Universidade de Brasília
Faculdade de Tecnologia

Departamento de Engenharia Elétrica
Brasília - DF

CRIPTOANÁLISE DE SINAIS DE VOZ CIFRADOS POR PERMUTAÇÃO DE SEGMENTOS TEMPORAIS

Autor: José Antonio Apolinário Junior

Prof. Orientador: Dr. Henrique Sarmento Malvar

Dissertação apresentada ao Departamento de Engenharia Elétrica da Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica.

Brasília, Junho de 1993.

TESE APRESENTADA AOS PESQUISADORES:

_____

Dr. Henrique Sarmento Malvar
UnB - Departamento de Engenharia Elétrica

_____

Dr. Luiz Antônio da Frota Mattos
UnB - Departamento de Matemática

_____

Dr. Joel Guilherme da Silva Filho
UnB - Departamento de Engenharia Elétrica

_____

Dr. Francisco Assis de Oliveira Nascimento
UnB - Departamento de Física

Vista e permitida a impressão.

_____

Joel Guilherme da Silva Filho

Coordenador de Pós-Graduação do Departamento de
Engenharia Elétrica da Universidade de Brasília - UnB

Brasília, Junho de 1993.

Aos meus pais,
José e Alaíde,
à minha esposa,
Ana Luisa,
e aos meus filhos,
Isabela e Eduardo.

AGRADECIMENTOS

O autor agradece ao Prof. Henrique Sarmento Malvar pela orientação, incentivo e amizade ao longo de todo o curso, bem como aos demais professores e companheiros do Departamento de Engenharia Elétrica da UnB que diretamente ou indiretamente contribuíram para este trabalho.

Agradece, ainda, ao Ministério do Exército - CIGE, por ter proporcionado a realização do Curso de Mestrado em regime de tempo parcial.

ÍNDICE

RESUMO

A Permutação de Segmentos Temporais, conhecida também pela sigla em inglês TSP ("time segment permutation"), é uma técnica tradicional de criptofonia usada em "scramblers" que, embora ofereça um grau relativamente pequeno de privacidade, não apresenta uma criptoanálise conhecida ou, pelo menos, amplamente divulgada.

Este trabalho apresenta uma proposta de criptoanálise baseada na mínima distância espectral entre as bordas de segmentos adjacentes. Esta abordagem deriva-se intuitivamente de métodos manuais usados no passado que tentavam reagrupar os segmentos embaralhados com base na visualização de seus espectrogramas.

O problema é o de como tornar inteligível um sinal de voz que foi dividido em blocos de N segmentos de mesmo tamanho e tais segmentos embaralhados bloco a bloco. São levantados parâmetros dos espectros do início e final de cada segmento, testadas algumas medidas de distância espectral e implementada uma busca exaustiva nas permutações possíveis.

Os resultados experimentais das simulações realizadas mostram que o esquema proposto é capaz de recuperar a inteligibilidade das mensagens cifradas pela permutação de segmentos temporais.

ABSTRACT

Time Segment Permutation (TSP) is a traditional method used in voice privacy systems (scramblers). Although TSP allows a relatively small degree of privacy, it doesn't have a known or, at least, widely spread cryptoanalysis.

This work presents a method of cryptoanalysis based in the minimum spectral distance between borders of adjacent segments. This approach is derived intuitively from manual methods used in the past, which tried to rearrange the scrambled segments based on the visualization of its spectrograms.

The problem is how to make intelligible a speech signal that was divided in blocks of N segments of the same size, with such segments shuffled one block after another. Spectral parameters at the beginning and end of each segment are estimated, some measures of spectral distance are tested, and an exhaustive search is implemented for the possible permutations.

The experimental results from simulations show that the proposed scheme is able to retrieve the intelligibility of messages ciphered by time segment permutation.

1. INTRODUÇÃO

A criptoanálise será passivamente entendida neste trabalho como sendo o ramo da criptologia que tenta extrair a informação existente numa mensagem cifrada sem o conhecimento da chave. Quando a mensagem em claro for um sinal de voz, o processo de cifrar esta mensagem (transformá-la num criptograma) é conhecido como criptofonia. A criptoanálise possui aplicações restritas a uns poucos aficcionados e, principalmente, a órgãos e agências governamentais que tratam esta disciplina de maneira sigilosa.

A criptofonia utilizando a permutação de segmentos temporais, conhecida também pela sigla em inglês TSP ("Time Segment Permutation"), é uma técnica tradicional usada em "scramblers" (misturadores de voz) que, embora sozinha ofereça um grau relativamente pequeno de segurança, não apresenta uma criptoanálise conhecida ou pelo menos amplamente divulgada.

O presente trabalho tem por objetivo apresentar uma proposta de criptoanálise de sinais de voz cifrados por permutação, bloco a bloco, de segmentos temporais de tamanho fixo ("TSP jumping window"). Trata-se de uma contribuição inicial para um possível sistema automático de criptoanálise de um sinal de voz que passou por um "scrambler".

A idéia básica é a escolha da permutação que apresenta a menor soma das distâncias espectrais das bordas de segmentos adjacentes. Para tanto, torna-se necessária a escolha de uma distância espectral que bem represente o sinal a ser criptoanalisado nas extremidades inicial e final de cada segmento temporal.

As principais técnicas de criptofonia serão abordadas no Capítulo 2, onde destaca-se a técnica de interesse para o presente trabalho. A seguir, encontra-se no Capítulo 3 um breve estudo das chamadas medidas de distância espectral que foram efetivamente usadas; tais medidas foram os parâmetros empregados no teste de encaixe entre segmentos para decidir qual a permutação que melhor se aproximaria do sinal em claro. O Capítulo 4, algoritmo de criptoanálise, apresenta o esquema proposto que consiste de uma preparação do sinal, da reordenação dos segmentos e de uma melhoria após a reordenação. O Capítulo

5 mostra alguns resultados obtidos em simulações bem como uma sucinta análise dos mesmos. Finalmente, o Capítulo 6 formaliza algumas conclusões.

As dificuldades encontradas para obter-se bons resultados neste caso simples de permutação de segmentos temporais apontam estratégias para a criptoanálise de outras variantes desta técnica de criptofonia.

2. TÉCNICAS DE CRIPTOFONIA

Têm-se notícia dos primeiros sistemas de criptofonia logo após o término da Primeira Guerra Mundial. Mais tarde, tais sistemas foram amplamente utilizados pelo Governo Americano, companhias telefônicas, meios diplomáticos e na Segunda Guerra Mundial [Cabral, 1987]. Contudo, somente após o advento dos semicondutores, e em particular dos sistemas digitais, veio a criptofonia a ter um grande impulso em direção às técnicas usadas nos dias atuais.

Podemos, inicialmente, dividir os sistemas criptofônicos em quatro classes, levando-se em consideração tipo de processamento e transmissão do sinal, conforme mostrado na Figura 2.1.

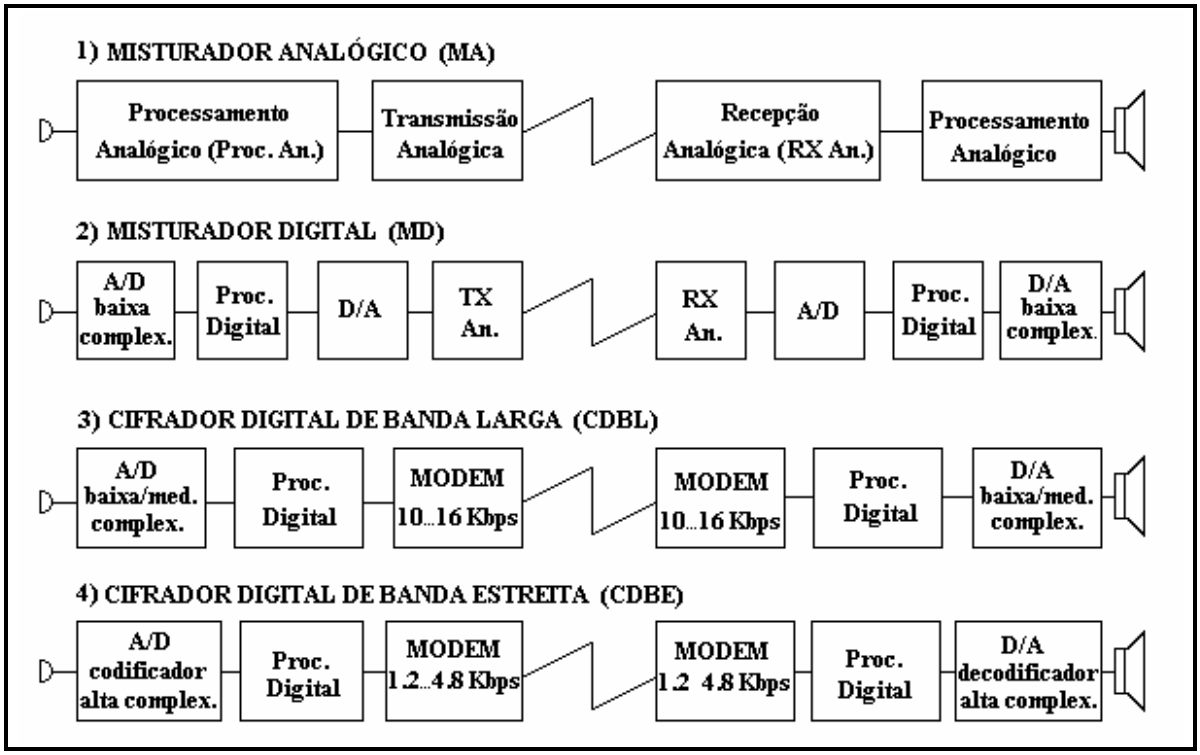

Fig. 2.1 - Classificação dos sistemas criptofônicos [Timmann, 1986].

Os sistemas das duas primeiras classes são conhecidos no mercado internacional pelos seus nomes originais em inglês SCRAMBLERS (misturadores). O sinal de voz original é transmitido após um processamento que mistura ou modifica alguns de seus parâmetros (amplitude, freqüência, seqüência, etc.). Isto possibilita um ataque criptoanalítico diretamente na forma de onda do sinal recebido e é admitido [Timmann,

1986] que a classe 1 (MA) é segura para um tempo menor que uma hora e a classe 2 (MD) para um tempo da ordem de vinte e quatro horas (supostamente, de acordo com os melhores recursos computacionais disponíveis em 1986). Tais sistemas são caracterizados por proporcionarem níveis de segurança de casual a tático[1] e são utilizados onde não se exija um nível de segurança estratégico ou onde não se deseja pagar o custo mais elevado de um outro sistema.

Os sistemas das classes 3 (CDBL) e 4 (CDBE) são conhecidos internacionalmente por VOICE CIPHER SETS (conjuntos cifradores de voz) ou mais informalmente por ComSec Systems; eles caracterizam-se por não transmitirem qualquer parte da voz original e podem, com uma transmissão digital, proporcionar a segurança de potentes algoritmos criptográficos (nível de segurança até estratégico). Os equipamentos da classe 3 (CDBL), embora podendo apresentar um excelente grau de segurança, possuem seu emprego limitado a canais que dispõem de largura de banda adequada para transmissões a velocidades elevadas.

Os equipamentos da classe 4 (CDBE) surgiram para vencer tal dificuldade, usando codificadores de voz de alta complexidade (VOice CODERS, por exemplo) e garantindo seu uso num canal de voz.

É interessante citar que os seguintes requisitos são desejáveis num sistema de criptofonia:

(1) transmissão dentro de um canal de voz (não deve haver expansão de banda além da permitida pelo canal);
(2) a voz cifrada deve ser ininteligível (não deve haver inteligibilidade residual) ao ouvido humano;
(3) resistência à criptoanálise;
(4) a voz decifrada deve ser de boa qualidade (inteligibilidade da voz e preservação das características do locutor);
(5) apresentar um retardo ("delay") de codificação máximo adequado a uma perfeita comunicação;
(6) custo compatível com o nível de segurança oferecido.

[1] O grau de segurança provido por um equipamento é medido pelo **nível de segurança** do processo, que pode ser classificado genericamente como **casual**, **tático** ou **estratégico**, conforme os recursos e tempo requeridos para a quebra do sistema ou criptoanálise de um criptograma.

Será apresentada a seguir uma outra classificação dos sistemas de criptofonia segundo o método utilizado [Cabral, 1987]:

CSI (Criptofonia por Segmentos de Informação):

- no domínio da freqüência (CSI-F)
- no domínio do tempo (CSI-T)
- bidimensionais (CSI-FT)

CBB (Criptofonia Bit a Bit)

CCI (Criptofonia de Características Informativas)

Esta classificação será melhor detalhada ao longo do capítulo, onde será enfatizada a técnica de interesse para este trabalho.

2.1. CRIPTOFONIA POR SEGMENTOS DE INFORMAÇÃO (CSI)

Neste grupo, enquadram-se aqueles procedimentos que dividem o sinal em elementos de informação e procuram, através de transformações e/ou permutações, eliminar a inteligibilidade do sinal de saída. A tentativa é de deixar o mais plano possível o espectro do sinal cifrado e ao mesmo tempo oferecer resistência a criptoanálise. Podemos, em princípio, alterar um dos seguintes parâmetros: amplitude, freqüência e posicionamento temporal. Como a alteração da amplitude, ou modulação, não é utilizada para fins de criptofonia [Cabral, 1987], veremis, então, tais técnicas no domínio da freqüência e no domínio do tempo.

2.1.1. CSI-F

Dentro desta classe encontramos desde um simples inversor de freqüências até os mais elaborados misturadores/inversores de sub-bandas ("band-splitters") variando no tempo de acordo com uma chave.

CSI-F pode apresentar inteligibilidade residual se o número de sub-bandas for muito pequeno e a dificuldade da criptoanálise depende tanto do número de sub-bandas como da

freqüência de troca da chave, controlada por um gerador de seqüência pseudo-aleatória. Pode-se afirmar que um sistema CSI-F de classe 1 (MA), exemplo típico dos primeiros equipamentos de criptofonia, oferece um baixo nível de segurança e seu emprego tende a ser limitado em situações onde deseja-se simplesmente proteção contra ouvintes casuais ou contra aqueles que não disponham de bons recursos para uma criptoanálise.

2.1.2. CSI-T

Estes sistemas efetuam a criptofonia através da permutação de N elementos temporais. Tais elementos podem ser amostras do sinal ("sample permutation") ou mais usualmente segmentos tomados dentro de blocos do sinal ("time-segment permutation"). Verifica-se que estes sistemas podem ser vistos como pertencentes à classe 2 (MD), oferecendo ainda níveis de segurança casuais ou táticos.

Observamos na Figura 2.2 os conceitos de bloco e segmentos. Os segmentos foram permutados dentro de um bloco; para que isto possa ser realizado, é necessário que todos os segmentos deste bloco sejam armazenados numa memória antes de serem transmitidos numa ordem diferente da original. Isto implica num retardo total de comunicação igual a duas vezes o tamanho do bloco (transmissão e recepção). Este retardo é uma das limitações do processo, bem como o número de segmentos em cada bloco: um número grande de segmentos diminuiria a inteligibilidade residual e aumentaria a resistência à criptoanálise, mas ao mesmo tempo causaria expansão de banda, necessidade de um sincronismo mais preciso e um efeito mais acentuado da superposição de segmentos[2] quando o sinal passa por um canal. Uma combinação de 8 segmentos num bloco de tamanho igual a 256 ms parece ser bem razoável para aplicações práticas e tais valores foram implementados nas simulações efetuadas. É interessante mencionar-se, também, que a permutação é diferente a cada bloco.

[2] Este efeito será comentado com mais detalhes no Capítulo 4.

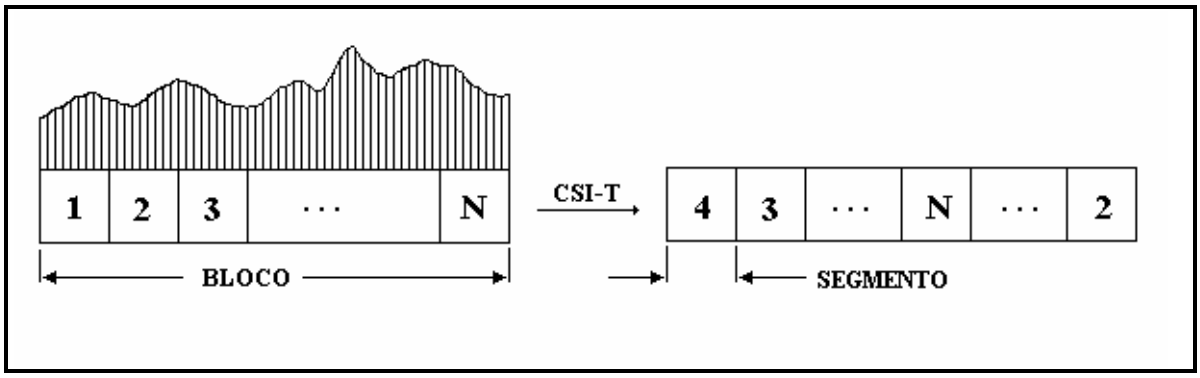

Fig. 2.2 - Sinal de voz dividido em blocos e segmentos.

Num sistema deste tipo, sabemos que o número máximo de permutações de N elementos é N!. Entretanto, devido à alta inteligibilidade de um sinal de voz permutado de maneira mais "simples", somente um número K << N! pode ser considerado efetivo. A estimação deste número efetivo de chaves K é difícil de definir ou avaliar pois o conceito de inteligibilidade é bastante subjetivo [Jayant, 1987]. Pode-se afirmar [Cabral, 1987] que para um sistema CSI-T usando um bloco com 8 segmentos, temos que das 8! = 40320 maneiras de se permutar os segmentos, somente cerca de 3000 serão efetivas. A busca somente nestas permutações efetivas torna-se uma estratégia de criptoanálise particularmente atraente quando a busca em todas permutações ou busca exaustiva não for viável (N muito grande).

A maneira como a permutação é feita pode variar, mas uma idéia básica é montar uma tabela com as chaves efetivas e usar tais chaves a comando de um gerador de seqüência pseudo-aleatória. Uma outra classe de permutações mais simples são as chamadas permutações uniformes ("U-permutations") definidas por:

$$s = k_1.r \pmod N; \quad r,s = 1, 2, ..., N \tag{2.1}$$

onde: N = tamanho do bloco (número de segmentos);

r = posição inicial do segmento;

s = posição final do segmento;

k_1 = chave (deve ser um número relativamente primo a N).

A decifração, neste caso, obedecerá a equação $r = k_2.s \pmod N$, onde $k_1.k_2 \pmod N = 1$. A Figura 2.3 ilustra este tipo de permutação para um bloco com 16 segmentos (N

= 16) e k_1 = 3. Observa-se que neste caso k_2 = 11 pois $k_1.k_2$ (mod N) = 3.11 (mod 16) = 1.

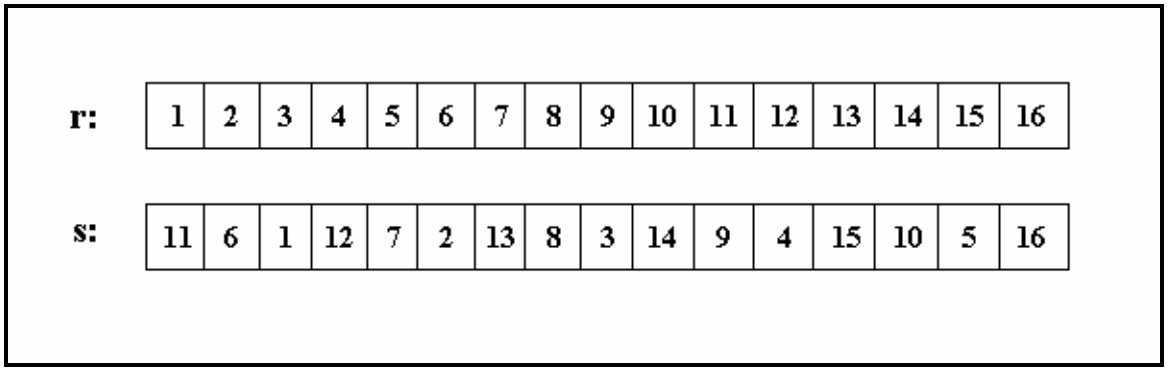

Fig. 2.3 - Exemplificação da permutação uniforme.

Esta permutação uniforme (de baixa resistência à criptoanálise) foi usada na simulação do sinal cifrado; entretanto, como veremos posteriormente, o esquema de criptoanálise que será proposto não leva isto em consideração e implementa uma busca exaustiva que torna a maneira como foi permutada irrelevante (poderíamos testar a criptoanálise num sinal em claro que o resultado seria o mesmo do cifrado).

Existem inúmeras variações do esquema apresentado ("jumping window") ; entre elas, pode-se citar o sistema que usa tamanhos de segmentos variáveis e aqueles que usam um bloco com um número fixo de segmentos e a saída por sorteio: o próximo elemento de entrada ocupa o local do segmento sorteado para transmissão e novo sorteio ocorre; a cada sorteio, um teste é feito para evitar que um segmento fique no bloco por um tempo superior ao correspondente a dois tamanhos de bloco ("sliding window").

A título de curiosidade, menciona-se que desde meados da década de 40 a criptofonia por segmentos de informação no domínio do tempo vem sendo utilizada. Naquela época, isto era feito com fitas de áudio em equipamento com várias cabeças magnéticas.

2.1.3. CSI-FT

Estes sistemas, ditos bidimensionais (geralmente é feito inicialmente uma inversão de freqüências ou deslocamento de sub-bandas, que pode ser processado analogicamente, e em seguida uma permutação de segmentos temporais), apresentam um bom desempenho em

relação aos CSI unidimensionais, em termos de inteligibilidade residual e resistência à criptoanálise, sendo que ainda hoje representam em muitos casos uma boa alternativa aos sistemas digitais; tanto que ainda são fabricados em considerável escala. Observa-se que, em casos especiais (comunicações em HF, por exemplo), justifica-se o uso de um sistema unidimensional (só CSI-T por exemplo) e a sua produção ainda existe hoje em dia.

Para finalizar-se os sistemas CSI, resta mencionar a possibilidade de inclusão de sinais de mascaramento. Um possível exemplo é a inclusão de uma sub-banda contendo ruído no caso de CSI-F.

2.2. CRIPTOFONIA DIGITAL

2.2.1. Criptofonia Bit a Bit (CBB)

Os sistemas CBB pertencem à classe 3 (CDBL) da classificação sugerida inicialmente e podem oferecer excelentes níveis de segurança, embora empregando uma elevada taxa de transmissão de bits. Estes codificadores de saída digital incluem tipicamente três elementos:

- um codificador de voz digital: transforma a voz em fluxo de bits;
- uma unidade de cifração e outra de decifração: cifra na transmissão e decifra na recepção o fluxo de bits;
- modem: transmite e recebe a informação digital cifrada sobre um canal analógico.

O sinal transmitido num sistema CBB é, pois, totalmente ininteligível (o sinal transmitido tem características de ruído) e a sua segurança depende do algoritmo criptográfico utilizado.

É característica destes sistemas o uso de codificadores de baixa complexidade (um modulador delta do tipo CVSD, por exemplo) e a principal limitação está na dificuldade de usá-los em canais de voz comuns dadas suas elevadas velocidades de transmissão.

2.2.2. Criptofonia de Características Informativas (CCI)

Os sistemas CCI usam técnicas de compressão de sinal de voz, tirando partido das possíveis redundâncias desse sinal e da complacência auditiva do ser humano, de modo a extrair e transmitir as características informativas em taxas menores que os sistemas CBB. Eles encaixam-se na classe 4 (CDBE), são considerados de elevado nível de segurança (até estratégico) e usam hardware e software de alta complexidade, o que eleva seus preços em relação aos demais.

São exemplos de tais sistemas os conhecidos VOCODERS cifrados, que alcançam taxas de transmissão de bits passíveis de serem usadas em canais de voz com banda de 0,3 - 3,4 KHz tais como telefone e rádio VHF, UHF. Pesquisas estão sendo desenvolvidas (já existindo alguns equipamentos no mercado internacional) visando a compressão do sinal de voz a taxas tão baixas que tornem os modens capazes de transmitir voz digital em linhas telefônicas de baixa qualidade ou links rádio de HF.

Além dos métodos tradicionais, vocoder de canal e vocoder LPC ("Linear Predictive Coding"), encontramos ADPCM com realimentação e CELP ("Code-Excited Linear Prediction") e suas variações com taxas de 8 a 4,8 Kbps.

Ainda caros e complexos, os sistemas CCI vem tornando-se mais atraentes à medida que progressos teóricos e tecnológicos reduzem seu custo e recuperam a qualidade da voz tida como sintética ou artificial.

2.3. A IMPORTÂNCIA DO SINCRONISMO

Crítico em alguns casos e não tão crítico em outros, o sincronismo é sempre um problema merecedor da atenção daqueles que trabalham com criptofonia ou com sua criptoanálise. Pode-se mencionar, a título de exemplo, que um perfeito sincronismo dos bits recebidos com os bits da seqüência pseudo-aleatória no caso de um receptor CBB módulo 2 é imprescindível para a recuperação do sinal.

Três maneiras consideradas básicas de efetuar-se o sincronismo em sistemas criptofônicos são as seguintes:

(1) Sincronismo inicial: todas as informações necessárias ao sincronismo são enviadas em uma salva no início da transmissão. Isto impede que uma estação que não tenha recebido a salva de sincronismo (supondo-se uma comunicação via rádio, por exemplo) venha a receber o resto da mensagem (a entrada atrasada não é permitida). Esta salva é composta de um conjunto de bits que no caso de CSI vêm normalmente modulados em FSK.

(2) Sincronismo contínuo: as informações de sincronismo são enviadas continuamente através de uma portadora piloto que ocupa uma pequena faixa da banda de áudio. Neste caso, a entrada atrasada é permitida a custo do sacrifício de uma fatia do espectro.

(3) Sincronismo intercalado: a transmissão do sinal de voz é periodicamente interrompida, total ou parcialmente e por uma pequena fração de segundo, para a transmissão de bits de sincronismo.

2.4. SUMÁRIO

Foram apresentados os diversos tipos de sistemas de criptofonia e algumas classificações que conduzem a uma divisão geral dos sistemas como aqueles que oferecem baixo grau de segurança (conhecidos no mercado mundial como SCRAMBLERS) e os que oferecem elevado grau de segurança (conhecidos como VOICE CIPHER SETS ou ComSec Systems). Os primeiros foram divididos em: aqueles que possuem processamento e

transmissão analógicos e os que possuem processamento digital e transmissão analógica.[3] Por outro lado, os sistemas criptofônicos que oferecem elevado grau de segurança foram divididos em: equipamentos com processamento digital e transmissão (digital) em taxas elevadas (larguras de banda maiores, mais simples e mais baratos) e os mais complexos que conseguem taxas de transmissão bem menores (larguras de banda menores e mais caros).

Neste capítulo, foi apresentada, também, a classificação segundo o método: CSI (Criptofonia por Segmentos de Informação nos domínios da freqüência, tempo e bidimensionais), CBB (Criptofonia Bit a Bit) e CCI (Criptofonia de Características Informativas). Foi ressaltado com um número maior de detalhes o método de interesse para este trabalho, CSI-T.

Tomando como fonte o Jane's Military Communications [1991-92], foi feito um levantamento dos equipamentos de criptofonia produzidos pelas grandes empresas mundiais. O resultado deste levantamento está resumido na Tabela 2.1.

PAÍS	FABRICANTE	CSI	CBB	CCI	NÃO ESPECIFICADO
Bélgica	Alcatel Bell STD SA	2	-	-	---
China	China National Elec. Imp.and Exp. Corp.	1	-	1	---
França	Thomsom CSF	1	1	-	---
Alemanha	Teltron GmbH	2	-	-	---
"	Tele Security Timmann	1	2	2	1 (usa espalham. de espectro)
Índia	Bharat Electr. Ltd.	1	1	-	---

[3]Não é descartada a existência de possíveis "scramblers" que possuem elevada complexidade de criptoanálise e possam ser considerados altamente seguros (CSI-FT com 12 sub-bandas e 16 segmentos temporais por blocos, por exemplo).

Tabela 2.1 - Equipamentos de criptofonia do mercado mundial (continua).

PAÍS	FABRICANTE	CSI	CBB	CCI	NÃO ESPECIFICADO
Iran	GAM Electronics and Comm. Ind.	3	-	-	---
Israel	Tadiran	1	-	-	1 (usa TDM eqp. multicanal)
Itália	Marconi Italiana	-	1	-	---
"	IRET	1	-	-	---
"	Bero Div. Elettron.	-	-	-	1
Holanda	Philips Crypto BV	-	1	1	---
Espanha	Tecnicas de Cifra SA	1	-	-	---
Suécia	Ericsson Radio Syst.	-	1	-	---
Suiça	Crypto AG	2	2	-	1 (multiplex)
Reino Unido	MEL	-	-	1	---
"	Marconi Secure Radio	-	6	1	---
"	Racal - Comsec Ltd.	3	3	1	---
"	Vigilant Comm. Ltd.	1	-	-	---
EUA	Time Space Proc. Inc	-	1	-	---
"	Racal Commun. Inc	-	1	-	---
"	Datotek Inc	1	2	2	---
"	Rockwell Internat. Collins Def. Comm.	2	1	1	---
"	Technical Com. Corp.	-	-	-	1

"	E-Syst., Garland Div.	-	-	4	---
"	Napco Intern. Inc.	1	1	-	---

Tabela 2.1 - Equipamentos de criptofonia do mercado mundial (continua).

PAÍS	FABRICANTE	CSI	CBB	CCI	NÃO ESPECIFICADO
EUA	Whittaker Elec.Syst.	-	-	-	1 (TDM)
"	Ocean Technology Inc.	-	1	-	---
"	Harris Corp, RF Div.	-	1	-	(com freq. hopping)
"	Intercon Syst. Corp.	-	1	-	---
"	Scientific Radio Sys.	1	-	-	---
"	GE Gov.Com. Syst.Div.	-	-	1	---
"	Motorola Inc.	-	1	1	---
"	ITT Defense Com.Div.	-	-	1	---
"	Honeywell	-	1	-	---
T O T A L		25	29	17	6
T O T A L G E R A L = 77 equipamentos					

Tabela 2.1 - Equipamentos de criptofonia do mercado mundial (continuação).

Da tabela acima, observa-se que 32,5% dos sistemas de criptofonia produzidos são CSI (a grande maioria bidimensionais e alguns com diferentes variantes), 37,6% CBB, 22,1% CCI e 7,8% não foram especificados ou não se encaixam exatamente na divisão apresentada. Não podemos esquecer que a tabela não apresenta níveis de demanda e, portanto, é possível que a cifra apresentada para os sistemas CSI aumentem consideravelmente caso seja levantada uma estatística do número de equipamentos produzidos, dado os baixos custos em comparação com os demais.

Uma outra observação é que a referência não considerou pequenas empresas e aquelas que produzem exclusivamente para uso interno em agências governamentais. Neste particular, pode-se mencionar a existência de equipamentos brasileiros e citar o CEPESC (Centro de Pesquisa para a Segurança das Comunicações) e a empresa ACRON como fabricantes de equipamentos de sigilo de voz na cidade de Brasília. Se compararmos a mesma fonte (1991-92) com a do ano de 1985, por exemplo, podemos verificar um razoável crescimento dos sistemas CBB devido em grande parte ao avanço tecnológico dos rádios táticos militares que passaram a dispor de transmissão de dados.

A Referência "SPEECH AND FACSIMILE SCRAMBLING AND DECODING" é um texto preparado para a publicação em 1946 pelo Summary Reports Group of the Columbia University Division of War Research que foi durante alguns anos considerado documento reservado e que apresenta algumas técnicas de criptoanálise de sistemas CSI-F e CSI-T de acordo com a tecnologia disponível na época. Sendo este o único documento encontrado a respeito de criptoanálise de sinais de voz cifrados e dada a quantidade de equipamentos CSI produzidos, conforme a tabela 2.1, justifica-se o presente esforço na criptoanálise do CSI-T, posto que afirma-se que tal sistema oferece um baixo grau de segurança mas não encontra-se, de acordo com o conhecimento do autor, publicada a quebra de tal técnica.

3. MEDIDAS DE DISTÂNCIA ESPECTRAL

Uma medida de distância espectral é a visualização através de um número não negativo do quanto o espectro de um sinal assemelha-se do espectro de um outro sinal: sinais iguais apresentam distância nula e tal medida aumenta quanto mais diferentes forem os espectros de tais sinais. Desde quando a predição linear tornou-se amplamente difundida como modelo para a produção de voz, o estabelecimento de uma medida de distância espectral baseada em dois conjuntos de coeficientes LPC ("linear prediction coefficients") passou a merecer uma grande atenção por parte dos pesquisadores.

Neste trabalho, é de interesse considerar-se a distância entre os espectros do final e do início de dois segmentos de sinal de voz. É, pois, necessário o levantamento de coeficientes capazes de descrever o espectro de um sinal numa análise a curto tempo para que seja possível comparar os espectros por meio da distância entre tais coeficientes nos dois instantes de interesse: antes e após a transição dos segmentos.

Encontra-se na literatura [Itakura, 1975, Gray e Markel, 1976, Tribolet, Rabiner e Sondhi, 1979, Gray, Buzo, Gray e Matsuyama, 1980, De Souza e Thomsom, 1982 e Brown e Rabiner, 1982] várias medidas de distância espectral dentre as quais destaca-se o trabalho de Itakura [1975] e de sua medida conhecida como "log likelihood ratio". Todas elas são amplamente usadas em processamento de voz e particularmente em criptofonia [Sridharan, Dawson e Goldburg, 1991] para medir objetivamente inteligibilidade residual e qualidade de voz recuperada.

Observa-se [Gray e Markel, 1976] que, para dois sinais de voz $\boldsymbol{x}$ e $\boldsymbol{y}$, a medida de distância d(x,y) deve satisfazer as seguintes propriedades:

1) $d(x,y) = d(y,x)$ (simetria);
2) $d(x,y) > 0$ para $x \neq y$ e $d(x,x) = 0$;
3) d(x,y) deve ter uma interpretação fisicamente significativa no domínio da freqüência;
4) Deve ser possível calcular-se d(x,y) num tempo compatível com a sua aplicação;
5) $d(x,y) \leq d(x,z) + d(y,z)$ (desigualdade triangular).

Serão apresentadas a seguir a maneira como foram estimados os coeficientes e as três medidas de distância espectral utilizadas neste trabalho: as distâncias Euclideana, de Itakura e Cepstral.

3.1. ESTIMAÇÃO DOS COEFICIENTES

Assumindo a estacionaridade em pequenos intervalos de tempo de um sinal de voz, a sua produção pelo aparelho fonador humano é usualmente modelada como a saída de um filtro "só polos" (modelo de síntese simplificado) excitado por um trem de pulsos quase periódico ou por ruído aleatório [Rabiner e Schaffer, 1978], como representado na Figura 3.1. A função de transferência de tal filtro é dada por:

$$H(z) = \frac{X(z)}{U(z)} = \frac{G}{A(z)} = \frac{G}{1 - \sum_{i=1}^{p} a_i . z^{-i}} \qquad (3.1)$$

onde p é a ordem do modelo, i.e., o número de coeficientes.

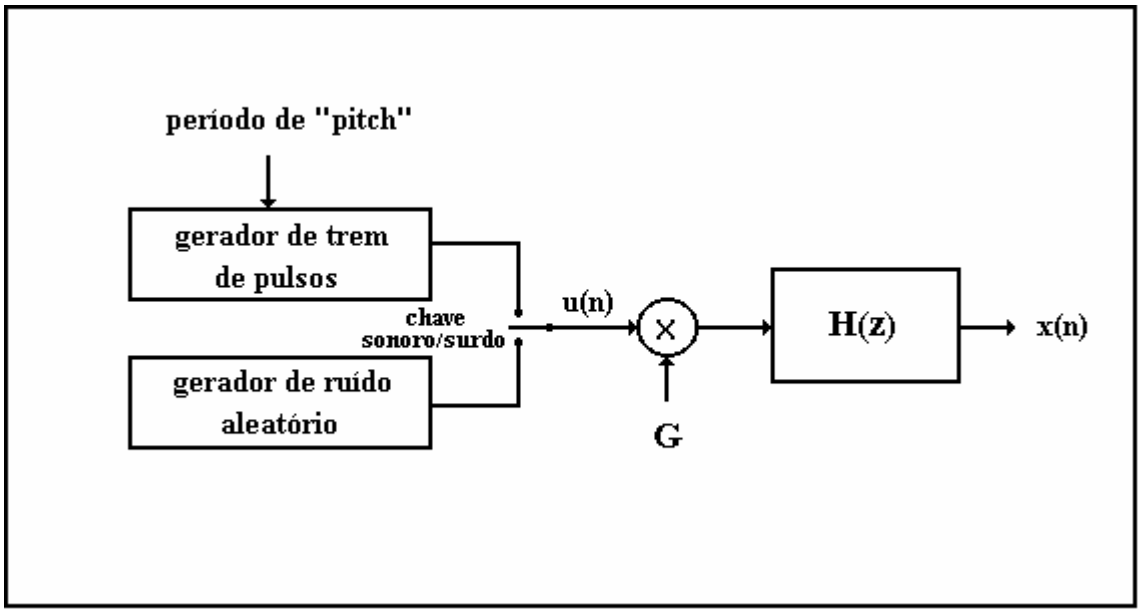

Figura 3.1 - Modelo simplificado de produção de voz.

O problema desta modelagem, conhecida como autoregressiva (AR), é a determinação do conjunto dos coeficientes $\{a_i\}$. Isto pode ser feito pela técnica de predição linear minimizando-se o erro médio quadrático de predição entre amostra atual x(n) e amostra predita $\hat{x}(n)$ (combinação de ***p*** amostras passadas). Este erro de predição corresponderá à excitação u(n) (multiplicada pelo ganho G) como vemos na equação de diferenças abaixo derivada da função de transferência (3.1):

$$e(n) = G.u(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^{p} a_i . x(n-i) \qquad (3.2)$$

A minimização de **e(n)** conduz a duas abordagens básicas, conhecidas como método da covariância e método da autocorrelação, para a estimação dos coeficientes preditivos lineares (LPC) $\{a_i\}$. Em ambos métodos algumas soluções são conhecidas, tais como a decomposição de Cholesky para o método da covariância e a solução recursiva de Levinson-Durbin para o método da autocorrelação [Rabiner e Schaffer, 1978].

Entretanto, os dois métodos consistem no cálculo dos valores de uma matriz de correlação e na solução de um conjunto de equações lineares. Dada a imperiosa necessidade de obtermos coeficientes que representam bem o espectro na extremidade de um segmento de voz e tendo em vista que nestes métodos os coeficientes são significativos para as amostras centrais da janela, o melhor que poderia ser feito é a utilização de uma janela assimétrica [Florêncio,1991] de modo a ponderar com maior peso as amostras mais próximas desta extremidade.

Optou-se, então, pela utilização de uma estrutura reticulada ("lattice") que, além de oferecer a possibilidade de obtenção dos coeficientes de modo adaptativo (amostra a amostra), apresenta encadeamento de seções idênticas, coeficientes com magnitude menor que 1, teste de estabilidade por inspeção e obtenção dos coeficientes direto das amostras de voz sem um cálculo intermediário da função de autocorrelação [Cowan e Grant, 1985].

Os coeficientes da estrutura reticulada, usualmente chamados coeficientes de reflexão $\{k_i\}$, independem da ordem **p** do filtro e podem ser transformados em coeficientes $\{a_i\}$.

A implementação em estrutura reticulada é apresentada na Figura 3.2. Observa-se que este filtro, conhecido em inglês como "feedforward lattice", implementa uma função de transferência só-zeros, como na seguinte equação do modelo de análise:

$$G.U(z) = X(z).A(z) \tag{3.3}$$

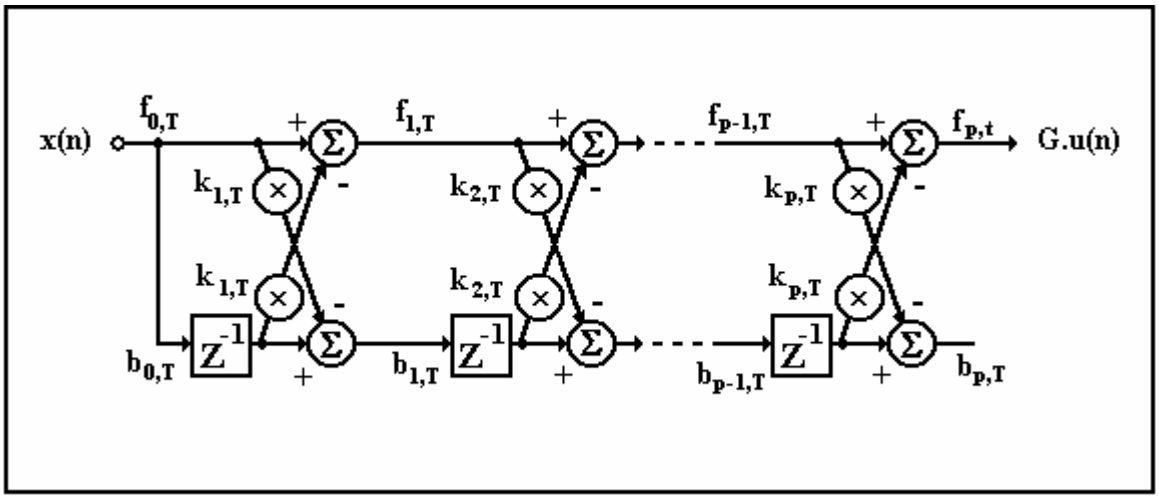

Figura 3.2 - Implementação em estrutura reticulada do filtro de análise.

Na estrutura reticulada, uma estimação recursiva gera a cada amostra de entrada novos coeficientes de reflexão e erros de predição, para cada estágio. A estimação recursiva dos coeficientes de reflexão pode ser feita usando diferentes algoritmos, dentre os quais menciona-se o algoritmo do gradiente adaptativo para estrutura reticulada e o algoritmo dos mínimos quadrados recursivo para estrutura reticulada (RLSL - "Recursive Least-Squares Lattice"). Este último apresenta uma solução exata ao problema dos mínimos quadrados e uma convergência mais rápida que os métodos baseados no gradiente [Cowan e Grant,1985].

O algoritmo utilizado para a estimação dos coeficientes de reflexão neste trabalho foi uma versão normalizada do RLSL conhecida como SQNLSL ("Square-Root-Normalized Least-Square Lattice"). Este algoritmo apresenta uma diminuição da complexidade das recursões[4] e uma melhoria nas propriedades numéricas das variáveis [Cowan e Grant, 1985]. Ele é apresentado na Figura 3.3.

[4]A forma é mais compacta embora a complexidade computacional seja maior.

O sinal de entrada x(n), antes de passar pelo algoritmo SQNLSL, sofreu uma preênfase por um filtro da forma $H_p(z) = 1 - \alpha.z^{-1}$, onde α = 0,9. É assumido, ainda, que o sinal de entrada x(n) possui média nula.

Parâmetros de entrada:

p = ordem máxima do filtro

μ = fator de ponderação exponencial ("memória") - foi usado o valor 0.98

σ = inicialização da variância do sinal para evitar divisão por zero (1 e-7 por exemplo)

x_T = amostra de entrada no instante T

Variáveis:

R_T = variância estimada de x no instante T

$k_{i,T}$ = coeficientes de reflexão do i-ésimo estágio no instante T

$fn_{i,T}$ = erro de predição para frente normalizado

$bn_{i,T}$ = erro de predição para trás normalizado

Inicialização:

$$R_0 = \sigma + x_0^2$$

$$fn_{0,0} = bn_{0,0} = \frac{x_0}{\sqrt{R_0}}$$

$$k_{i,0} = 0, \quad 1 \le i \le p$$

Para cada nova amostra de entrada, T=1 até o final

{ $R_T = \mu.R_{T-1} + x_T^2$

$$fn_{0,T} = bn_{0,T} = \frac{x_T}{\sqrt{R_T}}$$

Para cada estágio da estrutura reticula, i = 0 até min (T,p) - 1

{ $k_{i+1,T} = \sqrt{1 - fn_{i,T}^2}.\sqrt{1 - bn_{i,T-1}^2}.k_{i+1,T-1} + fn_{i,T}.bn_{i,T-1}$

$$fn_{i+1,T} = \frac{fn_{i,T} - k_{i+1,T}.bn_{i,T-1}}{\sqrt{1 - k_{i+1,T}^2}.\sqrt{1 - bn_{i,T-1}^2}}$$

$$bn_{i+1,T} = \frac{bn_{i,T-1} - k_{i+1,T}.fn_{i,T}}{\sqrt{1 - k_{i+1,T}^2}.\sqrt{1 - k_{i,T}^2}}$$

}

}

Figura 3.3 - O Algoritmo SQNLSL.

3.2. DISTÂNCIA EUCLIDEANA

Sejam **d** e **e** dois vetores contendo os coeficientes de reflexão estimados nos dois instantes de interesse (final de um segmento e início de um outro). A distância Euclideana simples (não ponderada) é dada por:

$$(\mathrm{de}(D,E))^2 = (\mathbf{d}-\mathbf{e})'.(\mathbf{d}-\mathbf{e}) \tag{3.4}$$

onde o apóstrofo (') significa transposição.

Com os mesmos vetores pode-se imaginar uma distância ponderada por uma matriz quadrada **W** escolhida de modo a minimizar a probabilidade de erro do processo[5]:

$$(\mathrm{dp}(D,E))^2 = (\mathbf{d}-\mathbf{e})'.\mathbf{W}.(\mathbf{d}-\mathbf{e}) \tag{3.5}$$

Esta nova medida é conhecida como distância de Mahalonobis quando **W** é o inverso da matriz de covariância do espaço de coeficientes. Entretanto, como essa inversa não é facilmente determinável, outras abordagens são preferidas [Brown e Rabiner, 1982].

Brown e Rabiner [Brown e Rabiner, 1982] sugerem a inclusão da informação de energia do sinal na medida de distância espectral. Esta informação será ***adicionada*** à distância não ponderada através de uma constante de escalonamento (α). Foi adotado neste trabalho o módulo do logaritmo da razão de variâncias dado por:

$$\mathrm{LER}(D,E) = \left| \log \frac{R(D)}{R(E)} \right| \tag{3.6}$$

onde: LER(D,E) é o módulo do logaritmo da razão entre as variâncias de D e E (do inglês "**L**og **E**nergy **R**atio"), R(D) é a variância do sinal em D e R(E) é a variância do sinal em E.

[5] Esta matriz deve ser simétrica e positiva definida para que a distância seja de acordo com as propriedades enumeradas no início deste capítulo.

Resultados práticos evidenciaram uma melhoria no desempenho da criptoanálise com o uso desta informação de energia na distância Euclideana:

$$de(D,E) = \sqrt{(\mathbf{d}-\mathbf{e})'.(\mathbf{d}-\mathbf{e})} + \alpha.LER(D,E) \qquad (3.7)$$

A Figura 3.4 mostra um trecho de programa em linguagem C que preenche uma matriz de distâncias espectrais entre as extremidades de segmentos para o caso de um bloco com oito segmentos. Observa-se que são admitidas todas as distâncias possíveis: extremidade direita do segmento i (0 a 8) para a extremidade esquerda do segmento j (1 a 8).

Os coeficientes da extremidade direita de um segmento foram calculados rodando-se um filtro SQNLSL no sentido direto das amostras do segmento e os da extremidade esquerda rodando-se o mesmo filtro no sentido inverso.

```
                    /*----------        Calculando as distâncias        ----------*/
/*  kd[m][n] = coeficiente índice m da extremidade direita  do segmento n
    ke[r][s]   = coeficiente índice r da extremidade esquerda do segmento s
    kd[m][0] = coeficiente índice m do ultimo segmento do bloco  anterior
    ke[0][s]  = informação de energia normalizada à esquerda do segmento s
    alfa         = constante de escalonamento  */
    alfa = 1.8 ;
    for (i=0 ; i<=8 ; i++)
   {  for (j=1 ; j<=8 ; j++)
     { /* d[i][j] = distância da direita do segmento i para a esquerda do segmento j  */
          d[i][j] = 0;
           for (z=1; z<=ordem; z++)
          { d[i][j] = d[i][j] + (kd[z][i]-ke[z][j])*(kd[z][i]-ke[z][j]);
           }
                 d[i][j] = sqrt(d[i][j]) + alfa * fabs(log10(kd[0][i])-log10(ke[0][j]));
     }  /* for j */
   }     /* for i */
```

Figura 3.4 - Cálculo das distâncias Euclideanas.

3.3. DISTÂNCIA DE ITAKURA

A medida de distância espectral desenvolvida por Itakura [1975] e conhecida como razão de verossimilhança logarítmica ("log likelihood ratio") tem, provavelmente, sido a mais popular medida de distância baseada nos coeficientes LPC. No nosso caso, esta distância é é dada por:

$$\mathrm{di}(D,E) = \mathrm{LLR}(D,E) = \log(\mathbf{a}_D \mathbf{V} \mathbf{a}'_D / \mathbf{a}_E \mathbf{V} \mathbf{a}'_E) \tag{3.8}$$

onde $\mathbf{a}_D$ e $\mathbf{a}_E$ são os vetores com os coeficientes LPC $(1, -a_1, \ldots, a_p)$ de D e E. **V** é a matriz de correlação de E (assumindo-se sinais de média nula) cujos elementos são definidos por:

$$v_{ij} = v_{ji} = \boldsymbol{E}[x(n).x(n+|i-j|)] \tag{3.9}$$

onde $\boldsymbol{E}[x]$ representa o valor esperado de x.

Tomando-se a equação (3.2) e achando-se $\boldsymbol{E}[e^2(n)]$ (resíduo de predição), verifica-se que este valor corresponde a **aVa'**. Logo, di(D,E) pode ser visto como o logaritmo da razão entre o resíduo de predição quando rodamos um filtro nas amostras de E usando coeficientes estimados em D, e o resíduo de predição de quando rodamos um filtro nas amostras de E usando coeficientes estimados em E. Evidentemente, o resíduo do numerador é maior e di[D,E] é maior que zero, sendo igual a zero somente se os coeficientes forem idênticos.

Observa-se em (3.8) que LLR(D,E) é diferente de LLR(E,D). A medida di(D,E) pode passar a ser simétrica se fizermos:

$$\mathrm{di}(D,E) = \frac{\mathrm{LLR}(D,E) + \mathrm{LLR}(E,D)}{2} \tag{3.10}$$

Tomando a interpretação de resíduo de predição e usando os coeficientes de reflexão $\{k_i\}$ ao invés de $\{a_i\}$, pode-se dizer que $a_D V a'_D$ corresponde a $\mathrm{Res}(X_D, K_E)$ ou resíduo obtido quando rodamos uma estrutura reticulada nas amostra X_E usando os k's estimados a partir das amostras X_D. A medida di(D,E) simétrica passa a ser dada por:

$$di(D,E)=\frac{1}{2}\log\left[\frac{\mathrm{Re}\,s(X_D,K_E)}{\mathrm{Re}\,s(X_D,K_D)}.\frac{\mathrm{Re}\,s(X_E,K_D)}{\mathrm{Re}\,s(X_E,K_E)}\right] \tag{3.11}$$

Finalmente, acrescentando a informação de energia nesta medida, chega-se à medida de distância espectral, baseada no trabalho de Itakura, a ser usada neste trabalho:

$$di(D,E)=\frac{1}{2}\log\left[\frac{\mathrm{Re}\,s(X_D,K_E)}{\mathrm{Re}\,s(X_D,K_D)}.\frac{\mathrm{Re}\,s(X_E,K_D)}{\mathrm{Re}\,s(X_E,K_E)}\right]+\alpha.\left|\log\frac{R(D)}{R(E)}\right| \tag{3.12}$$

A constante α, tanto para a distância Euclideana quanto para a de Itakura, deverá ser ajustada para obter-se o melhor resultado para cada sinal. O valor $\alpha = 2$ é, em geral, um bom início para constatar-se de imediato uma melhoria de desempenho em relação ao caso $\alpha = 0$, que corresponde ao uso só dos coeficientes LPC.

3.4. DISTÂNCIA CEPSTRAL

O cepstro de um sinal (ou *cepstro real* para contrastar da definição de cepstro complexo) é definido como a transformada de Fourier inversa do logarítmo da magnitude da transformada de Fourier deste sinal [Oppenheim e Schafer, 1989]; isto é:

$$c(n)=\frac{1}{2\pi}.\int_{-\pi}^{\pi}\log\left|X(e^{jw})\right|e^{jw}.dw \tag{3.13}$$

Por outro lado, considerando-se a modelagem do sinal de voz conforme (3.1), vemos que o espectro em magnitude logarítmico de x(n) é dado por:

$$\ln\left|X(e^{jw})\right|^2 = \ln\left|\frac{G.U(e^{jw})}{A(e^{jw})}\right|^2 = \ln\frac{G^2}{\left|1-\sum_{i=1}^{p}a_i e^{-jw}\right|^2} \tag{3.14}$$

Baseado em (3.14), podemos definir uma distância que representa a média quadrática das diferenças de dois espectros em magnitude logarítmicos [Markel e Gray, 1976]:

$$dc(D,E)=\frac{1}{2\pi}.\int_{-\pi}^{\pi}\left|\ln\left|X_D(e^{jw})\right|^2-\ln\left|X_E(e^{jw})\right|^2\right|^2.dw \qquad (3.15)$$

A integral acima pode ser aproximada para uma expressão bem mais simples se considerarmos os termos X_D e X_E pelas suas transformadas discretas de Fourier (DFT) e usarmos os coeficientes cepstrais $\{c_i\}$ [Markel e Gray, 1976][6]:

$$dc(D,E) = \left[c_D(0)-c_E(0)\right]^2 + 2.\sum_{i=1}^{\infty}\left[c_D(i)-c_E(i)\right]^2 \qquad (3.16)$$

Os coeficientes cepstrais serão obtidos a partir dos coeficientes $\{\mathbf{a_i}\}$ por:

$$c(n) = -a(n) - \frac{1}{n}\sum_{i=1}^{n-1}(n-i).c(n-i).a(i) \qquad (3.17)$$

para n > 0, com c(0) = 2.lnG, c(-i) = c(i), a(0) = 1 e a(i) = 0 para i > p.

De posse dos $\{k_i\}$, podemos achar os $\{a_i\}$ e destes os $\{c_i\}$ de forma recursiva a partir de (3.17). Isto é mostrado no trecho de programa em linguagem C da Figura 3.5.

Como pode-se observar em (3.17), os coeficientes $\{c_i\}$, $1 \le i \le p$, são unicamente descritos pelos coeficientes $\{a_i\}$ e c(0) contém a informação de energia. Isto, aliado ao fato da distância ser simétrica (dc(D,E) = dc(E,D)) e de ter-se originado a partir de um conceito significativo no domínio freqüencial, torna a distância cepstral bem interessante do ponto de vista das propriedades desejáveis de medidas de distâncias espectrais. É interessante ressaltar que, uma vez que os coeficientes cepstrais c(i) para i > p são nulos, o somatório de (3.16) é finito (i variando de 1 a p).

---

[6] Esta referência relaciona os coeficientes cepstrais com o espectro em magnitude logarítmico do seguinte modo: $\ln\left|X(e^{jw})\right|^2 = \sum_{i=-\infty}^{\infty} c(i).e^{-jwi}$

```
                    /* Efetuando transformações de coeficientes */
for (seg=1 ; seg<(nr_segs+1) ; seg++)   /* nr_segs = número de segmentos no bloco */
{   /* Step-up (gerando os ad's a partir dos kd's) */
    ad[1][seg] = 1;
    ad[2][seg] = kd[1][seg];
    for (inc=2; inc<=ordem; inc++)
    {  for (j=1; j<=inc; j++)
       {  jb = inc-j+1;
          b[j] = ad[jb][seg];
       }
       for (i=2; i<=inc; i++)  ad[i][seg] = ad[i][seg] + kd[inc][seg]*b[i-1];
       ad[inc+1][seg] = kd[inc][seg];
    }
    /* Step-up (gerando os ae's a partir dos ke's ) */
    ae[1][seg] = 1;
    ae[2][seg] = ke[1][seg];
    for (inc=2; inc<=ordem; inc++)
    {  for (j=1; j<=inc; j++)
       {  jb = inc-j+1;
          b[j] = ae[jb][seg];
       }
       for (i=2; i<=inc; i++)  ae[i][seg] = ae[i][seg] + ke[inc][seg]*b[i-1];
       ae[inc+1][seg] = ke[inc][seg];
    }
    /*  CÁLCULO DOS COEFICIENTES CEPSTRAIS
        Ex: cd[1][1] = informação de energia do segmento 1
            ce[2][3] = c(1) do segmento 3                    */
```

Figura 3.5 - Transformando coeficientes de reflexão $\{k_i\}$ nos coeficientes cepstrais$\{c_i\}$ (continua).

```
    /* Calculando os coeficientes cepstrais à direita */
     cd[1][seg] = log(kd[0][seg]+0.0000001);
     cd[2][seg] = -ad[2][seg];
     for (i=2; i<=ordem; i++)
    {  LP = i + 1;
       soma = i*ad[LP][seg];
       for (j=2; j<=i; j++)
       {  jb = i - j + 2;
          soma = soma + ad[j][seg] * cd[jb][seg] * (jb-1);
       }
       cd[LP][seg] = -soma/i;
    }
    /* Calculando os coeficientes cepstrais à esquerda */
    ce[1][seg] = log(ke[0][seg]+0.0000001);
    ce[2][seg] = -ae[2][seg];
    for (i=2; i<=ordem; i++)
    {  LP = i + 1;
       soma = i*ae[LP][seg];
       for (j=2; j<=i; j++)
       {  jb = i - j + 2;
          soma = soma + ae[j][seg] * ce[jb][seg] * (jb-1);
       }
       ce[LP][seg] = -soma/i;
    }
} /* for seg */
```

Figura 3.5 - Transformando coeficientes de reflexão $\{k_i\}$ nos coeficientes cepstrais $\{c_i\}$ (continuação).

3.5. SUMÁRIO

Este capítulo teve por objetivo apresentar as três medidas de distância espectral que serão utilizadas neste trabalho: as distâncias Euclideana, de Itakura e Cepstral. Foram apresentadas a modelagem do sinal e o algoritmo SQNLSL usado na estimação dos coeficientes de reflexão. A partir destes coeficientes, as medidas de distância foram adaptadas à necessidade de compararmos as extremidades esquerda e direita de segmentos de voz adjacentes.

4. ESQUEMA PROPOSTO

O esquema de criptoanálise proposto neste capítulo é aplicável num sinal de voz que foi cifrado pela permutação de N segmentos temporais de mesmo tamanho dentro de blocos transmitidos sequencialmente. Será assumido neste trabalho que N = 8. Esta informação, apesar das considerações do Capítulo 2, não é facilmente obtida diretamente do sinal. Na prática, existe ainda a incerteza sobre o método utilizado: "JUMPING WINDOW" ou "SLIDING WINDOW", como também visto no Capítulo 2. Contudo, observa-se que nos manuais de vários equipamentos ("SCRAMBLERS") de grandes fabricantes constam o método, o tamanho do segmentos em milisegundos e o número N de segmentos por bloco.

Os sinais de teste usados neste trabalho foram criptofonados com segmentos de 32ms que, amostrados a 8 KHz, resultaram em 256 amostras por segmento e 2048 amostras por bloco. Estes sinais, após serem transmitidos por um canal, devem ser preparados, criptoanalisados e melhorados ou tornados mais inteligíveis.

Em 4.1. apresentamos a metodologia utilizada para levantar o número de amostras por segmento e obter o sincronismo com os inícios dos segmentos. Em seguida (4.2.), mostramos como o sinal preparado passa pelo algoritmo de criptoanálise onde terá seus segmentos reordenados de acordo com a mínima distância espectral entre segmentos adjacentes. Em 4.3. apresentamos como podemos melhorar a inteligibilidade do sinal de voz criptoanalisado pela minimização do efeito de superposição de segmentos.

4.1 PREPARANDO O SINAL

O objetivo desta fase é a obtenção de um arquivo contendo o sinal de voz cifrado onde sabe-se que seu início coincide com o início de um segmento e conhece-se exatamente o número de amostras de cada segmento.

Inicialmente, são estimados os coeficientes LPC (k's e variância) amostra a amostra, via SQNLSL (vide Capítulo 3), do arquivo com o sinal cifrado e do arquivo com o sinal cifrado invertido no tempo. Isto feito, é calculado e gravado num outro arquivo (sinal.dst) as distâncias entre as amostras adjacentes conforme a Figura 4.1. Estas distâncias são obtidas da seguinte maneira:

(1) com os coeficientes estimados do sinal cifrado, obtém-se a distância (d_d) entre os coeficientes estimado na amostra atual e os coeficientes estimados na amostra anterior ;

(2) no caso dos coeficientes estimados a partir do sinal cifrado invertido no tempo, reinverte-se os mesmos primeiro e obtém-se uma nova distância (d_i) entre amostra atual e anterior;

(3) será considerada como distância (variação do espectro do sinal no instante daquela amostra em relação ao instante da amostra anterior) o maior valor entre d_d e d_i.

A idéia é a obtenção de picos de distância que possam indicar variações abruptas no espectro do sinal (prováveis transições de segmentos).

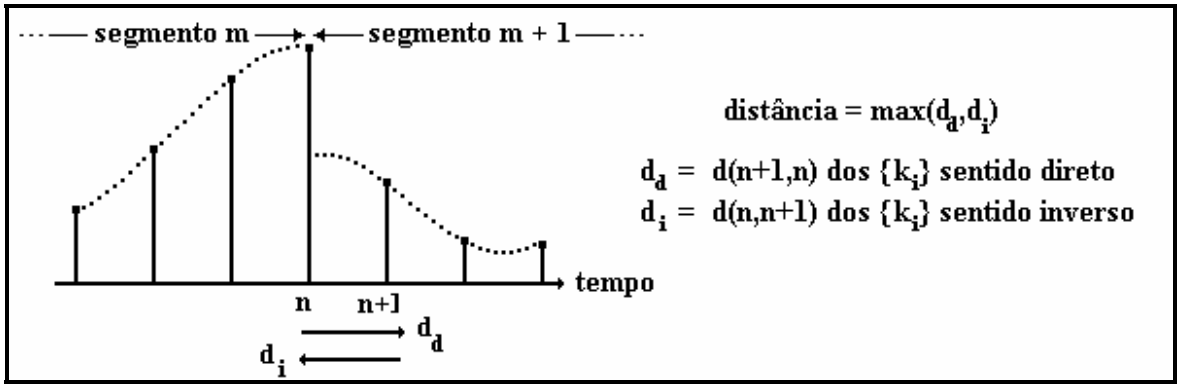

Figura 4.1 - Obtendo as transições de segmentos.

De posse do arquivo do sinal original e deste arquivo contendo as possíveis transições, consegue-se visualizar os segmentos e obter-se o tamanho dos mesmos. Um exemplo de um sinal CSI-T e suas correspondentes distâncias espectrais amostra a amostra são mostrados na Figura 4.2.

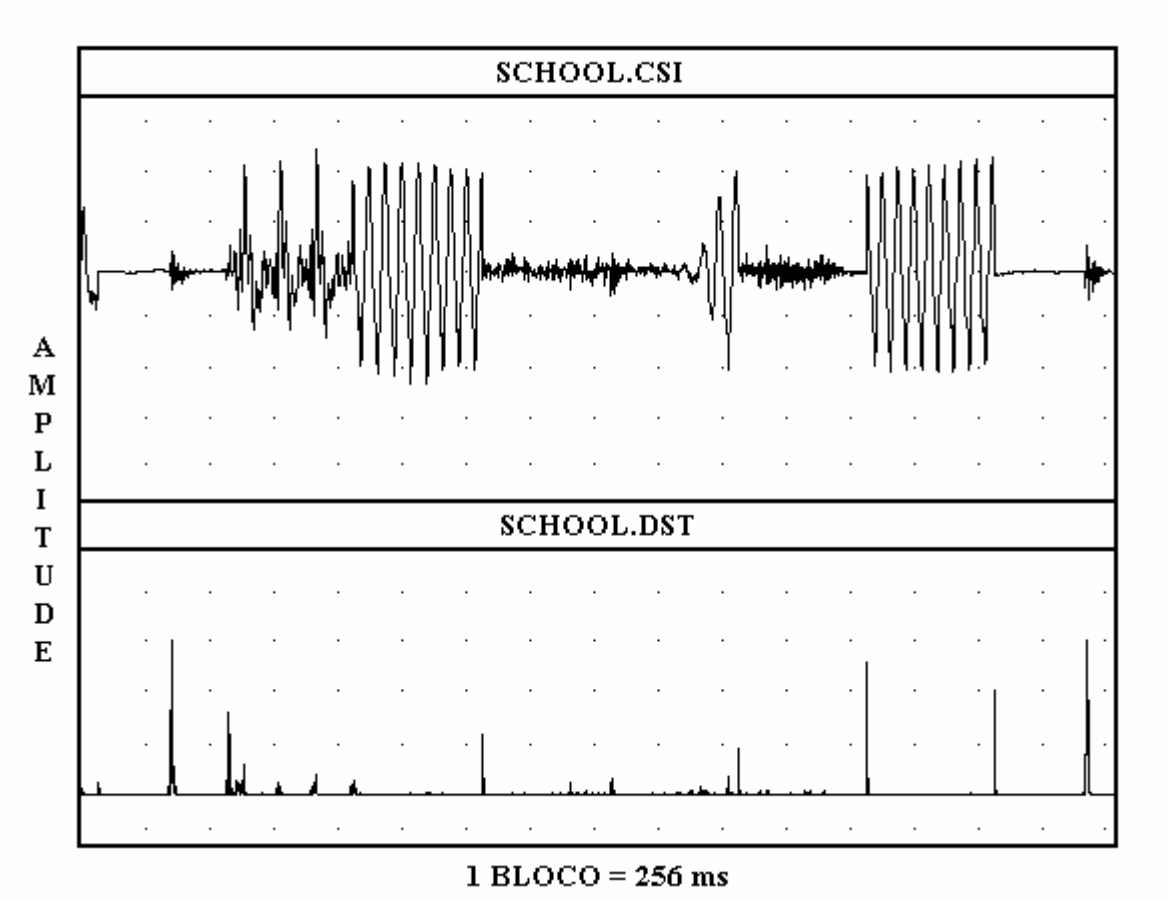

Figura 4.2 - Sinal de voz criptofonado CSI-T e suas distâncias espectrais.

Observa-se na Figura 4.2 que poderia ser estabelecido um certo valor limiar para decidir sobre a existência ou não de uma transição entre segmentos: se numa dada amostra a distância fosse maior que este limiar e um máximo local, seria decidido que haveria naquele instante uma transição. Se assim fosse feito, ocorreria provavelmente um considerável número de erros ao longo do sinal[7]. Entretanto, como sabemos que todos segmentos possuem o mesmo tamanho, podemos fazer a correlação do sinal school.dst com um trem de impulsos periódicos. À medida que o trem de impulsos "desliza", conforme indicado na Figura 4.3, haverá um certo momento onde a correlação será máxima, ou seja, os impulsos estarão em cima das transições dos segmentos.

[7] Estes erros são conhecidos na literatura clássica de deteção (teste de hipóteses) como de dois tipos: existia uma transição que não foi detectada ("miss") e não existia uma transição que foi detectada ("false alarm") [Van Trees, 1968].

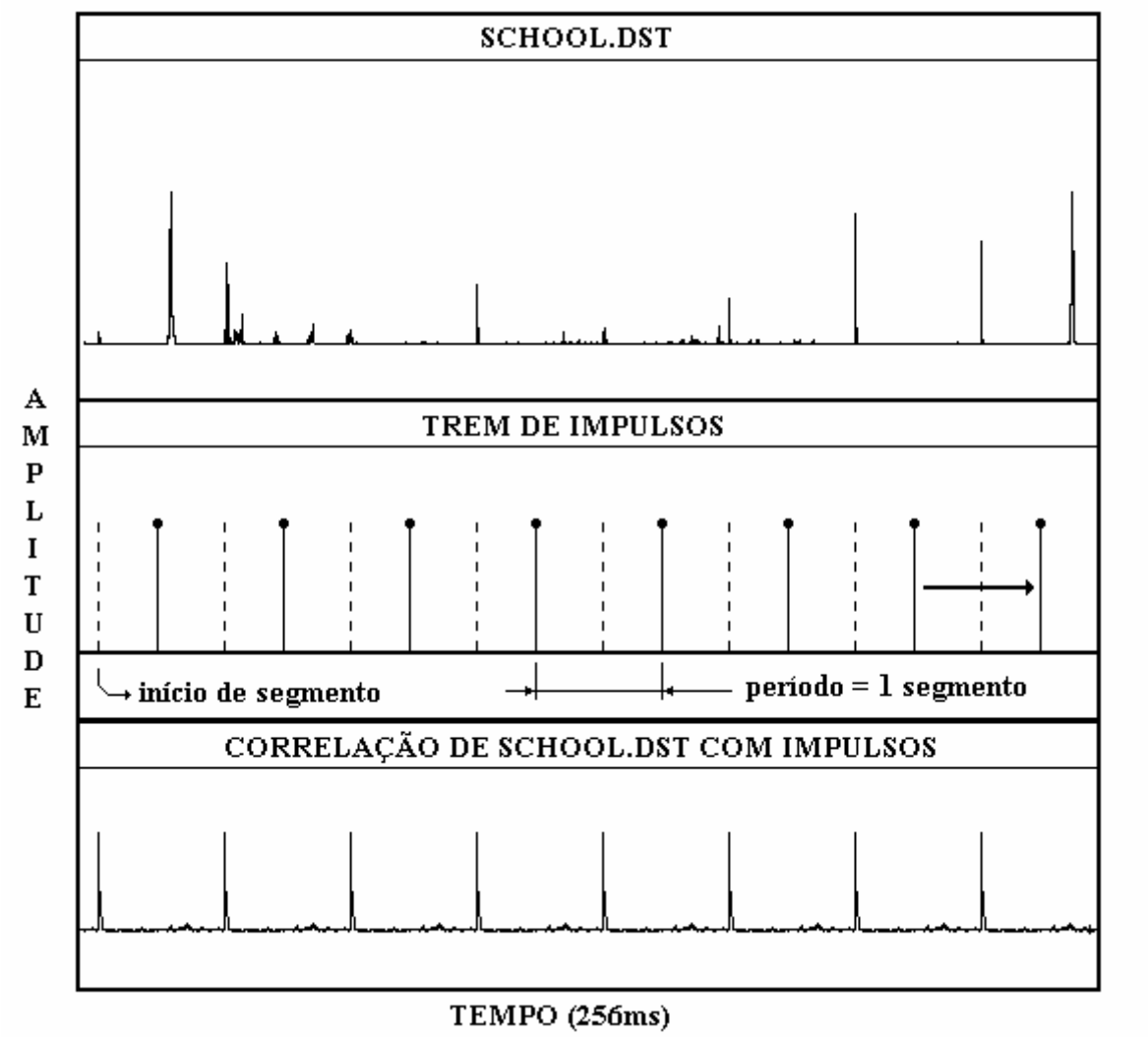

Figura 4.3 - Correlação de school.dst com um trem de impulsos periódico. Os impulsos (linhas contínuas) "deslizam" até encontrar os inícios de segmentos (linhas tracejadas) onde haverão máximos de correlação.

Fazendo-se um período variável, dentre limites estabelecidos por uma inspeção visual, e escolhendo-se aquele que corresponde à maior correlação, pode-se ter uma boa estimativa do tamanho do segmento, bem como do número de amostras entre o início do sinal e o início do primeiro segmento. Se o sinal a ser criptoanalisado for longo, experimentações mostraram que as pequenas diferenças entre as freqüências de amostragem dos equipamentos de transmissão e recepção ocasionam um tamanho de segmento não inteiro em número de amostras; este fato pode exigir acertos do tipo: eliminar uma ou duas amostras a cada intervalo de vários blocos.

Até aqui, tem-se um arquivo com o sinal preparado de tal forma que sabe-se precisamente o início de um segmento e o tamanho deste segmento. Entretanto, não é conhecido o início de um bloco. Esta informação, embora fundamental, não chega a ser uma preocupação, pois podemos executar a fase seguinte (reordenação dos segmentos) supondo que o um bloco inicia-se no primeiro segmento obtido, depois no segundo e assim por diante até o oitavo. Desta forma, um dos resultados será o melhor e corresponderá a uma suposição correta do início do bloco. O único inconveniente é efetuar a criptoanálise para oito arquivos sem saber qual deles está preparado adequadamente. Como veremos no próximo capítulo, o tempo de processamento é pequeno para o caso de oito segmentos e, portanto, este inconveniente não justifica uma análise do sincronismo do sinal cifrado para descobrir-se o primeiro segmento a priori.

4.2 REORDENANDO OS SEGMENTOS

O algoritmo apresentado na Figura 4.4 mostra como será processada a reordenação dos segmentos. Para cada bloco (8 segmentos) lido são estimados os coeficientes $\{k_i\}$ à direita e à esquerda de cada segmento. Para os coeficientes à esquerda, roda-se um filtro SQNLSL do final para o início do segmento. Os coeficientes à direita são obtidos com o mesmo filtro rodando do início para o final do segmento.

Os coeficientes estimados são armazenados e as distâncias do final para o início de dois segmentos são calculadas e armazenadas numa matriz contendo todas as distâncias possíveis entre os segmentos do bloco.

```
        /*---------- ALGORITMO CSITOVOZ  (CRIPTOANÁLISE DE CSI-T) ----------*/
INÍCIO  CSITOVOZ (ENTRA:SINAL.CSI, SAI:SINAL.VOZ)
    "DECLARAÇÃO DE ARQUIVOS E VARIÁVEIS";
    "ABERTURA DE ARQUIVOS";
    "LEITURA  E ESCRITA DE CABEÇALHOS";
    "INICIALIZAÇÃO DE VARIÁVEIS";
    ENQUANTO  NÃO (FIM DE ARQUIVO)
        "LER  UM BLOCO DO ARQUIVO DE ENTRADA":
        PARA  SEG DE  1  ATÉ  NR_SEGS PASSO 1
            "CALCULAR  Kd  (COEFICIENTES À DIREITA DO SEGMENTO )";
             "CALCULAR  Ke  (COEFICIENTES À ESQUERDA DO SEGMENTO)";
        FIM-PARA;
        "CALCULAR AS DISTÂNCIAS d[i][j]  (DIR SEG i PARA ESQ SEG j)";
        "ACHAR A PERMUTAÇÃO DE MENOR DISTÂNCIA (USANDO A FUNÇÃO VISIT ( ) )";
        "GUARDAR OS COEFICIENTES DO FINAL DO BLOCO";
        "REORDENAR OS  SEGMENTOS  E ESCREVER NO ARQUIVO DE SAÍDA";
    FIM-ENQUANTO;
   "FECHAR ARQUIVOS";
FIM  {CSITOVOZ}.
```

Figura 4.4 - Algoritmo de reordenação dos segmentos (BLOCO com NR_SEGS = 8 segmentos).

A seguir, são testadas as 8! (40320) permutações possíveis (busca exaustiva) entre segmentos do mesmo bloco. É usado para isto a função recursiva visit( ). Esta função é sugerida por Sedgewick [Sedgewick, 1946], sendo mostrada uma adaptação da mesma escrita em linguagem C na Figura 4.5. Para cada permutação, acha-se uma distância total que corresponde a soma das distâncias das transições de segmentos desta permutação. Nesta distância total é levada em consideração a transição entre o último segmento do bloco anterior e o primeiro segmento do bloco atual. Dentre as 8! permutações, é escolhida aquela que apresenta a menor distância total.

```
/*---------------------------------------- FUNÇÃO  VISIT ( ) ----------------------------------------*

        Esta função gera todas as nr_segs! permutações possíveis. Trata-se de uma função recursiva que foi
adaptada para calcular as distâncias totais de cada permutação e escolher a de menor valor (chave). Sua
chamada é da seguinte maneira:
            int chave[p + 1];                    ( p é a ordem do modelo )
            int nr_segs = 8;                     ( número de segmentos)
            int id, val[9];                  (variáveis usadas pela função)
            float distmin;        (menor valor de distância encontrado)
            float  d[9][9];                  (matriz contendo as distâncias)
                (as variáveis acima deverão ser definidas globalmente)
              int i;
            id = -1;
            for (i=0;i<=nr_segs;i++)
            {  val[i] = 0;
            }
            distmin = MAXFLOAT;
            visit(0);
    *----------------------------------------------------------------------------------------------------*/
void visit(int k)
{  int i,t;
  float distancia;
  int *origem, *destino;
  id = id + 1;
  val[k] = id;
  if (id == nr_segs)
  {  /* achando a distância para uma dada permutação val[.] */
    distancia = 0;
    for (i=1;i<nr_segs;i++)
    {  distancia = distancia + d[val[i]][val[i+1]];
    }
    /* distância entre o último segmento do bloco anterior e o primeiro do bloco atual */

    distancia = distancia + d[0][val[1]];
    if (distancia < distmin)
    {  origem  = val;
       destino = chave;
       distmin = distancia;
       memcpy(destino,origem,(nr_segs+1)*sizeof(int));
    }
  }
  for (t=1;t<=nr_segs;t++)
  {  if (val[t]==0) visit(t);
  }
  id = id - 1;   val[k] = 0;
}
      /*-------------------------------------- FIM DA FUNÇÃO VISIT ( ) --------------------------------------*/
```

Figura 4.5 - Função que gera as permutações de segmentos possíveis (nr_segs!).

A permutação escolhida é guardada num vetor que será a ***chave*** para a reordenação dos segmentos. Uma vez efetuada a reordenação, é escrito o bloco no arquivo de saída e o processamento continua com a leitura do próximo bloco.

4.3 MELHORANDO O SINAL

Em casos reais, quando o sinal criptofonado passa por um canal, ocorre uma superposição de segmentos, ou seja, o final de um segmento avança, por efeito do canal, para o início do segmento seguinte. O resultado é o aparecimento de um ruído indesejável que degrada a inteligibilidade do sinal quando recuperado. Para minimizar este efeito, são normalmente zeradas algumas amostras do final de cada segmento (cerca de 1 a 2 ms) antes da transmissão [manuais dos equipamentos Cryptophon 1100 da Brown, Boveri & Company, Ltd. e TST 7595 da Tele Security Timmann].

Apesar disto, ainda notamos esta superposição de segmentos. Este efeito é percebido diretamente na forma de onda como alguns picos nas transições dos segmentos ou ouvindo o sinal ("zumbido" de fundo). Para melhorar o resultado pode-se, então, diminuir a amplitude das primeiras amostras de cada segmento. Isto irá permitir uma melhoria na qualidade do sinal criptoanalisado pois, como foi observado com experimentações, o "zumbido" causado pelos possíveis vales que ocorrem quando o sinal tem amplitude alta prejudica menos a inteligibilidade do que o "zumbido" causado pelos picos de superposição.

4.4 SUMÁRIO

Este capítulo apresentou o esquema de criptoanálise de CSI-T proposto. Foi visto inicialmente como preparar o sinal cifrado para a criptoanálise. Em seguida, foi apresentado

o algoritmo de reordenação de segmentos baseado na mínima distância espectral entre segmentos adjacentes de um mesmo bloco. Por fim, foi comentado como minimizar na recepção o efeito da superposição de segmentos melhorando, desta forma, a inteligibilidade do sinal criptoanalisado.

No capítulo seguinte, veremos a apresentação dos resultados das simulações e testes feitos a partir deste esquema. Vários sinais de voz serão cifrados e criptoanalisados supondo-se transmissão num canal ideal e num canal telefônico. Serão usadas medidas objetivas de desempenho para comparar o sinal criptoanalisado com o sinal original (em cada medida de distância espectral usada na criptoanálise).

5. RESULTADOS OBTIDOS

Serão apresentados neste capítulo os resultados experimentais obtidos pelo esquema de criptoanálise de CSI-T ("jumping window") proposto. Alguns sinais de voz foram cifrados por este método e a criptoanálise foi feita usando-se as três distâncias espectrais discutidas nos capítulos anteriores.

Os resultados serão apresentados por meio de medidas objetivas de desempenho que procuram quantificar a proximidade dos espetros dos sinais criptoanalisado e original, apesar de poder-se conjecturar que o objetivo da criptoanálise pretendida é a obtenção do conteúdo da mensagem e não a recuperação perfeita do sinal de voz que foi cifrado. Neste particular, foi feita uma medida subjetiva para o caso de canal ideal visando a constatação de que os sinais criptoanalisados ficaram ***inteligíveis*** (alguém que ouviu é capaz de escrever o conteúdo da mensagem) em quase a sua totalidade para as pessoas consultadas.

Na Seção 5.1. serão vistos os sinais de testes usados, suas origens e como foram cifrados. Na Seção 5.2. serão apresentados os resultados obtidos quando passamos os sinais de teste por um canal ideal. Os resultados serão apresentados em termos de relação sinal-ruído (SNR), distorção espectral, taxa de acertos (objetiva) e, somente neste caso de canal ideal, em termos de uma taxa de acertos subjetiva. A Seção 5.3. mostra os resultados dos sinais de testes criptoanalisados após passarem por um canal telefônico simulado. Finalmente, na Seção 5.4., será feita uma análise dos resultados obtidos.

5.1. SIMULAÇÃO DO SINAL CIFRADO

Os sinais de voz usados nos testes passaram por um simulador de CSI-T "jumping window" (CSIT.EXE) onde o sinal de entrada é dividido em blocos de 8 segmentos de 256 amostras e os segmentos são "embaralhados" por meio de uma permutação uniforme, conforme visto no Capítulo 2.

Foram utilizados os seguintes arquivos de voz todos digitalizados com uma freqüência de amostragem de 8 KHz a 12 bits por amostra:

- SCHOOL.VOZ : voz masculina, idioma inglês, gravada em fita K-7;
- SINTO.VOZ : voz feminina, idioma português e gravada de um aparelho de televisão;
- TELEF2.VOZ : voz masculina, idioma português e proveniente de um canal telefônico;
- VEGA2.VOZ : voz feminina, idioma inglês e gravada em fita K-7. Apresenta uma certa constância na sua intensidade (amplitude quase constante).

No caso da avaliação subjetiva do canal ideal, foram usadas dez frases (sinais) transcritas abaixo para eventuais correlações de fonemas com desempenho da criptoanálise:

1) VENDO UMA TELEVISÃO COLORIDA PELO MELHOR PREÇO DA PRAÇA.
2) CUMPRINDO DETERMINAÇÕES DO MINISTÉRIO DA SAÚDE A SECRETARIA DE SAÚDE DO RIO GRANDE DO SUL COMEÇOU A RECOLHER HOJE PELA MANHÃ TODOS OS MEDICAMENTOS ELABORADOS À BASE DE CONFREI.
3) EU ME SINTO ASSIM NA OBRIGAÇÃO DE BANCAR SUA ADVOGADA.
4) O RIO GRANDE DO SUL VEM AUMENTANDO O CONSUMO DE ENERGIA ELÉTRICA ENTRE 5 E 6 % AO ANO.
5) ENGENHEIROS ELETRÔNICOS ALEMÃES ACABARAM POR DESCOBRIR UMA NOVA MANEIRA DE TRANSMITIR PROGRAMAS DE RÁDIO BASTANTE SUPERIOR ÀS ONDAS MÉDIAS.
6) ELA VAI GANHAR UM KIT RAINHA DO LAR SÍLVIO.
7) MUITAS VEZES AS PRÓPRIAS MÃES REAGEM A ESTAS AGRESSÕES TAMBÉM COM MUITA VIOLÊNCIA.
8) SERÁ EFETUADO UM CHECK-UP METABÓLICO DOS PARTICIPANTES POSSIBILITANDO A PRESCRIÇÃO DE UMA MEDICAÇÃO PERSONALIZADA.
9) BRASILEIRAS E BRASILEIROS BOM DIA AQUI VOS FALA O PRESIDENTE JOSÉ SARNEY EM MAIS UMA CONVERSA AO PÉ DO RÁDIO.
10) CHANCE DE SEGUNDO TURNO COM LULA E BRIZOLA AGITA O BLACK E OURO.

5.2. CANAL IDEAL

Começamos esta seção definindo uma das medidas objetivas de desempenho que será usada. A relação sinal-ruído (SNR) é uma medida objetiva bastante utilizada para avaliar a qualidade de voz recuperada após um certo processamento. Ela corresponde ao logaritmo da razão entre a variância do sinal original e a variância do sinal reconstruído[8] [Jayant e Noll, 1984], sendo dada por:

$$\mathrm{SNR(dB)} = 10\log(\sigma_x^2 / \sigma_r^2) \tag{5.1}$$

onde o índice **x** refere-se ao sinal original e **r** ao erro (reconstruido menos original). O logaritmo acima é de base 10.

Uma outra medida objetiva, que é na verdade um refinamento da SNR, é a relação sinal-ruído segmentar (SNRSEG), muito utilizada na avaliação de codificadores de voz e definida por [Jayant e Noll, 1984]:

$$\mathrm{SNRSEG(dB)} = \boldsymbol{E}\left[\mathrm{SNR(m)(dB)}\right] \tag{5.2}$$

SNR em (5.2) é a relação sinal-ruído convencional para o segmento m, e o valor esperado é, na prática, uma média sobre todos os segmentos de interesse. Neste trabalho, usaremos somente a relação sinal-ruído uma vez que no caso da SNRSEG não haveria ruído quando um segmento estivesse em seu local correto.

Entende-se por canal ideal aquele com resposta plana e fase nula em todo o espectro ou, em outras palavras, o sinal foi criptoanalisado logo após ser cifrado. Os resultados, em termos de relação sinal-ruído SNR para cada sinal, em função de cada medida de distância espectral, são apresentados na Tabela 5.1.

---

[8] No nosso caso, sinal criptoanalisado.

	SCHOOL.VOZ	SINTO.VOZ	TELEF2.VOZ	VEGA2.VOZ
EUCLIDEANA	3,31 dB	0,72 dB	1,71 dB	9,97 dB
ITAKURA	1,80 dB	0,84 dB	0,78 dB	1,55 dB
CEPSTRAL	2,22 dB	2,09 dB	3,41 dB	16,59 dB

Tabela 5.1 - SNR de sinais criptoanalisados após transmissão por canal ideal, em função de cada distância espectral usada.

Observa-se na Tabela 5.1 baixos valores de SNR, se comparados com o desempenho de codificadores dos mais simples [Jayant e Noll, 1984]. Entretanto, a medida realizada é apenas mais uma indicação do quanto uma distância é melhor que a outra no esquema de criptoanálise proposto. Toma-se como exemplo a reordenação de um bloco onde houve apenas um segmento fora da ordem: o último segmento passou a ser o primeiro e todos os outros ficaram, pois, deslocados para a direita de um segmento. Este bloco, dependendo das amplitudes das amostras de cada segmento, poderia apresentar um elevadíssimo componente de erro para o cálculo da SNR embora num teste subjetivo o erro seria provavelmente pouco percebido.

A seguir, será apresentada na Tabela 5.2 uma medida de distorção espectral (foi chamada de distorção para não confundir com as medidas de distância espectral usadas na criptoanálise) para os sinais cifrados em função das medidas de distância usadas. Esta medida procura avaliar o quanto o espectro do sinal criptoanalizado assemelha-se ao do sinal original; trata-se da distância Euclideana média não ponderada dos coeficientes $\{a_i\}$ (estimados **segmento** a **segmento** pelo algoritmo de Levinson-Durbin) dos sinais original e criptoanalisado em relação aos coeficientes do sinal original e é dada por:

$$\text{distorção espectral relativa} \quad = \quad \frac{(\mathbf{a}_1 - \mathbf{a}_2)'.(\mathbf{a}_1 - \mathbf{a}_2)}{\mathbf{a}_1'.\mathbf{a}_1} \text{x} 100\% \qquad (5.3)$$

onde $\mathbf{a}_1$ é o vetor contendo os coeficientes $\{a_i\}$ do sinal original e $\mathbf{a}_2$ é o vetor contendo os coeficientes do sinal criptoanalisado.

	SCHOOL.VOZ	SINTO.VOZ	TELEF2.VOZ	VEGA2.VOZ
EUCLIDEANA	29,4 %	36,2 %	29,1 %	7,2 %
ITAKURA	33,8 %	32,5 %	26,1 %	18,6 %
CEPSTRAL	23,3 %	24,1 %	21,3 %	5,3 %

Tabela 5.2 - Distorção espectral relativa de sinais criptoanalisados após transmissão por canal ideal em função de cada distância espectral usada.

Uma outra medida objetiva de desempenho, a ***taxa de acertos*** (número de segmentos recolocados em seus locais de maneira acertada pelo número total de segmentos do sinal), será mostrada na Tabela 5.3. Esta taxa, embora simples e de cálculo imediato, mostrou-se bastante significativa, pois além de ser intuitiva apresentou resultados que bem poderiam representar uma possível medida subjetiva. Ela é dada por:

$$\text{taxa de acertos} = \frac{\text{número de segmentos em seus locais corretos}}{\text{número total de segmentos}} \text{x} 100\% \qquad (5.4)$$

	SCHOOL.VOZ	SINTO.VOZ	TELEF2.VOZ	VEGA2.VOZ
EUCLIDEANA	70,3 %	51,0 %	52,5 %	89,7 %
ITAKURA	61,7 %	58,3 %	51,2 %	66,2 %
CEPSTRAL	72,7 %	65,6 %	63,1 %	91,2 %

Tabela 5.3 - Taxa de acertos de sinais criptoanalisados após serem transmitidos por canal ideal em função de cada distância espectral usada.

A seguir, a Tabela 5.4 apresenta de forma resumida o tempo de processamento do programa de criptoanálise (CSITOVOZ.EXE). Os tempos apresentados foram os tempos médios calculados durante a reordenação dos arquivos de teste num microcomputador com CPU 80486 DX2 a 66 MHz.

Observa-se que, como o tempo de processamento com as distâncias Euclideana e Cepstral é apenas sete vezes o tempo do sinal, seria factível implementar-se a criptoanálise em ***tempo real*** com o uso de processador de sinais do tipo TMS 320C50 (20 MIPS). Entende-se por tempo real a capacidade da implementação de reproduzir um bloco criptoanalisado na mesma velocidade que o sinal cifrado chega, logo após o preenchimento

de uma memória de bloco (8 x tamanho do segmento = 2.048 amostras). Isto só seria possível se conseguíssemos determinar o sincronismo (início de um bloco) a priori.

Medida de Distância Espectral Usada no Algoritmo de Reordenação	Tempo de Processamento (segundo de processamento / segundo de sinal)
Euclideana	7,15
Itakura	54,05
Cepstral	7,10

Tabela 5.4 - Tempo de Processamento do programa de Criptoanálise em segundos de processamento por segundo de sinal (CPU 80486 DX2 a 66 MHz).

Por fim, teremos na Tabela 5.5 a apresentação dos resultados de uma avaliação subjetiva realizada da seguinte maneira: dez pessoas ouviram quantas vezes desejaram dez frases criptoanalisadas (via distância Cepstral) e escreveram numa folha de respostas, com uma palavra em cada espaço, todo o texto que conseguiram compreender. A tabela mostra uma taxa de acertos subjetiva que corresponde à porcentagem das palavras que as pessoas conseguiram acertar. Aparece, também, a taxa de acertos objetiva para fins de comparação.

FRASE	NÚMERO DE PALAVRAS	TAXA DE ACERTOS OBJETIVA (% DE SEGMENTOS CERTOS)	TAXA DE ACERTOS SUBJETIVA (% DE PALAVRAS CERTAS)
1	9	56/512 = 50,0 %	85/90 = **94,4** %
2	29	239/376 = 63,6 %	288/290 = **99,3** %
3	10	77/96 = 80,2 %	83/100 = **83,0** %
4	19	101/160 = 63,1 %	189/190 = **99,5** %
5	19	123/288 = 42,7 %	134/190 = **70,5** %
6	9	80/96 = 83,3 %	83/90 = **92,2** %
7	13	126/192 = 65,6 %	129/130 = **99,2** %
8	14	133/240 = 55,4 %	112/140 = **80,0** %
9	20	139/240 = 57,9 %	200/200 = **100,0** %
10	12	121/152 = 79,6 %	117/120 = **97,5** %

Tabela 5.5 - Resultados da avaliação subjetiva do canal ideal.

Da Tabela 5.5, pode-se verificar que a taxa de acertos objetiva geral (ponderada pelo número total de segmentos) é de 61,2 %, enquanto que a taxa de acertos subjetiva geral (ponderada pelo número total de palavras) é de 92,2 %.

5.3. CANAL TELEFÔNICO

Neste caso, o sinal de teste a ser criptoanalisado deve passar por um canal telefônico. O canal telefônico foi simulado por um filtro digital com os coeficientes obtidos do software TLN (Simulador de Canal Telefônico por Filtro Digital) [Chiaratto e Santos, 1991]. O simulador de canal utilizado provoca três tipos de distorção, em graus que variam de "0" a "6"; um exemplo é apresentado a seguir [Montoro, 1990]:

"F = 1: Distorção de fase equivalente à passagem do sinal por um sistema multiplexador de canais de voz por divisão de freqüência (FDM). Num sistema desse tipo, o sinal sofre uma translação para a freqüência portadora e outra translação para sua posição original na faixa de voz, sendo a distorção provocada principalmente pelos filtros separadores de canais;

C = 1: Distorção de fase equivalente à passagem do sinal por 50 Km de linha condicionada com bobinas de 80 mH (pupinização) espaçadas a cada 1.700 metros;

A = 1: Distorção de amplitude equivalente à passagem do sinal por um sistema FDM tal qual descrito para F = 1".

Procurou-se usar um canal telefônico com os valores médios de F, C e A de modo a obeter-se um caso próximo de uma linha real típica. Foi escolhida, então, a linha tipo 333 (F = 3, C = 3 e A = 3) cuja curva de resposta em freqüência é apresentada na Figura 5.1.

O sinal de teste, já cifrado, foi preparado conforme visto no Capítulo 4 (foram zeradas algumas amostras do final de cada segmento) de modo a minimizar o efeito de superposição de segmentos. Este efeito pode ser observado na Figura 5.2 onde temos o sinal em claro, o mesmo criptofonado após passar pelo canal simulado e, por último, "decifrado" (segmentos colocados em seus devidos lugares mas sem a melhoria mencionada no Capítulo 4).

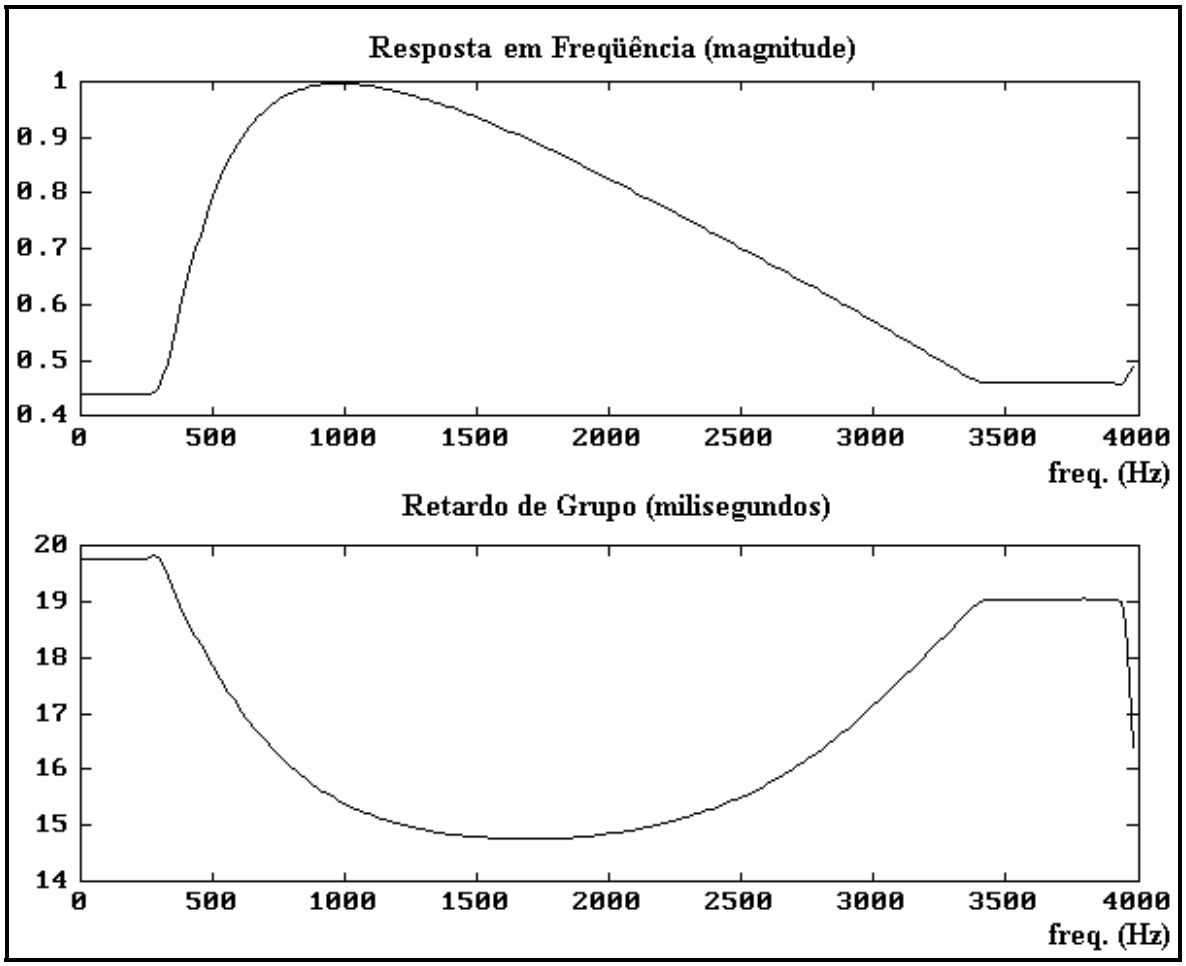

Figura 5.1 - Resposta em freqüência e retardo de grupo do canal telefônico simulado (linha 333) por filtro digital.

A Tabela 5.6 apresenta a taxa de acertos dos sinais criptoanalisados em função das distâncias usadas. A SNR deixou de ser usada pois a preparação do sinal e a passagem pelo canal ocasionaram sinais bastantes diferentes do original. Tanto que, numa primeira tentativa, obteve-se um valor de SNR para o sinal decifrado (recuperado a partir do conhecimento da chave de cada bloco) menor que os obtidos pela criptoanálise (em todas as diferentes distâncias). Uma outra idéia é o cálculo da SNR em relação ao sinal decriptofonado e não em releção ao original. Como os valores obtidos foram muito baixos (relações sinal-ruido negativas), optou-se pela apresentação somente da taxa de acertos.

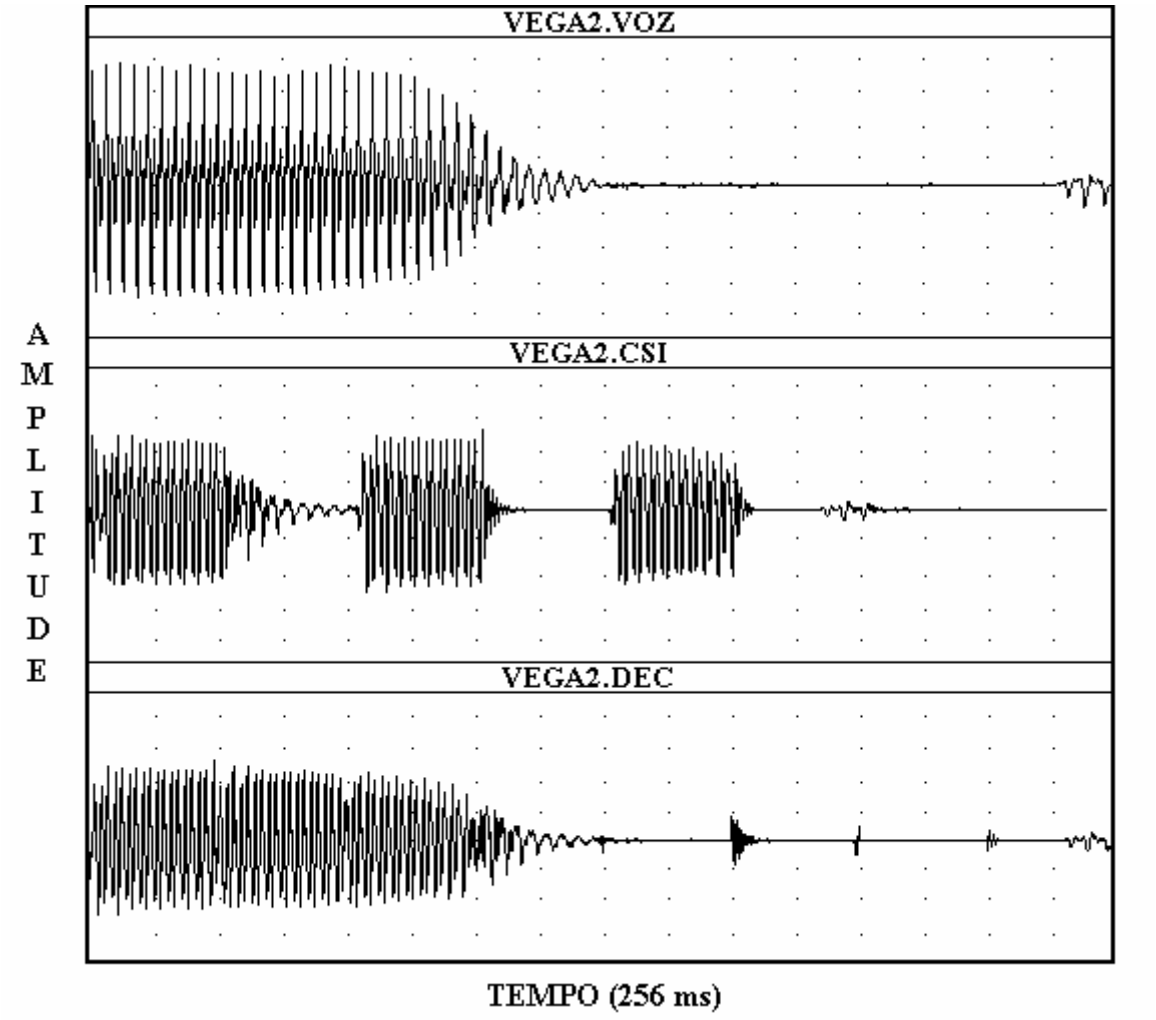

Figura 5.2 - O efeito da superposição de segmentos. Sinal VEGA2 em claro, cifrado após passar pelo canal telefônico e decifrado.

	SCHOOL.CSI	SINTO.CSI	TELEF2.CSI	VEGA.2CSI
EUCLIDEANA	36,7 %	39,6 %	23,1 %	43,4 %
ITAKURA	30,5 %	44,8 %	21,9 %	28,7 %
CEPSTRAL	32,0 %	19,8 %	21,9 %	32,4 %

Tabela 5.6 - Taxa de acertos de sinais criptoanalisados após serem transmitidos por canal telefônico em função de cada distância espectral usada.

Os valores observados na Tabela 5.6 mostram que o desempenho da criptoanálise em sinais que passam por canais telefônicos considerados típicos não foi bom. Para melhorar este desempenho, foi feita uma equalização do canal para elevar a taxa de acerto a um patamar mais próximo do caso do canal ideal. Esta equalização pode ser feita de modo adaptativo, segundo o esquema mostrado na Figura 5.3 [Widrow e Stearns, 1985], onde

procura-se obter o modelo inverso do canal. A diferença entre o sinal de saída da filtragem inversa e o sinal original, ou seqüência de treinamento (antes de passar pelo canal), é usada pelo algoritmo de adaptação para gerar os coeficientes $\mathbf{W}_k$ do filtro inverso. Após um certo tempo o filtro converge e tais coeficientes podem ser usados para equalizar o canal.

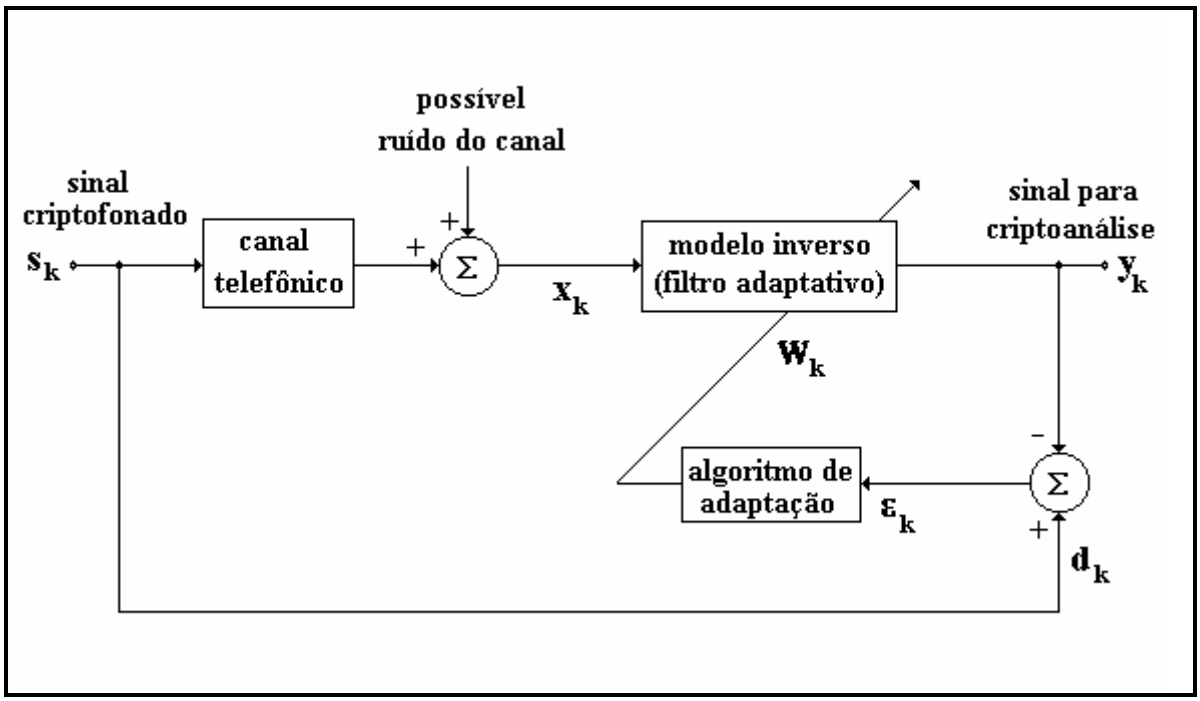

Figura 5.3 - Equalização do canal telefônico usando modelagem inversa com filtragem adaptativa

A Figura 5.4 apresenta as curvas de resposta em freqüência e retardo de grupo do filtro inverso usado na equalização da linha 333. Observa-se que a multiplicação da resposta em freqüência da linha 333 pela do filtro inverso é próximo da unidade e que a soma dos retardos de grupo dos mesmos é aproximadamente constante ao longo do espectro.

Após a equalização do canal, foram feitas novas simulações cujos resultados serão apresentados a seguir. A Tabela 5.7 mostra o desempenho da criptoanálise de sinais de voz após a equalização do canal em termos de distorção espectral relativa. Nesta tabela, a distorção espectral foi calculada tendo como referência os coeficientes LPC do sinal decriptofonado. Tomando-se como referência os coeficientes LPC do sinal de voz original, a Tabela 5.8 mostra novos resultados e, na sua última linha, as distorções do sinal decifrado.

	SCHOOL.VOZ	SINTO.VOZ	TELEF2.VOZ	VEGA2.VOZ
EUCLIDEANA	44,3 %	35,3 %	34,9 %	18,4 %
ITAKURA	48,8 %	39,6 %	41,4 %	43,8 %
CEPSTRAL	45,1 %	40,7 %	33,7 %	14,2 %

Tabela 5.7 - Distorção espectral relativa de sinais criptoanalisados após transmissão por canal telefônico equalizado em função de cada distância espectral usada (referência sinal decifrado).

A Tabela 5.9 apresenta a taxa de acertos dos sinais criptoanalisados após a passagem por um canal telefônico que foi equalizado. Os valores desta tabela evidenciam a melhoria resultante da equalização do canal.

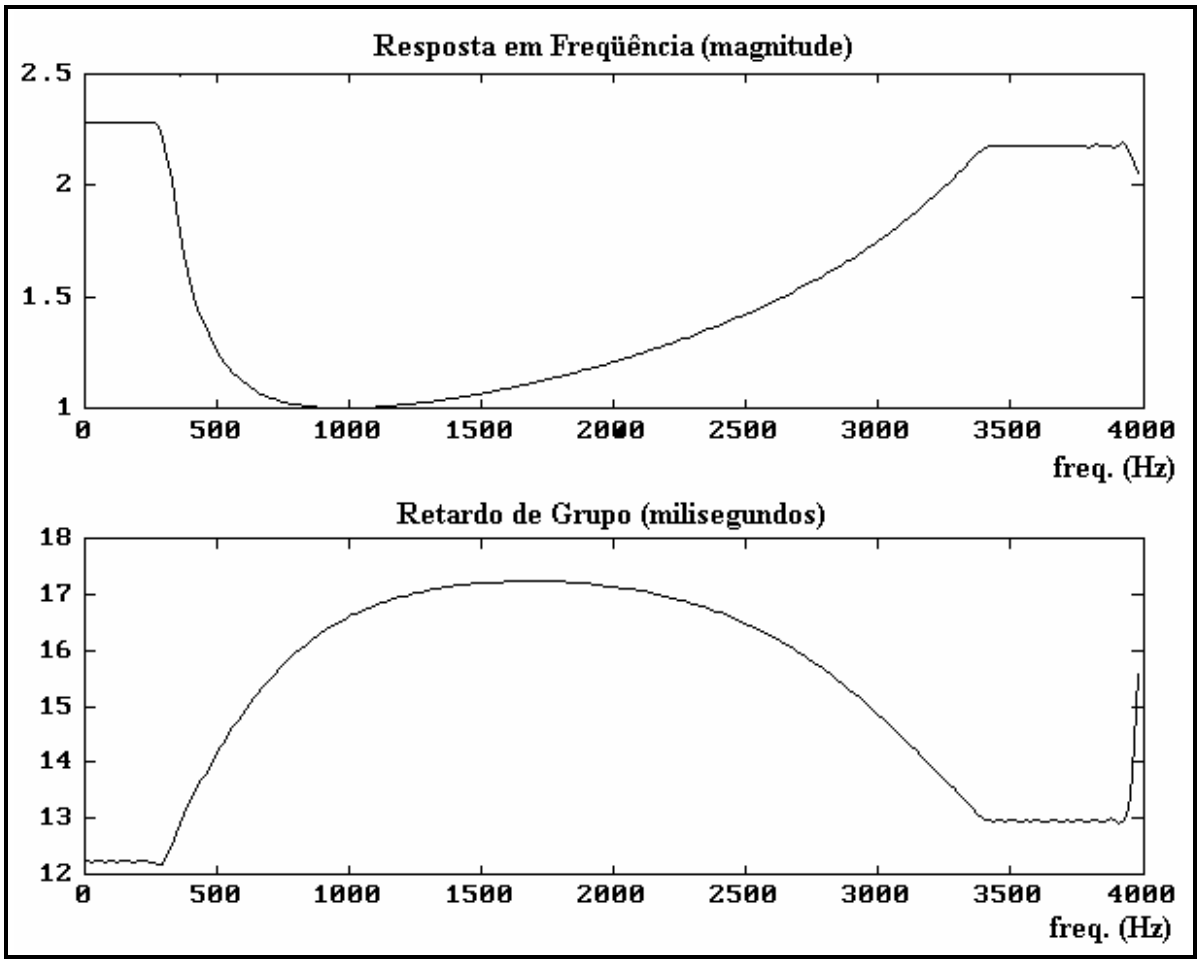

Figura 5.4 - Resposta em freqüência e retardo de grupo do filtro inverso (equalização da linha 333).

	SCHOOL.VOZ	SINTO.VOZ	TELEF2.VOZ	VEGA2.VOZ
EUCLIDEANA	50,3 %	38,0 %	39,0 %	27,1 %
ITAKURA	53,4 %	43,6 %	43,1 %	48,6 %
CEPSTRAL	51,6 %	44,0 %	37,8 %	24,2 %
DECIFRADO	11,8 %	7,14 %	8,0 %	17,2 %

Tabela 5.8 - Distorção espectral relativa de sinais criptoanalisados e decifrados após transmissão por canal telefônico equalizado em função de cada distância espectral usada e do próprio sinal decifrado (referência sinal original).

	SCHOOL.CSI	SINTO.CSI	TELEF2.CSI	VEGA.2CSI
EUCLIDEANA	50,0 %	49,0 %	40,0 %	67,6 %
ITAKURA	39,8 %	54,2 %	23,8 %	40,4 %
CEPSTRAL	45,3 %	46,9 %	41,9 %	75,7 %

Tabela 5.9 - Taxa de acertos de sinais criptoanalisados após serem transmitidos por canal telefônico equalizado em função de cada distância espectral usada.

Pode-se mencionar que o processamento usado na equalização do canal é mínimo dada a melhoria do desempenho. Além disso, cabe ressaltar que a mesma abordagem usada na equalização de canais telefônicos pode, a princípio, ser usada em canais rádio.

5.4. ANÁLISE DOS RESULTADOS

Dos resultados objetivos obtidos e mostrados nas Tabelas da Seção 5.2., verifica-se que a medida espectral que apresentou a melhor performance foi a cepstral. Os resultados foram quase unânimes, sendo que a única contradição, Tabela 5.1 (SNR do sinal SCHOOL.VOZ criptoanalisado), deveu-se provavelmente a termos segmentos reagrupados erroneamente

que baixaram a relação sinal-ruido do sinal criptoanalisado pela distância cepstral em relação ao sinal criptoanalisado pela distância Euclideana: segmentos de altas amplitudes colocados em locais que apresentavam originariamente amplitudes baixas. Além disto, a relação sinal-ruído mostrou-se inadequada para medir objetivamente o desempenho da criptoanálise.

Percebeu-se, também, que a distância de Itakura, embora seja tradicionalmente considerada uma boa medida de distância espectral [Brown e Rabiner, 1982], não é muito adequada ao esquema proposto, uma vez que a simples distância Euclideana mostrou-se melhor na maioria dos testes. Esta inadequação deve-se ao fato de termos que rodar filtros em estrutura reticulada (SQNLSL) com coeficientes constantes num certo intervalo (foi usado um intervalo correspondente a um pouco mais de um terço de segmento) e tomarmos os resíduos no final para computarmos a distância. Estes resíduos terão embutidos uma grande quantidade de informação não relevantes àquilo que necessitamos, uma vez que os coeficientes foram estimados de maneira a representar o espectro do sinal na extremidade do segmento e não ao longo do mesmo.

Dos três critérios objetivos apresentados em 5.2. (SNR, distorção espectral e taxa de acertos), é possível que a distorção espectral seja o mais correlacionado com a qualidade subjetiva. Entretanto, a taxa de acertos (Tabela 5.3) apresentou resultados qualitativamente iguais (apontou os mesmos melhores desempenhos apesar de diferentes quantificações) e com a vantagem de podermos comparar melhor os resultados de diferentes sinais. Logo, a taxa de acertos parece ser o indicador objetivo mais adequado para a avaliação do desempenho da criptoanálise.

Os resultados evidenciaram, também, a dependência do desempenho do esquema de criptoanálise proposto no tipo de sinal: VEGA2.VOZ destacou-se por ter um ritmo constante (ausência de transições abruptas do espectro do sinal original) e apresentar pequenas variações de amplitude.

A avaliação subjetiva mostrada na Tabela 5.5 mostrou que mais que 90% da informação é recuperada pela criptoanálise a partir de pessoas não treinadas para executar a tarefa do criptoanalista. É possível afirmar-se que consegue-se facilmente chegar aos 100%

(para uma mesma taxa de acertos objetiva) se tomarmos mais de uma pessoa e treiná-las antes para ouvir sinais criptoanalisados.

Em 5.3. foram apresentados os resultados da criptoanálise em sinais que passaram por um simulador de canal telefônico. A Tabela 5.6 mostra, como já era esperado, resultados bem inferiores em relação ao canal ideal. Observa-se que a distância cepstral não apresenta, neste caso de canal não ideal, o melhor desempenho; ao contrário disso, a distância que apresenta melhor rendimento fica, aparentemente, bastante dependente do sinal, o que sugere um ajuste fino de todos os parâmetros do algoritmo a cada sinal. Além do ajuste de parâmetros, torna-se fundamental a preparação do sinal. Nos casos práticos, melhores resultados são alcançados com tentativas e erros e, portanto, com a experiência do criptoanalista.

Da análise dos resultados para o caso do canal ideal, conclui-se que o desempenho médio do esquema de criptoanálise proposto pode ser considerado muito bom. Somando-se a este desempenho o reduzido tempo de processamento (conforme Tabela 5.4) para as distâncias Euclideana e cepstral, pode-se afirmar que seu uso é compensador e caso consiga-se, a partir de um possível sinal de sincronismo, determinar o início dos blocos, tal esquema pode ser realizável em tempo real (com hardware específico) a menos do tempo correspondente a um bloco (tal como ocorre num "scrambler" real).

No caso do sinal passar por um canal que apresente distorções de fase não linear, como é o caso do canal telefônico usado na simulação, verificou-se a necessidade de um pré-processamento do sinal a ser criptoanalisado (equalização do canal). Os resultados obtidos após a equalização do canal (Tabelas 5.7 e 5.8) demonstram esta necessidade. Em casos reais, a experiência do criptoanalista em conseguir uma boa filtragem inversa é, mais uma vez, fundamental pois em muitos casos não poderíamos contar com uma seqüência de treinamento.

5.5. SUMÁRIO

Neste capítulo foi mostrado o desempenho do esquema de criptoanálise proposto. Foram feitas tabelas de medidas de desempenho objetivas para os casos de transmissão por um canal ideal e por um canal telefônico típico.

Os sinais de testes usados foram de vários tipos e todos passaram por um simulador de sistema criptofônico CSI-T "jumping window" de 8 segmentos por bloco e 256 amostras por segmento.

Os resultados apresentados nas Seções 5.2. e 5.3. foram analisados na Seção 5.4., onde observou-se que a medida de distância espectral que apresentou o melhor desempenho médio foi a cepstral. Observou-se, também, que no caso de um canal com distorção de fase (fase não linear), uma equalização é necessária. Foi visto que esta equalização pode ser feita com uma filtragem inversa adaptativa a partir do conhecimento do sinal antes do canal telefônico ou de uma seqüência de treinamento (mensagem conhecida). Os resultados após a equalização aproximaram-se mais dos bons resultados do canal ideal.

6. CONCLUSÃO

O objetivo deste trabalho foi a apresentação de um esquema de criptoanálise de uma técnica de criptofonia conhecida como permutação de segmentos temporais "jumping window" ou, como chamado aqui, CSI-T bloco a bloco com segmento de tamanho fixo.

Uma apresentação das diferentes técnicas de criptoanálise foi feita no Capítulo 2, onde foi definido qual é exatamente o sinal que será criptoanalisado e mostrado uma justificativa deste esforço em função da quantidade relativa de equipamentos de sigilo de voz que são produzidos mundialmente usando esta técnica.

O Capítulo 3 tratou das medidas de distância espectral escolhidas para comparar final e início de possíveis segmentos adjacentes quando da reordenação, bem como mostrou como foram estimados os coeficientes LPC usados para o cálculo de tais medidas. As distâncias utilizadas foram a Euclideana, a de Itakura modificada e a cepstral. Os coeficientes usados (de reflexão) foram estimados pelo algoritmo SQNLSL ("Square-Root-Normalized Least-Square Lattice").

O esquema proposto, dividido em preparação do sinal, reordenação dos segmentos e melhoria do sinal, foi apresentado no Capítulo 4. Neste capítulo foi assumido que o canal por onde trafegou o sinal não introduziu sérias distorções de fase que justificassem um pré-processamento mais elaborado. Foi, também, assumido que a parcela do sincronismo já foi retirada do sinal a ser criptoanalisado; isto é normalmente feito pela simples edição do sinal no caso de sincronismo inicial ou filtragem ("notch") quando o sincronismo contínuo é percebido como um tom piloto ocupando uma pequena faixa da banda de áudio.

Dos resultados apresentados no Capítulo 5, pode-se concluir que o esquema é capaz de apresentar um sinal criptoanalisado com distorções aceitáveis (que não comprometem a inteligibiidade) em relação ao sinal decifrado. Do ponto de vista prático, o esquema foi eficaz no sentido de proporcionar a inteligibilidade da mensagem cifrada (criptograma) conforme visto na avaliação subjetiva do canal ideal.

Os resultados da Seção 5.2. (canal ideal) evidenciam uma superioridade da distância cepstral em relação às demais. Por outro lado, a distância de Itakura, embora tradicionalmente considerada uma boa medida de distância espectral, não é muito adequada ao esquema proposto uma vez que a simples distância Euclideana apresentou um melhor desempenho na maioria dos testes realizados.

Quando o sinal passa por um canal telefônico com especificações mais severas no que diz respeito às distorções que foram simuladas (distorção de fase simétrica, distorção de fase assimétrica e distorção de amplitude), verificou-se na Seção 5.3. a necessidade da equalização do canal anterior à criptoanálise. Pelas experimentações realizadas, pode-se concluir, também, que, após uma boa equalização, os resultados, tanto em termos das medidas de desempenho objetivas feitas quanto das subjetivas não formalizadas, aproximam-se dos níveis obtidos no caso do canal ideal e que os pequenos intervalos feitos nulos nas transições de segmentos representam, agora, uma queda na qualidade do sinal criptoanalisado. Portanto, o que foi mencionado como preparação do sinal no Capítulo 4 não mais se aplica.

Foi verificado nas experiências realizadas que o desempenho da criptoanálise é sensível ao sinal; este fato obriga o criptoanalista a ajustar alguns parâmetros do algoritmo de criptoanálise para obter melhores resultados para cada sinal. Estes parâmetros são: fator de ponderação exponencial (em torno de 0,98), coeficiente de pré-ênfase (em torno de 0,8), peso da informação de energia (em torno de 2) e ordem do modelo (geralmente entre 10 e 20).

O esquema de criptoanálise proposto abre várias alternativas para a continuação do trabalho tanto nas variantes de CSI-T (tamanho de segmentos variáveis e "sliding window") quanto nos processos bidimensionais: CSI-FT, por exemplo. Um argumento simples que comprova a importância destas técnicas afins é o fato deste autor não ter encontrado na cidade de Brasília um equipamento que utilizasse somente CSI-T bloco a bloco com segmentos de tamanho fixo para efetuar testes com sinais cifrados por um "scrambler" real.

Um outro ponto que merece atenção é o fato de não ser mais possível em tempo computacional a busca exaustiva de N! permutações possíveis quando este número N

começa a crescer. Torna-se, pois, interessante um estudo sobre até que valor de N esta busca exaustiva seria factível e que outras alternativas ou estratégias poderiam ser adotadas caso contrário. É importante frisar, mais uma vez, que N = 8 adotado para testes é um valor dentro dos mais utilizados no mundo real tendo em vista as razões apontadas no capítulo 2. O maior valor de N visto em equipamentos reais, 32, não é o "default" do equipamento (trata-se de um opcional que deverá ser adquirido à parte), só é sugerido quando pode-se dispor de excelentes canais de comunicações e, ainda assim, está sujeito à uma maior degradação da voz recuperada no destino.

Finalmente, resta reafirmar a necessidade fundamental de poder-se contar com a experiência, "feeling" e perseverança do criptoanalista para que uma criptoanálise obtenha sucesso.

REFERÊNCIAS BIBLIOGRÁFICAS

Brown, M. K. **e Rabiner**, L. R., On the Use of Energy in LPC-Based Recognition of Isolated Words, in "The Bell System Technical Journal", vol. 61, n°. 10, dezembro **1982**, pp 2971-2987.

Brown, Bovery & Company, Ltda., Manual do Cryptophon 1100, Suiça.

Cabral Jr., Euvaldo F., Uma Incursão pelos Domínios da Criptofonia, in "Revista Militar de Ciência e Tecnologia", vol IV, n°. 2, abril/junho **1987**, pp 26-52.

Chiaratto, André R. L. **e Santos**, Ângela M. dos, Simulação de Canais Telefônicos Através de Filtros Digitais, Relatório de Estágio Supervisionado, Universidade de Brasília, Brasília, julho **1991**.

Cowan, C.F.N. **e Grant**, P.M., Adaptive Filters, Englewood Cliffs, Prentice-Hall, **1985**.

De Souza, Peter **e Thomson**, Peter J., LPC Distance Measures and Statistical Tests with Particular Reference to the Likelihood Ratio, in "IEEE Transactions on Acoustic, Speech, and Signal Processing", vol. ASSP-30, n°. 2, abril **1982**, pp 304-315.

Florêncio, Dinei A. F., On the Use of Asymmetric Windows for Reducing the Time Delay in Real-Time Spectral Analysis, in "Proc. ICASSP'91", **1991**, pp 3261-3264.

Gray Jr., Augustine H. **e Markel**, John D., Distance Measures for Speech Processing, in "IEEE Transactions on Acoustic, Speech, and Signal Processing", vol. ASSP-24, n°. 5, outubro **1976**, pp 380-391.

Gray, Robert M., **Buzo**, Andrés, **Gray** Jr., Augustine H. **e Matsuyama**, Yasuo, Distortion Measures for Speech Processing, in "IEEE Transactions on Acoustic, Speech, and Signal Processing", vol. ASSP-28, n°. 4, agosto **1980**, pp 367-376.

Itakura, Fumitada, Minimum Prediction Residual Principle Applied to Speech Recognition, in "IEEE Transactions on Acoustic, Speech, and Signal Processing", vol. ASSP-23, n°. 1, fevereiro **1975**, pp 67-72.

Jane's Military Communications, **1991-92**.

Jayant, N.S., Effective Number of Keys in a Voice Privacy System Based on Permutation Scrambling, in "AT&T Techical Journal", vol. 66, n°.1, janeiro/fevereiro **1987**, pp 132-136.

Jayant, N.S., **McDermott**, B. J., **Christensen**, S. W. **e Quinn**, A. M. S., A Comparison of Four Methods for Analog Speech Privacy, in "IEEE Transactions on Communications", vol. COM-29, n°. 1, janeiro **1981**, pp 18-23.

Jayant, N.S. **e Noll**, Peter, Digital Coding of Waveforms - Principles and Applications to Speech and Video, Englewwod Cliffs, Prentice-Hall, Inc., **1984**.

Markel, J.D. **e Gray** Jr., A.H., Linear Prediction of Speech, Berlin, Springer-Verlag, **1976**.

Montoro, Fábio de Azevedo, Transmissão de Dados e Modem, São Paulo, Érica, **1990**.

Oppenheim, Alan V. **e Schafer**, Ronald W., Discrete-Time Signal Processing, Englewood Cliffs, Prentice-Hall, Inc., **1989**.

Rabiner, Lawrence R. **e Schafer**, Ronald W., Digital Processing of Speech Signals, Englewood Cliffs, Prentice-Hall, Inc., **1978**.

Sedgewick, Robert, Algorithms, USA, Addison-Wesley Publishing Company, **1946** (2a. edição em 1988).

Speech and Facsimile Scrambling and Decoding, Laguna Hills, Aegean Park Press, 1981 (original de **1946**).

Sridharan, S., **Dawson**, E. e **Goldburg**, B., Fast Fourier Transform Based Speech Encryption System, in "IEE Proceedings-I", vol. 138, n°. 3, junho **1991**, pp 215-223.

Tele Security Timmann, Manual do TST 7595, Alemanha.

Timmann, Klaus P., Secure Voice: Advanced ComSec systems for digital voice coding and high speed modems, in "Miltronics", vol. 7, n°. 2, abril/maio **1986**.

Tribolet, José M., **Rabiner**, Lawrence R. e **Sondhi**, Man Mohan, Statistical Properties of an LPC Distance Measure, in "IEEE Transactions on Acoustic, Speech, and Signal Processing, vol. ASSP-27, n°. 5, outubro **1979**, pp 550-558.

Van Trees, Harry L., Detection, Estimation, and Modulation Theory, USA, John Wiley and Sons, Inc., **1968**.

Widrow, Bernard **e Stearns**, Samuel D., Adaptive Signal Processing, Englewood Cliffs, Prantice-Hall, Inc., **1985**.