdot \text{arc cos} K' \right] \\
 &= \frac{D^{3/2}}{(2\pi)^2 D_{22}^2} \cdot \left[\frac{D_{22}^2}{D_{22}^2 - D_{24}^2} - \frac{D_{24} D_{22}^2}{(D_{22}^2 - D_{24}^2)^{3/2}} \text{arc cos} \frac{D_{24}}{D_{22}} \right],
 \end{aligned}$$

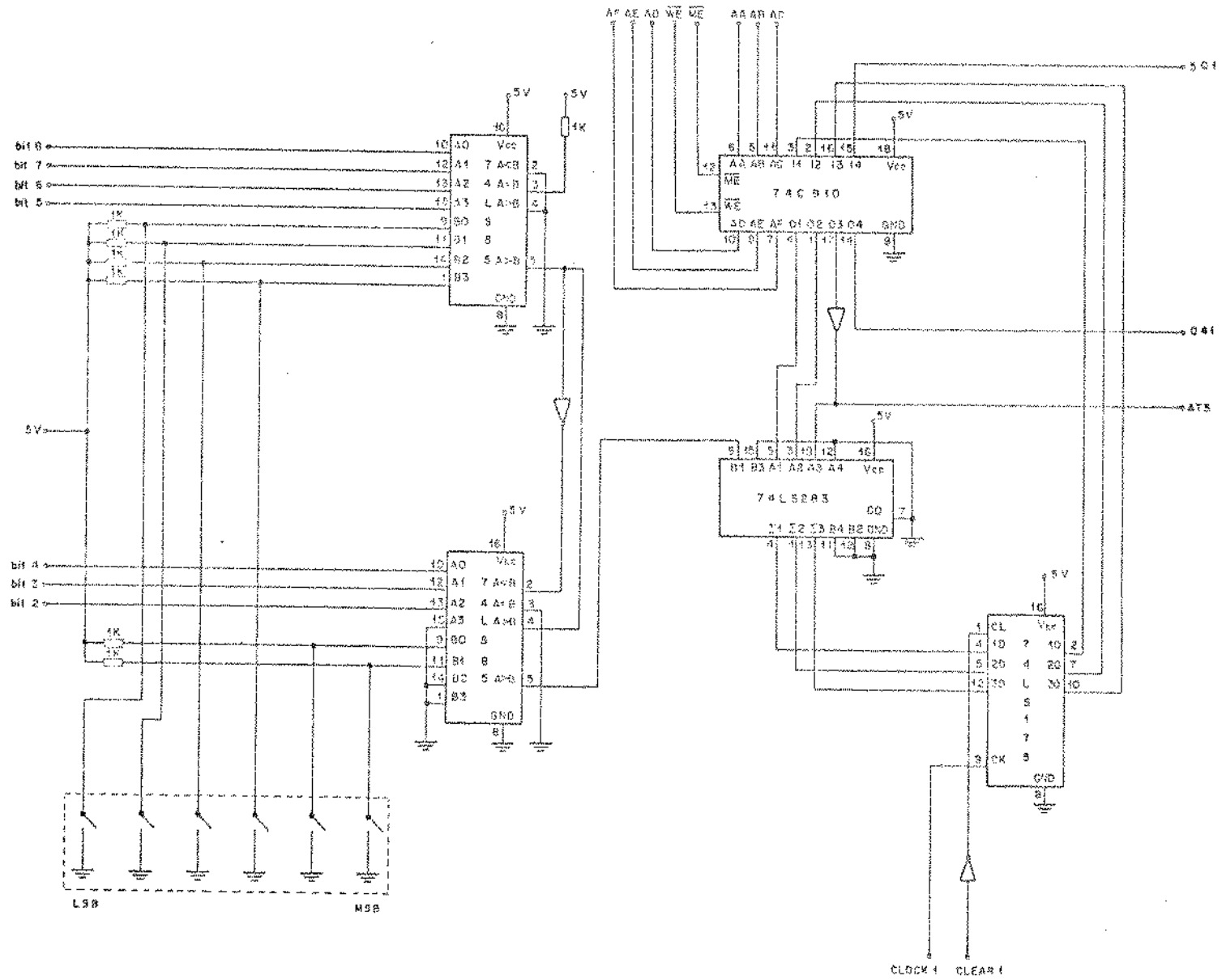
para $K' < 1$.

Substituindo-se esta expressão na eq.(A.1), e tendo em vista as eqs.(A.5) e (A.7), tem-se finalmente:

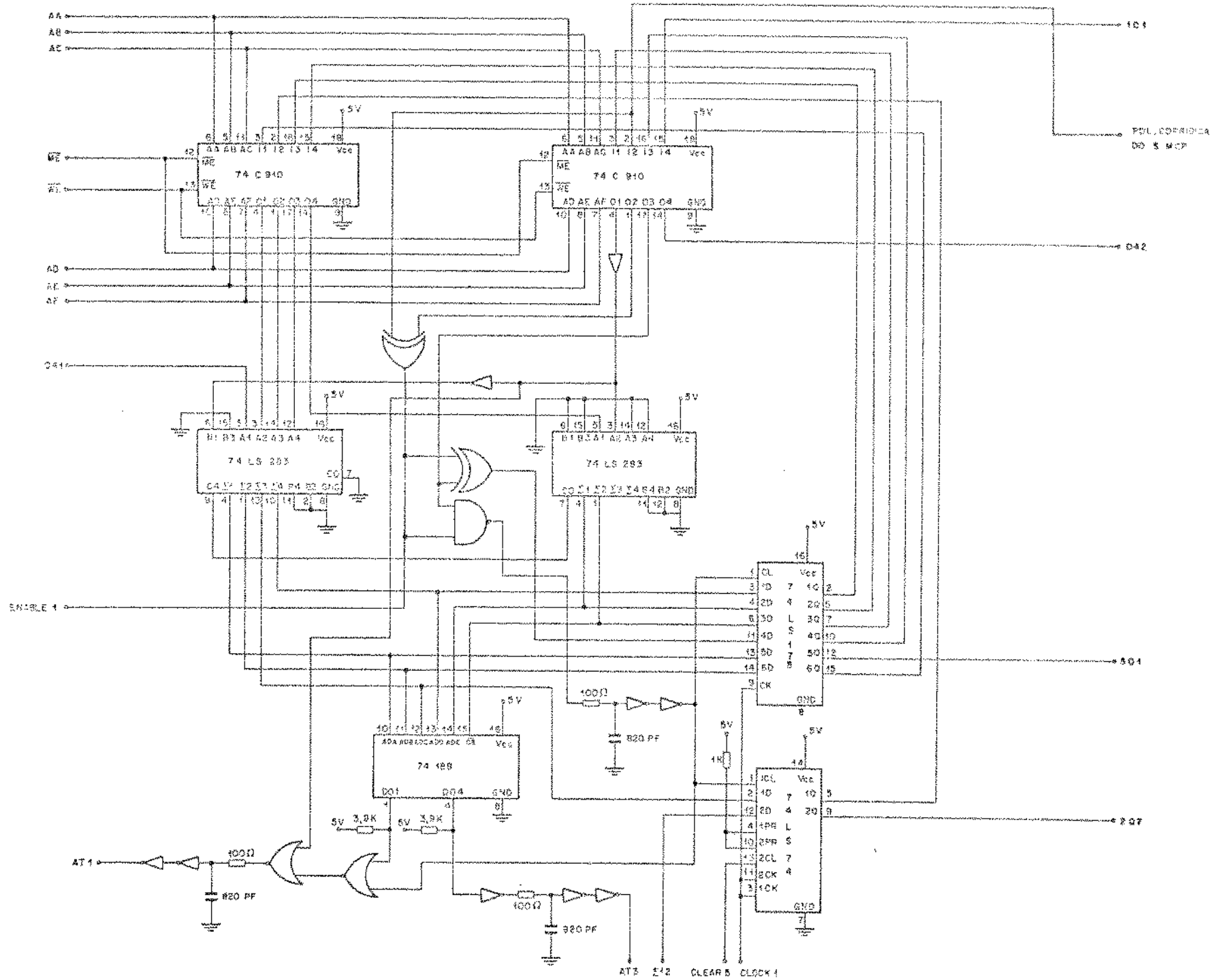
$$\begin{aligned}
 D_p = \alpha_r &= \left(\frac{\omega_1 T}{\pi} - \frac{\omega_1^2 T^2}{\pi^2} + \frac{T}{\pi^2} \int_0^T \left(1 - \frac{\tau}{T} \right) \frac{D_{24}}{(1-R^2)^{3/2}} \cdot \left[-\text{arc cos} \frac{D_{24}}{D_{22}} + \right. \right. \\
 &\quad \left. \left. = \frac{D_{22}^2 - D_{24}^2}{D_{24}} \right]^{1/2} d\tau \right)
 \end{aligned}$$

APÊNDICE B

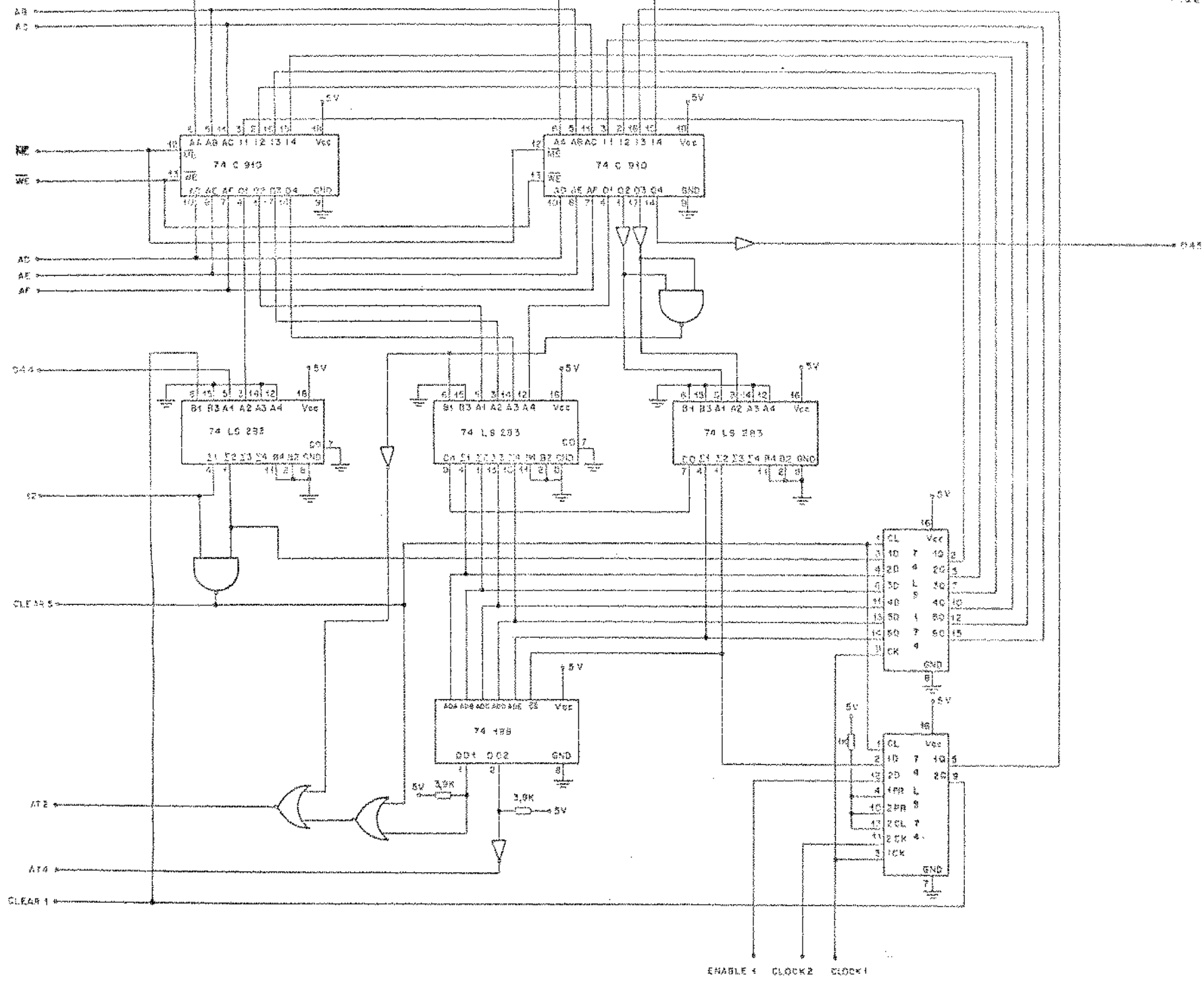
DIAGRAMAS DETALHADOS DE CIRCUITOS
NO DETECTOR DE VOZ



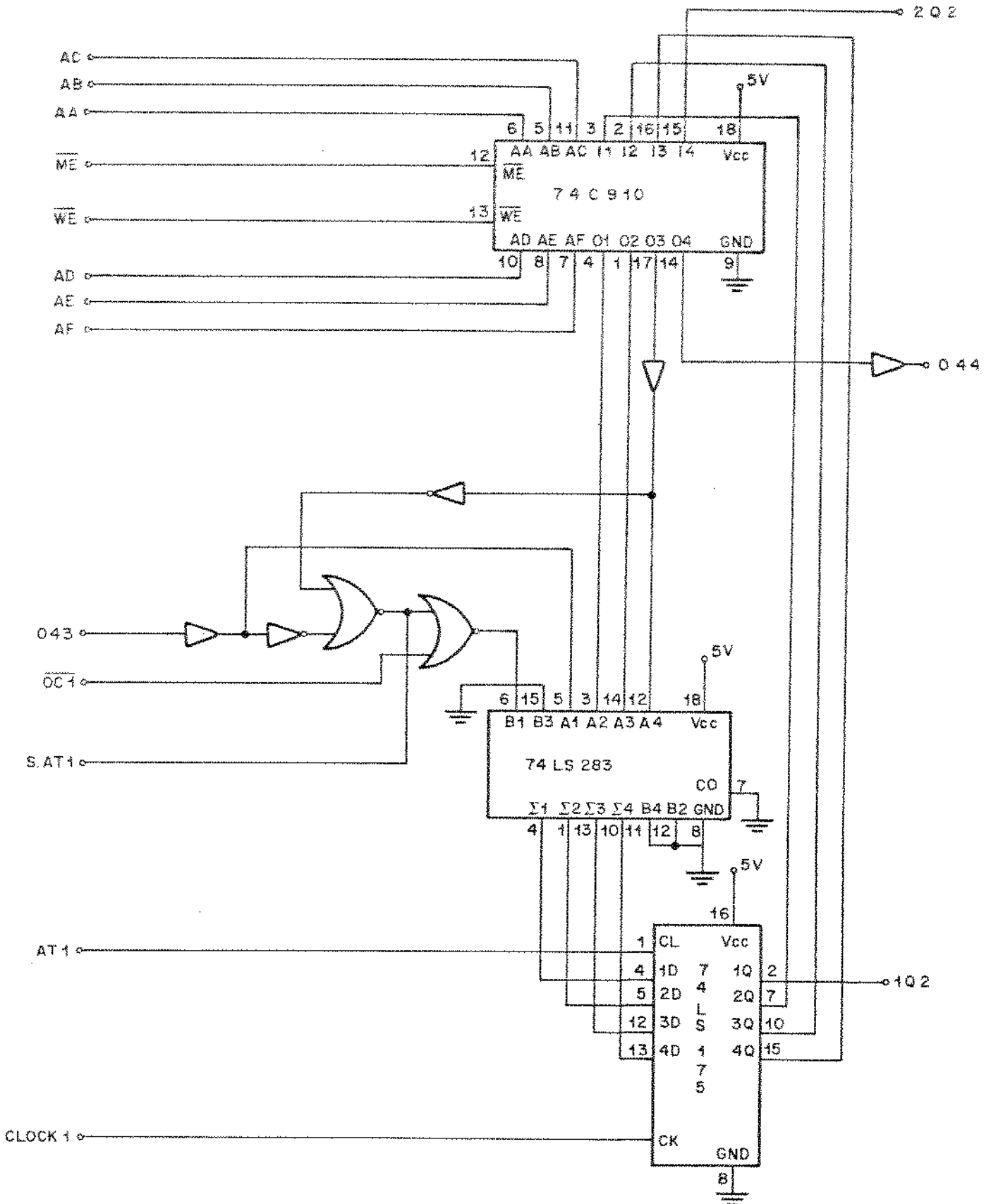
DETECTOR DE PICO



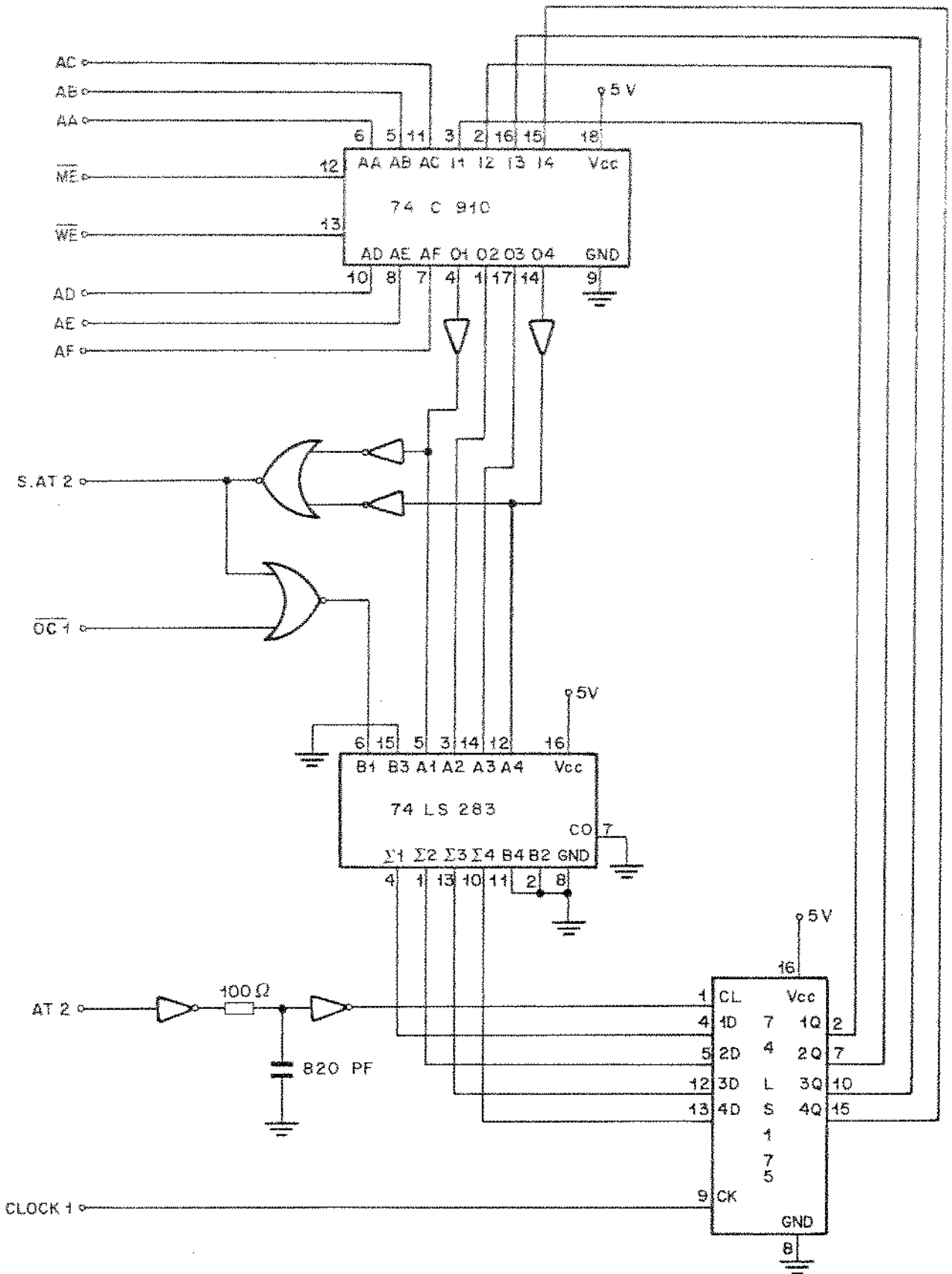
ANALISADOR DE CRUZAMENTOS DE ZERO ①



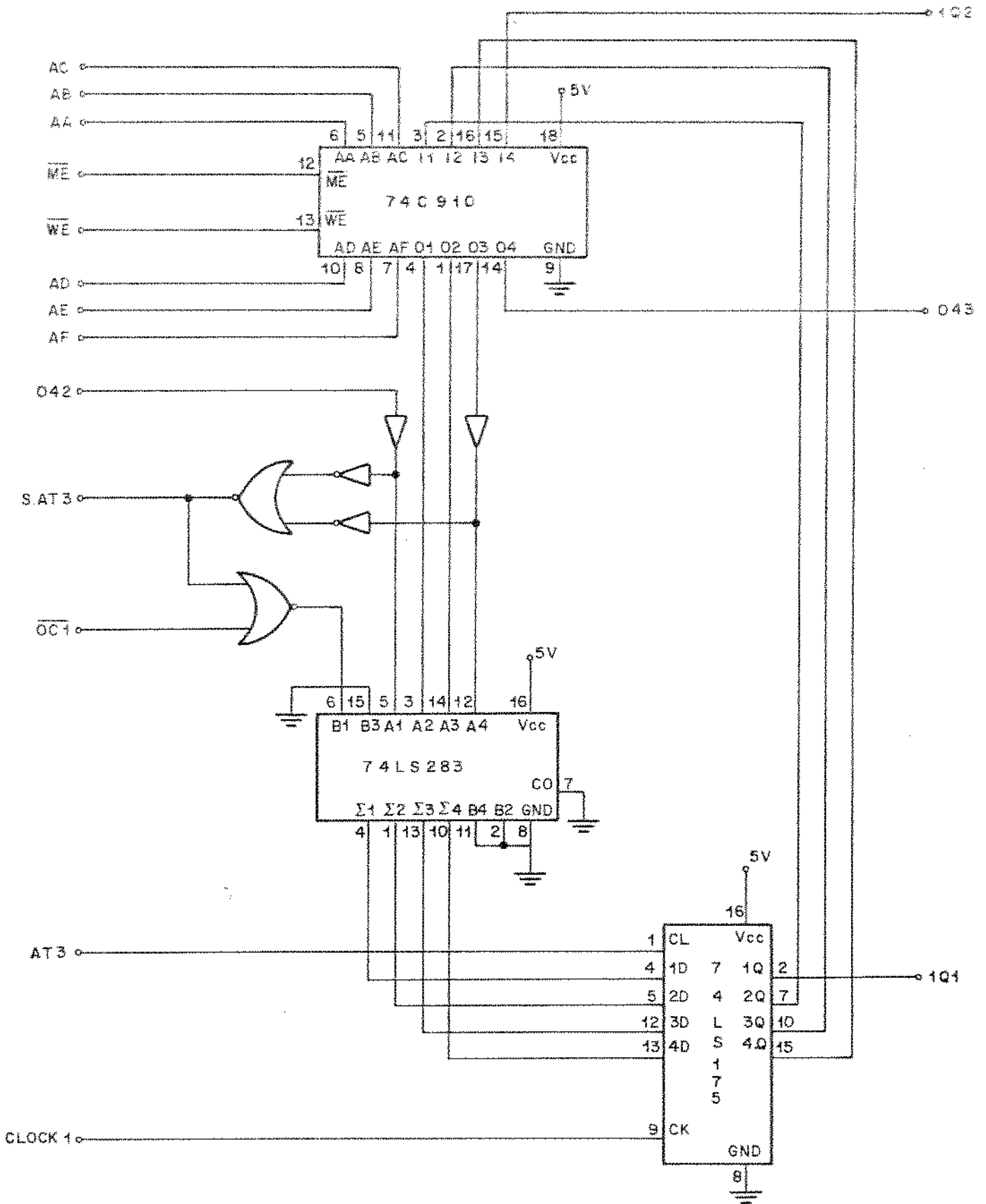
ANALISADOR DE CRUZAMENTOS DE ZERO (2)



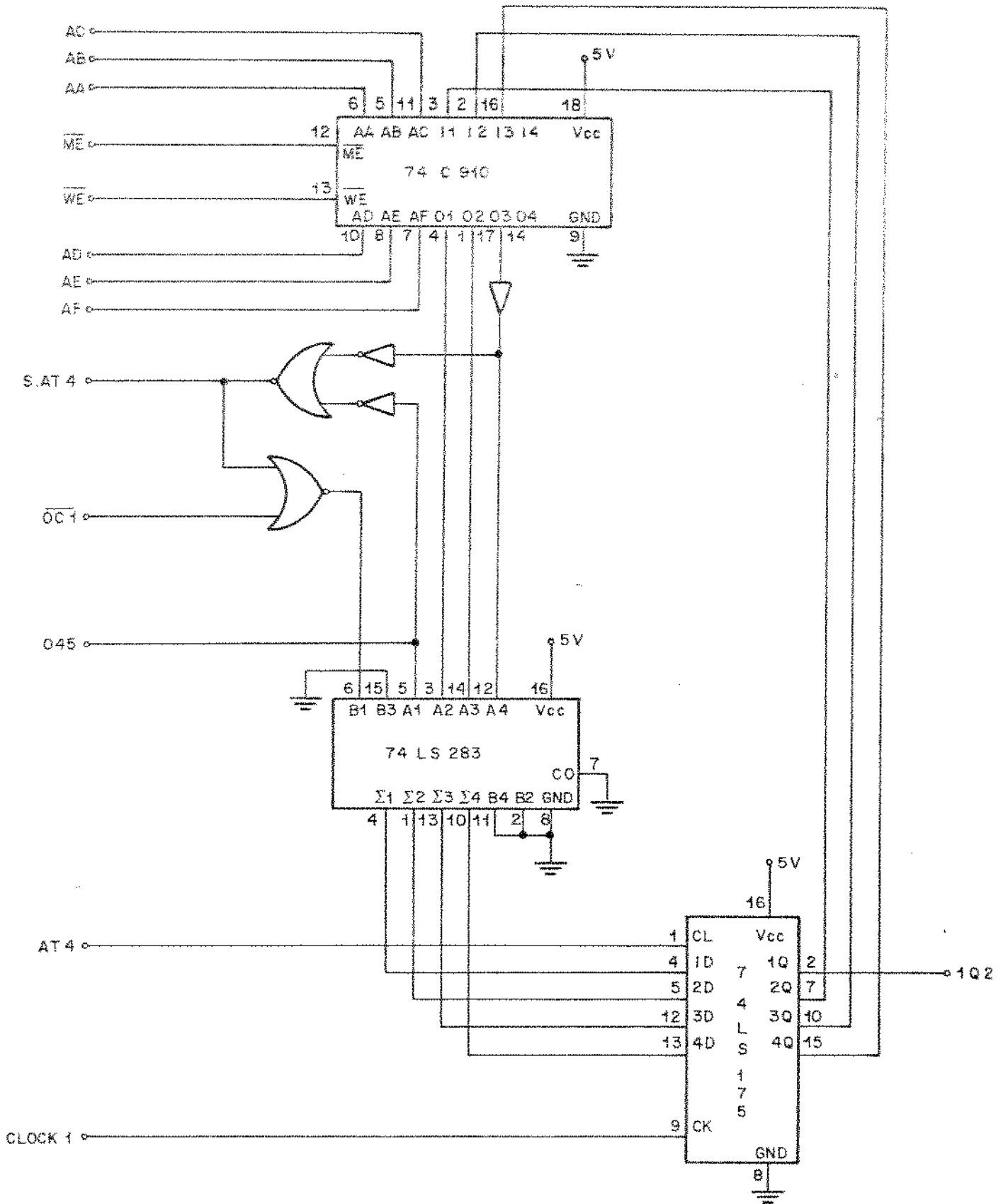
INIBIDOR DE RUIDO ①



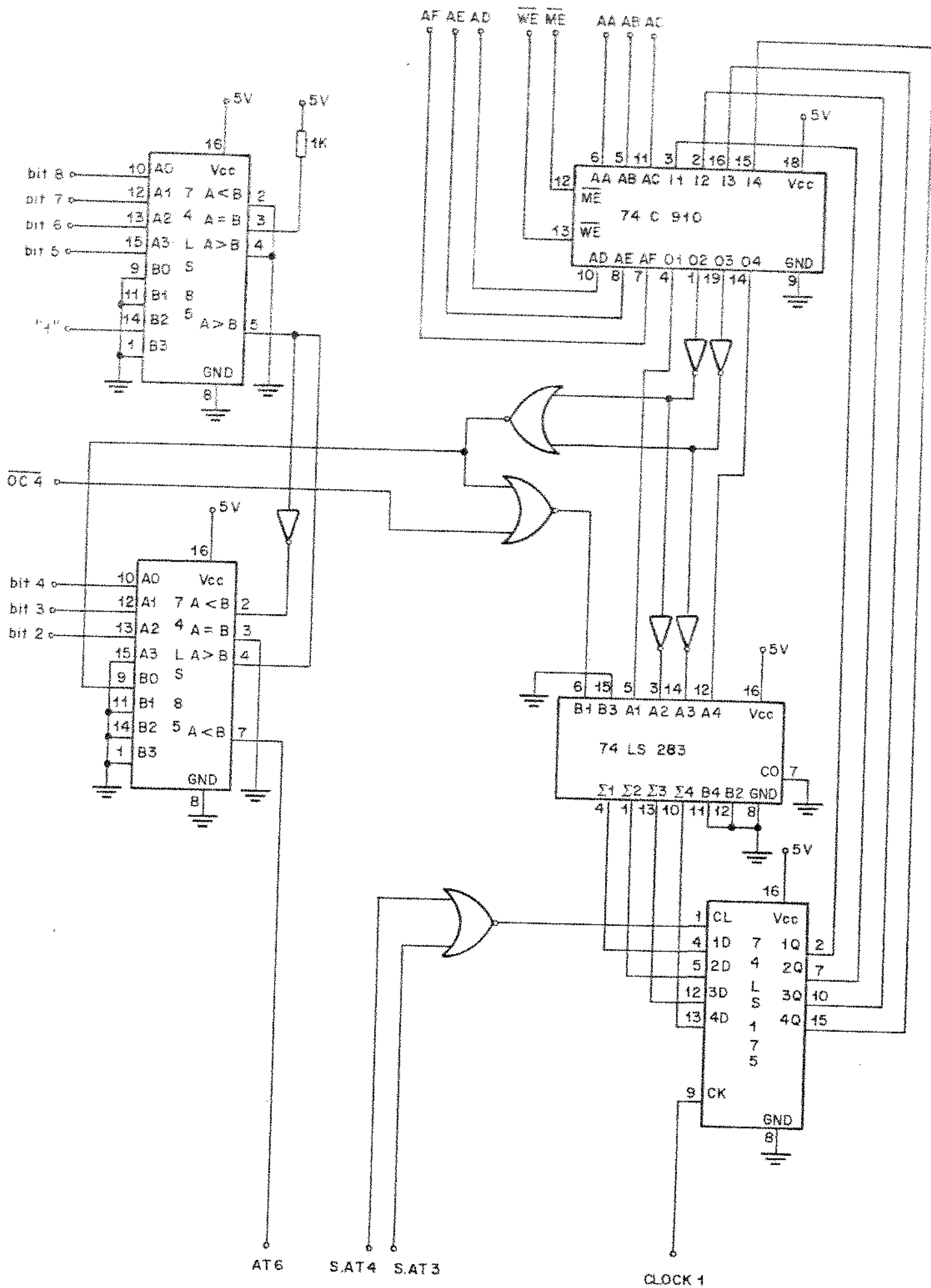
INIBIDOR DE RUIDO ②



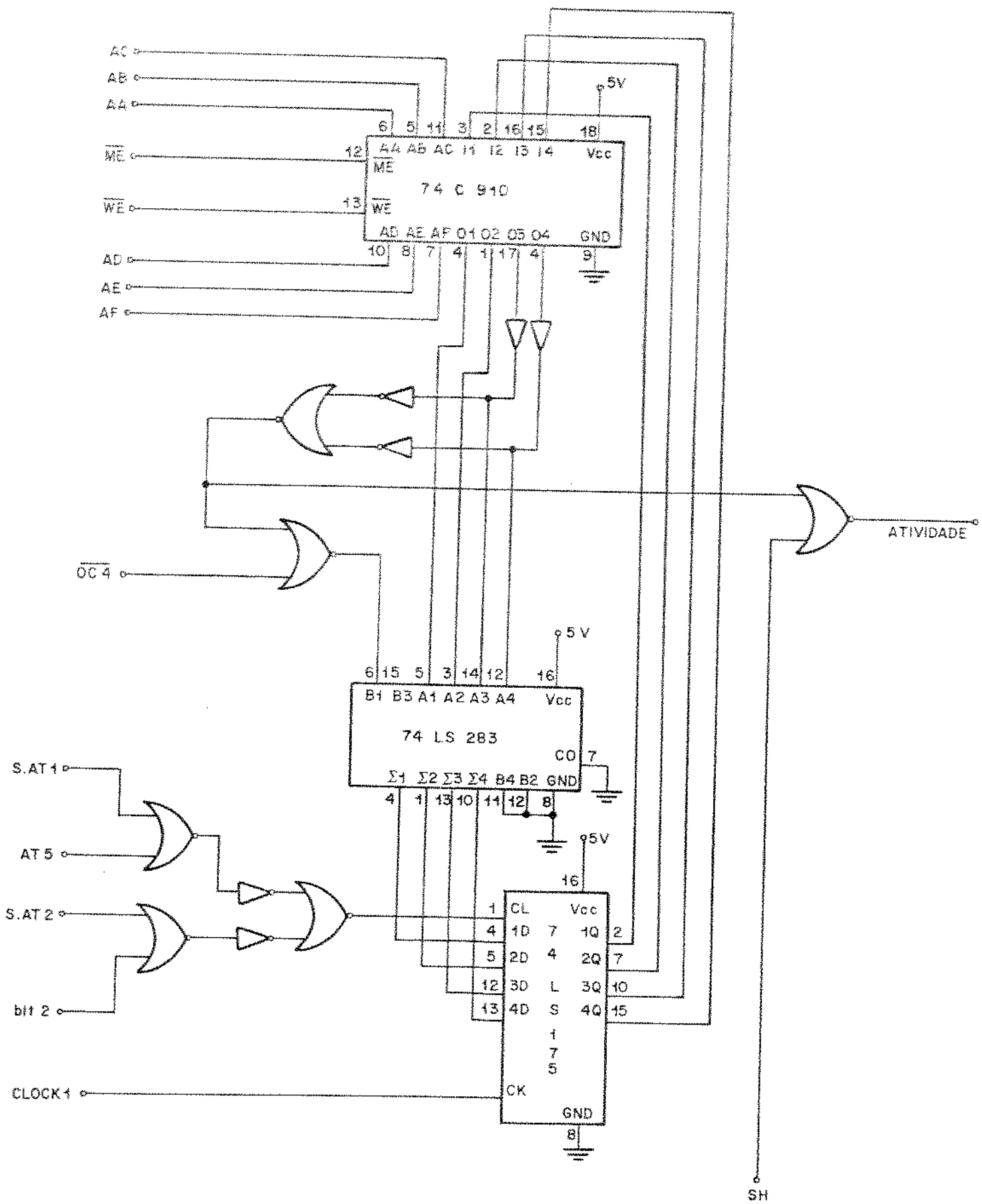
INIBIDOR DE RUIDO ③



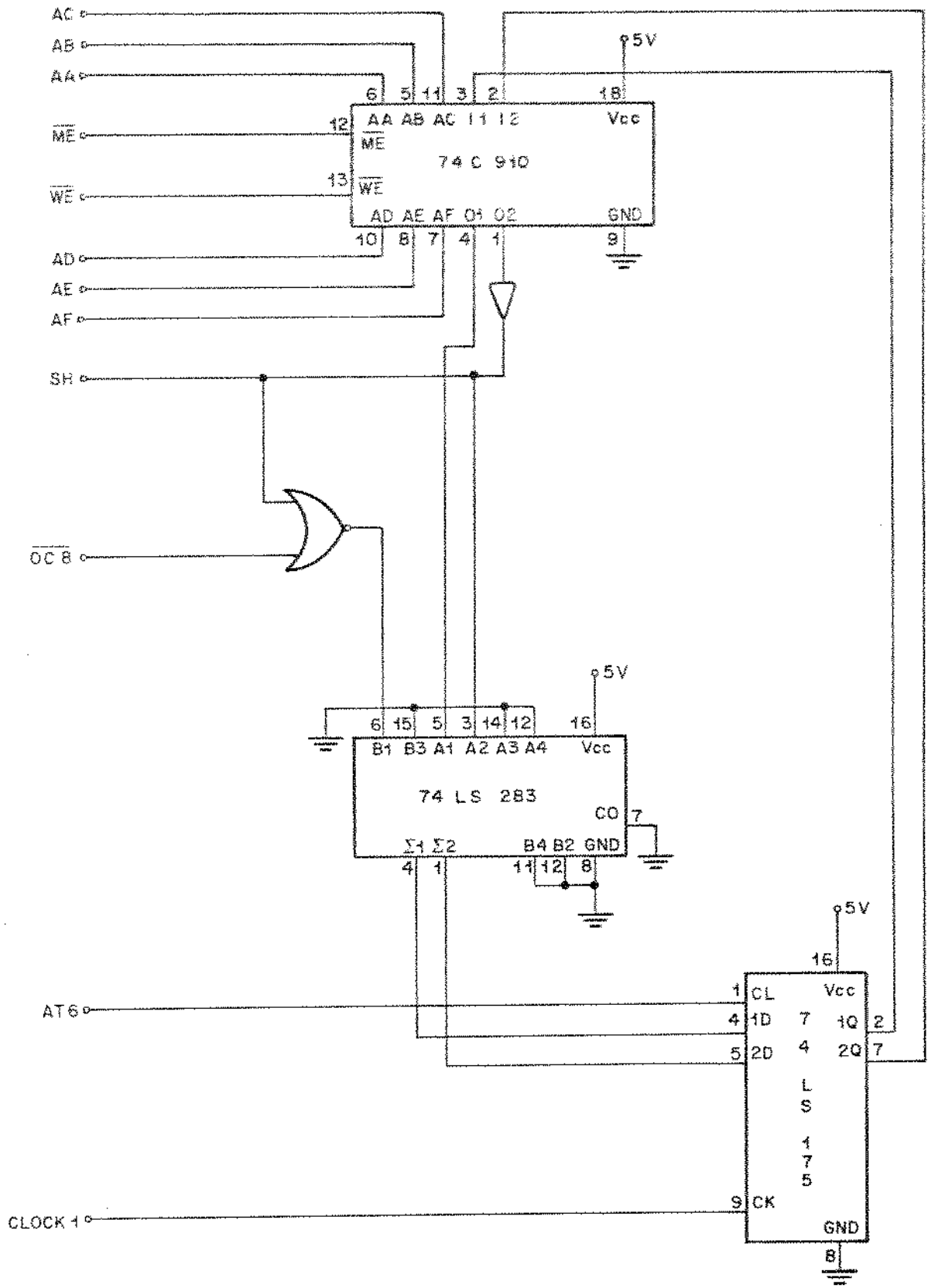
INIBIDOR DE RUIDO (4)



CIRCUITO DE PROTEÇÃO COM LIMIAR DE COMPARAÇÃO VARIÁVEL



HANGOVER ①



HANGOVER (2)

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] - D.L.RICHARDS, "Statistical Properties of Speech Signals", Proc. IEEE, Vol. 111, Nº 5, Maio 1964.
- [2] - "Transmission Systems for Communications", 4ª Ed., Bell Telephone Laboratories, Inc., 1970.
- [3] - J.M.FRASER; D.B.BULLOCK; N.G.LONG, "Over-All Characteristics of a TASI System", The Bell System Technical Journal, Julho 1962.
- [4] - K.BULLINGTON; J.M.FRASER, "Engineering Aspects of TASI", The Bell System Technical Journal, Março 1959.
- [5] - A.VIVALDI RODRIGUES; G.KANNO, "Sistema de Multiplexagem MCP de 60 canais à taxa de 2 Mbits", Dept. de Engenharia Elétrica da Faculdade de Engenharia de Campinas, RT-69, Fevereiro 1980.
- [6] - A.A.TERADA; A.V.RODRIGUES; G.KANNO, "Descrição do Sistema MCP-60 com 4 bits de prefixo por quadro", Dept. de Engenharia Elétrica da Faculdade de Engenharia de Campinas, RT-88, Outubro 1981.
- [7] - E.FARIELLO, "A Novel Digital Speech Detector for Improving Effective Satellite Capacity", IEEE Trans. on Comm., Fevereiro 1972.
- [8] - P.G.DRAGO/A.L.MOLINARI; F.C.VAGLIANI, "Digital Dynamic Speech Detectors", IEEE Trans. on Comm., Vol. 26, Nº 1, Janeiro de 1978.
- [9] - E.LYGHOUNIS; I.PORETTI; G.MONTI, "Speech Interpolation in Digital Transmission System", IEEE Trans. on Comm., Vol. 22, Nº 9, Setembro 1971.

- [10] - E.M.RIZZONI, "Compander Loading and Noise Improvement in Frequency Division Multiplex Radio-Relay Systems", Proc. of the IRE, Fevereiro 1960.
- [11] - The International Telegraph and Telephone Consultative Committee (CCITT), "Line Transmission", Blue book, 113 (International Telecommunication Union, 1965).
- [12] - The International Telegraph and Telephone Consultative Committee (CCITT), "Line Transmission", Orange Book, Vol. 111.1 (International Telecomm. Union, 1977).
- [13] - H.FLETCHER, "Speech and Hearing in Communication", D. Van Nostrand Co., Inc., New York, 1953.
- [14] - L.R.RABINER; R.W.SCHAFER, "Digital Processing of Speech Signals", Prentice-Hall, Inc., New Jersey, 1978.
- [15] - D.H.HAMSHER, "Communication System Engineering Handbook", McGraw-Hill.
- [16] - B.LÉVINE, "Fondements Théoriques De La Radiotechnique Statistique", Vol. 1, Éditions Mir, Moscou, 1973.
- [17] - D.F.HOTH, "Room Noise Spectra at Subscribers' Telephone Locations", Journal Acoustical Society America, Vol. 12, Abril 1941.
- [18] - D.F.SEACORD, "Room Noise at Subscribers' Telephone Locations", Journal Acoustical Society America, Vol. 12, Julho 1940.
- [19] - A.M.MOLINARI; F.C.VAGLIANI, ""Bit Rate per Channel Halving in PCM Multiplexes by Speech Interpolation and Adaptive Quantization", 1976, International Zurich Seminar on Digital Communications, Zurich, Março 1976.
- [20] - L.X.NEPOMUCENO, "Acústica", Ed. Edgard Blucher Ltda., S. Paulo, 1977.

[11. - M. NIEDEMA; M. G. SCHACHTMAN. "TASQ Quality-Effect of Speech
Detector and Interpolation", The Bell System Technical
Journal, July 1962.