

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**COMPRESSÃO DE SINAIS ELETROMIOGRÁFICOS
USANDO CODIFICADOR MULTITAXA PREDITIVO
ADAPTATIVO**

JOSÉ HAROLDO DE ASSIS CAVALCANTE

ORIENTADOR: LÚCIO MARTINS DA SILVA

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGENE.DM – 369/09

BRASÍLIA/DF: FEVEREIRO – 2009

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**COMPRESSÃO DE SINAIS ELETROMIOGRÁFICOS USANDO
CODIFICADOR MULTITAXA PREDITIVO ADAPTATIVO**

José Haroldo de Assis Cavalcante

DISSERTAÇÃO DE MESTRADO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

LÚCIO MARTINS DA SILVA, DR. ENE/UNB
(ORIENTADOR)

FRANCISCO DE ASSIS DE OLIVEIRA NASCIMENTO, DR. ENE/UNB
(EXAMINADOR INTERNO)

PEDRO DE AZEVEDO BERGER, DR., CIC/UNB
(EXAMINADOR EXTERNO)

BRASÍLIA/DF, 20 de fevereiro de 2009

FICHA CATALOGRÁFICA

CAVALCANTE, J. H. A. (2009).

Compressão de Sinais Eletromiográficos usando Codificador Multitaxa Preditivo Adaptativo [Distrito Federal] 2009.

xvi, 81 p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2009).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

- | | |
|------------------------------------|-------------------------|
| 1. Compressão | 2. Sinais |
| 3. Eletromiografia | 4. Codificação |
| 5. Engenharia Elétrica | 6. Engenharia Biomédica |
| 7. Processamento Digital de Sinais | 8. Biomecânica |
| 9. EMG | 10. AMR |
| I. ENE/FT/UnB | II. Título |

REFERÊNCIA BIBLIOGRÁFICA

CAVALCANTE, J. H. A. (2009). Compressão de Sinais Eletromiográficos usando Codificador Multitaxa Preditivo Adaptativo. Dissertação de Mestrado em Engenharia Elétrica, publicação PPGENE.DM-369/09, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 81 p.

CESSÃO DE DIREITOS

AUTOR: José Haroldo de Assis Cavalcante.

TÍTULO: Compressão de Sinais Eletromiográficos usando Codificador Multitaxa Preditivo Adaptativo.

GRAU: Mestre

ANO: 2009

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

José Haroldo de Assis Cavalcante
SQS 108 – Bl. A – Ap. 106
70347-010 – Brasília – DF – Brasil
Email: joseharoldoac@yahoo.com.br

DEDICATÓRIA

A meus pais, José e Maria Teresinha.

A meus sobrinhos, Maria Luísa, Pedro Ângelo e Mariana.

*O temor de Iahweh é princípio de conhecimento:
os estultos desprezam sabedoria e disciplina.
Pv 1,7.*

AGRADECIMENTOS

Acima de todos, a Deus.

A meu orientador, professor Lúcio Martins da Silva, por todo o apoio, amizade e imensa paciência.

Aos professores, funcionários e colegas do ENE. Em especial, aos professores Adson da Rocha, Francisco Assis, Leonardo RAX e, *in memoriam*, Sérgio Barroso e Lourenço Chehab. Ao Fabiano Soares, pelos sinais e rotinas. E também aos professores Pedro Berger (CIC, pelos sinais), Jake do Carmo (FEF) e Carlos Gonçalves (IB).

A meus pais José e Teresinha, minha irmã Maria Odete, Wenner e os sobrinhos, meus avós (*in memoriam*), Levy e Odette, João e Crescência. Meu pai e meu avô João, em particular, por me inculcaram o gosto pela eletricidade, ondas e coisas assim. E à Letícia e família, por todo o carinho e ajuda. A meus gerentes no Tesouro Nacional, Fábio Lima e Mônica Pacheco. E à Juliana, Marcos Sisnando, Otávio, Bosco, Petra, Francisca (“meu Deus, tem gráfico pela casa toda”), Marinalva, Rosirene, Janaína, o pessoal do Sarah, do Catecumenato e todos os amigos e aqueles que de alguma forma me inspiraram, acreditaram em mim, estimularam, apoiaram, comemoraram, oraram, torceram e/ou me desafiaram durante o período de elaboração desta dissertação ou em algum momento de minha vida e de minha trajetória acadêmica e profissional.

Eu quase que nada não sei. Mas desconfio de muita coisa.

(Guimarães Rosa)

RESUMO

COMPRESSÃO DE SINAIS ELETROMIOGRÁFICOS USANDO CODIFICADOR MULTITAXA PREDITIVO ADAPTATIVO

Autor: José Haroldo de Assis Cavalcante

Orientador: Lúcio Martins da Silva

Programa de Pós-Graduação em Engenharia Elétrica

Brasília, fevereiro de 2009

O desenvolvimento de codificadores de voz altamente eficientes tem despertado o interesse na sua utilização em outros tipos de sinal, tais como os sinais eletrofisiológicos. Este trabalho avalia o codec AMR-ACELP, desenvolvido para a telefonia celular de terceira geração, e propõe alterações que melhoram seu desempenho na compressão de sinais de eletromiografia, considerando a qualidade objetiva do sinal reconstruído. Tais alterações foram utilizadas em diversos novos modos de operação do codec que foram avaliados usando sinais de EMG isométricos. Obtiveram-se taxas de compressão entre 81,4% e 90,3%, e um erro quadrático médio mínimo de 1,81%. Verificou-se que a qualidade objetivamente medida do sinal aumenta com a exclusão dos elementos de enriquecimento da qualidade subjetiva e do dicionário adaptativo e com a inclusão de um segundo dicionário fixo. Os resultados foram compatíveis, em termos de erro e taxa de compressão, com trabalhos anteriores sobre compressão de EMG sob condições similares, usando *wavelets* ou o codec AMR. Diversos dos novos modos mostraram-se mais eficientes, pelos critérios apresentados, que o trabalho anterior utilizando AMR.

ABSTRACT

COMPRESSION OF ELECTROMYGRAPHIC SIGNALS USING ADAPTIVE PREDICTIVE MULTI-RATE CODEC

Author: José Haroldo de Assis Cavalcante

Supervisor: Lúcio Martins da Silva

Programa de Pós-Graduação em Engenharia Elétrica

Brasília, February, 2009

The development of highly efficient speech codecs has raised the interest in their use in other types of signal, such as electrophysiological ones. This study evaluates the AMR-ACELP codec, developed for third generation mobile telephony, and proposes changes to improve its performance in compression of electromyographical signals, considering the objective quality of reconstructed signal. Such changes were used in several new modes of operation of the codec, which was evaluated using EMG signals from isometric contractions. There were obtained compression rates from 81.4% to 90.3%, and a minimum mean square error of 1.81%. It was found that the objectively measured quality of the signal increases with the exclusion of the elements of subjective quality enhancement and the adaptive dictionary and with the inclusion of a second fixed dictionary. The results are consistent, in terms of compression rate and error, with previous works on compression of EMG under similar conditions, using wavelets or the AMR codec. Some of the new modes were more efficient, under the mentioned criteria, than the previous work using AMR.

SUMÁRIO

1 –	INTRODUÇÃO.....	1
1.1 –	PROCESSAMENTO E COMPRESSÃO DE SINAIS ELETROFISOLÓGICOS.....	2
1.2 –	CODIFICAÇÃO DE VOZ	4
1.3 –	OBJETIVO.....	6
1.4 –	ORGANIZAÇÃO DO TRABALHO	6
2 –	CONTRAÇÃO MUSCULAR E ELETROMIOGRAFIA	9
2.1 –	FISIOLOGIA DA CONTRAÇÃO MUSCULAR.....	9
2.2 –	AQUISIÇÃO E CARACTERÍSTICAS DO SINAL ELETROMIOGRÁFICO... ..	12
2.3 –	COMPARAÇÃO COM SINAIS DE VOZ	13
3 –	O CODIFICADOR DE VOZ AMR-ACELP	15
3.1 –	PRINCÍPIOS DE CODIFICAÇÃO CELP	15
3.2 –	CODIFICAÇÃO MULTITAXA.....	18
3.3 –	FLUXO DO SINAL NO CODIFICADOR.....	20
3.4 –	A ALOCAÇÃO DE BITS DOS MODOS DO CODEC AMR.....	22
3.5 –	DESCRIÇÃO FUNCIONAL DO CODIFICADOR AMR.....	24
3.5.1 -	Pré-processamento.....	24
3.5.2 -	Predição de curto termo.....	24
3.5.3 -	Análise tonal em laço aberto	28
3.5.4 -	Resposta impulsional.....	29
3.5.5 -	Determinação do sinal-alvo.....	30
3.5.6 -	Dicionário adaptativo	30
3.5.7 -	Dicionário fixo	31

3.5.8 -	Quantização do ganho dos dicionários	37
3.5.9 -	Atualização das memórias	37
3.6 -	FLUXO DO SINAL NO DECODIFICADOR.....	37
3.7 -	DESCRIÇÃO FUNCIONAL DO DECODIFICADOR AMR.....	38
3.7.1 -	Decodificação e Síntese.....	38
3.7.2 -	Pós-processamento anti-esparsidade	39
3.7.3 -	Cálculo da voz reconstruída	39
3.7.4 -	Proteção contra instabilidades	40
3.7.5 -	Pós-processamento	40
3.8 -	Implementação em software.....	42
3.9 -	Formatos de arquivo	43
4 -	VERSÃO DO AMR-ACELP ADAPTADA PARA SINAIS EMG.....	47
4.1 -	ESTRATÉGIA PARA IMPLEMENTAÇÃO DE NOVOS MODOS	47
4.2 -	MODIFICAÇÕES GERAIS.....	47
4.3 -	ALTERAÇÃO DO CRITÉRIO BÁSICO DE OTIMIZAÇÃO	48
4.4 -	RETIRADA DE PROCEDIMENTOS PARA MELHORIA SUBJETIVA DA VOZ.....	49
4.5 -	RETIRADA DO DICIONÁRIO ADAPTATIVO	49
4.6 -	UTILIZAÇÃO DE DOIS DICIONÁRIOS FIXOS.....	50
4.7 -	RESUMO DOS NOVOS MODOS	50
4.8 -	CONSIDERAÇÕES SOBRE OUTROS PROCEDIMENTOS	52
5 -	SIMULAÇÕES E RESULTADOS	53
5.1 -	IMPLEMENTAÇÃO	53
5.2 -	SINAIS UTILIZADOS NAS SIMULAÇÕES.....	53
5.3 -	MEDIDAS DE DESEMPENHO.....	54
5.4 -	DESEMPENHO DO CODIFICADOR AMR-ACELP ORIGINAL	55

5.5 –	DESEMPENHO DO CODIFICADOR AMR-ACELP MODIFICADO	57
5.6 –	EFEITO DA RETIRADA DOS PROCEDIMENTOS DE MELHORIA DA QUALIDADE SUBJETIVA DA VOZ.....	65
5.7 –	EFEITO DA RETIRADA DO DICIONÁRIO ADAPTATIVO.....	67
5.8 –	EFEITO DA ASSOCIAÇÃO DE DOIS DICIONÁRIOS FIXOS E DO USO DE UM OU DOIS CONJUNTOS DE PARÂMETROS LINEARES	67
5.9 –	COMPARAÇÃO COM TÉCNICA DE COMPRESSÃO USANDO ACELP.....	68
5.10 –	COMPARAÇÃO COM TÉCNICAS DE COMPRESSÃO USANDO WAVELETS E H.264	70
6 –	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS.....	71
6.1 –	CONCLUSÕES GERAIS	71
6.2 –	SUGESTÕES PARA TRABALHOS FUTUROS	71
	REFERÊNCIAS BIBLIOGRÁFICAS	73

LISTA DE TABELAS

Tabela 3.1 – Modos utilizados no AMR	20
Tabela 3.2 – AMR - Alocação de bits por modo, por parâmetro.	23
Tabela 3.3 - Posições potenciais de pulsos no dicionário fixo - modo 12,2 kbit/s.....	32
Tabela 3.4 - Posições potenciais de pulsos no dicionário fixo - modo 10,2 kbit/s.....	33
Tabela 3.5 - Posições potenciais de pulsos no dicionário fixo - modos 7,95 e 7,40 kbit/s.	33
Tabela 3.6 - Posições potenciais de pulsos no dicionário fixo - modo 6,70 kbit/s.....	33
Tabela 3.7 - Posições potenciais de pulsos no dicionário fixo - modo 5,90 kbit/s.....	34
Tabela 3.8 - Posições potenciais de pulsos no dicionário fixo - modos 5,15 e 4,70 kbit/s.	34
Tabela 3.9 – tipos de arquivo do AMR.	44
Tabela 3.10 - Formato do quadro do arquivo de saída do codificador.....	45
Tabela 3.11 – Quadro codificado no formato ETSI.	45
Tabela 4.1 - Propriedades dos novos modos.	50
Tabela 5.1 – Desempenho do modo original AMR_122 (CF = 87,3%) para sinais de eletromiografia e de voz.	55
Tabela 5.2 – Desempenho (valores de PRD) dos novos modos de codificação para sinais de EMG.	57
Tabela 5.3 - Desempenho (valores de MSE) dos novos modos de codificação para sinais de EMG.	58

Tabela 5.4 –Desempenho (valores de PRD) dos novos modos de codificação para
sinal de voz..... 58

Tabela 5.5 - Desempenho (valores de MSE) dos novos modos de codificação para
sinal de voz..... 58

LISTA DE FIGURAS

Figura 1.1 - Qualidade x Taxa de Bits para diferentes técnicas de codificação de voz (adaptado de Hanzo et al, 2007).....	5
Figura 2.1 - Representação esquemática dos mecanismos básicos de controle motor e da unidade motora e seus componentes (adaptado de Merletti, 2004).....	10
Figura 2.2 - Características contráteis de fibras rápidas (esquerda) e lentas (direita), mostrando a estimulação elétrica na junção neuromuscular (acima) e a resposta biomecânica (abaixo) (adaptado de Merletti, 2004).....	11
Figura 2.3 - Relação entre percentual de contração voluntária máxima (CVM), frequência média (FM) e espectro de potência de EMG (adaptado de Merletti, 2004).....	13
Figura 2.4 - Forma de onda e espectro do som sonoro /U/ (adaptado de Goldberg, 2000).....	14
Figura 2.5 - Forma de onda e espectro do som surdo /sh/ (adaptado de Goldberg, 2000).....	14
Figura 3.1 - Estrutura do codificador CELP (adaptado de Silva, 1996).....	16
Figura 3.2 - Diagrama de blocos simplificado do codificador adaptativo multitaxa (adaptado de Santos, 2006).....	21
Figura 3.3 - Janelamento no modo 12.2.	25
Figura 3.4 - Diagrama de blocos simplificado do decodificador adaptativo multitaxa.....	37
Figura 4.1 – Relações entre os modos originais do AMR e os novos modos.	51

Figura 5.1 – Sinal de EMG, original e reconstruído após compressão no modo AMR_122.....	56
Figura 5.2 - Sinal de voz, original e reconstruído após compressão no modo AMR_122.....	56
Figura 5.3 – Formas de onda do sinal de EMG original e reconstruídos a partir da codificação usando o modo AMR_122 e os novos modos.	59
Figura 5.4 - Formas de onda do sinal de voz original e reconstruídos a partir da codificação usando o modo AMR_122 e os novos modos.	60
Figura 5.5– Comparação do desempenho dos novos modos de codificação para os sinais de voz e EMG.....	61
Figura 5.6 – PRD <i>versus</i> taxa de compressão para os novos modos de codificação aplicados a sinais EMG.	62
Figura 5.7 - Formas de onda de sinal de EMG original e reconstruído após codificação usando os modos mais eficientes em termos da relação entre o erro e o fator de compressão.	63
Figura 5.8 - Formas de onda de sinal de voz original e reconstruído após codificação usando os modos mais eficientes em termos da relação entre o erro e o fator de compressão.	64
Figura 5.9 – Variação de desempenho devida às rotinas para a melhoria da qualidade subjetiva.....	65

Figura 5.10 – Formas de onda do sinal de EMG original e reconstruído a partir da codificação com os modos AMR_122 e 2LP/DA/DF10, evidenciando o efeito da retirada das rotinas de melhoria da qualidade subjetiva da voz do codec original.....	66
Figura 5.11 - Formas de onda do sinal de voz original e reconstruído a partir da codificação com os modos AMR_122 e 2LP/DA/DF10, evidenciando o efeito da retirada das rotinas de melhoria da qualidade subjetiva da voz do codec original.....	66
Figura 5.12 - Comparação entre o desempenho da compressão de EMG obtido por Carotti e por alguns dos novos modos de codificação.	69
Figura 5.13 - Comparação entre os resultados dos algoritmos empregados neste trabalho (CELP), por Berger et al (Wavelet) e Costa <i>et al</i> (H.264/AVC).....	70

LISTA DE ABREVIACÕES E SIGLAS

3G	Terceira geração de telefonia celular
3GPP	<i>3rd Generation Partnership Project</i>
AbS	<i>Analysis by Synthesis</i>
ACELP	<i>Algebraic code excited linear prediction</i>
AGC	<i>Adaptative Gain Control</i>
AMR	<i>Adaptive-Multi-Rate (codec)</i>
AMR-WB	<i>Adaptive Multi-Rate Wideband (codec)</i>
ANSI	<i>American National Standards Institute</i>
CELP	<i>Code Excited Linear Predictive (codec)</i>
CODEC	Codificador-decodificador
DTX	<i>Discontinuous transmission</i>
CF	<i>Compression Factor</i>
CVM	Contração voluntaria máxima
ECG	Eletrocardiografia
EDGE	<i>Enhanced Data rates for GSM Evolution</i>
EEG	Eletroencefalografia
EFR	<i>Enhanced Full Rate</i>
EMG	Eletromiografia
ETSI	<i>European Telecommunications Standards Institute</i>
GPRS	<i>General Packet Radio Service</i>
GNU	<i>GNU is Not Unix</i>
GSM	<i>Global System for Mobile Communications</i>
LP	<i>Linear Prediction</i>
LPC	<i>Linear Predictive Coding</i>
LSB	<i>Least significant byte</i>
LSF	<i>Line Spectral Frequencies</i>
LSP	<i>Line Spectral Pairs</i>
MIME	<i>Multipurpose Internet Mail Extensions</i>
MMS	<i>Multimedia Messaging Service</i>
MSB	<i>Most significant byte</i>
MSE	<i>Mean Squared Error</i>
MUAP	<i>Motor unit action potential</i>
PC	<i>Personal Computer</i>
PCM	<i>Pulse Code Modulation</i>
PDC	<i>Personal Digital Cellular</i>
PRD	<i>Percent Residual Difference</i>
RMS	<i>Root mean square</i>
SID	<i>Silence Indicator</i>
SMQ	<i>Split Matrix Quantization</i>
SVQ	<i>Split Vector Quantization</i>
TDMA	<i>Time division multiple access</i>
VAD	<i>Voice activity detection</i>
WCDMA	<i>Wideband Code Division Multiple Access</i>

1 – INTRODUÇÃO

Sinais obtidos a partir do registro da atividade eletrofisiológica, tais como os eletrocardiográficos, eletroencefalográficos e eletromiográficos, são de indiscutível relevância no contexto clínico, de pesquisa ou mesmo desportivo.

A partir de conceitos como o da interface homem-máquina (Lévy, 1993), do corpo biocibernético — o “corpo plugado” (Santaella, 2003) —, e de computadores “invisíveis” que utilizem sensores de sinais biológicos (Kaku, 2001), a literatura técnica tem recentemente apresentado propostas para sistemas de *eHealth*. Tais sistemas incluem aplicações em rede utilizando processamento de imagens, sinais e dados distribuídos (Maglogiannis, 2007), sistemas “vestíveis” e portáteis (Gatzoulis, 2007) e, em particular, de telemedicina, inclusive com o emprego de dispositivos móveis (Shimizu, 1999). Essa última aplicação vem sendo facilitada pela crescente difusão de tecnologias de comunicação de dados em alta velocidade, em especial a Internet, e tecnologias sem fio, baseadas em redes sejam celulares, quanto *Wi-Fi* ou *WiMax*, aliada à constante miniaturização e integração de sistemas eletrônicos.

Por outro lado, fatores limitadores da transmissão de dados, técnicos ou econômicos, tornam interessante ou mesmo indispensável a aplicação de técnicas de compressão aos sinais eletrofisiológicos. Entretanto, de modo diverso à transmissão de sinais de voz, para a qual a qualidade subjetivamente percebida é a mais importante a ser garantida, a compressão dos sinais eletrofisiológicos pode exigir um maior compromisso entre a taxa de compressão obtida e a qualidade objetiva do sinal reconstruído. De fato, para muitos sinais eletrofisiológicos, embora o volume de dados possa ser alto — em especial se o tempo total de registro for longo —, deve-se garantir que o sinal reconstruído possua uma fidedignidade acima de um mínimo com o original, para minimizar a produção de erros que poderiam acarretar uma incorreta interpretação diagnóstica.

Atenção especial tem sido dada à compressão de sinais de imagens, sons e voz, com aplicações que incluem seu armazenamento e transmissão — o que, no caso da voz, se deve, ao menos em parte, à pressão técnica e comercial para acomodar o crescente número de usuários de serviços de telefonia — celular, em especial. Deste modo, parece razoável supor que, pelo grau de sofisticação atingido por tais técnicas, seu aproveitamento possa ser vantajoso em outros contextos, tal como em sinais obtidos a partir do registro da

atividade eletrofisiológica, com a finalidade de armazenamento ou de transmissão, ainda que após o devido processo de adaptação e otimização. Assim, ao se avaliar se uma técnica de compressão originalmente desenvolvida para sinais de voz pode ser aproveitada para determinado sinal eletrofisiológico, deve ser levada em conta sua compatibilidade com o novo tipo de sinal – além de elementos como a largura de banda, taxa de amostragem, características espectrais — e a necessidade e factibilidade de introduzir alterações.

1.1 – PROCESSAMENTO E COMPRESSÃO DE SINAIS ELETROFISOLÓGICOS

O processamento de sinais eletrofisiológicos vem sendo pesquisado com ênfase em sinais cardíacos (ECG) (Jalaledine, 1990; Al-Fahoum, 2006) e eletroencefalográficos (EEG) (Argwal, 2001; Wongsawat, 2006), inclusive com aplicações em telemedicina (Kong, 2001), em função de estarem diretamente relacionados ao funcionamento de órgãos essenciais para a sustentação da vida.

Entretanto, o processamento e análise de sinais eletromiográficos (EMG) é também relevante em diversos contextos, encontrando diversas aplicações clínicas. Variáveis eletromiográficas podem apresentar relações inusitadas com outros estados fisiológicos (Soares, 2007). A interpretação dos registros de eletromiografia dos músculos dos membros inferiores durante a locomoção pode contribuir para o diagnóstico e tratamento de patologias neurológicas ou musculares (Cavalcante, 1991). Em reabilitação, sinais de eletromiografia podem ser utilizados como estímulo-resposta em técnicas de *biofeedback* (Basmajian, 1989; Cavalcante, 1994). Registros de EMG de longa duração são utilizados clinicamente para estimação de espasticidade em pacientes com lesão medular completa (Tepavac, 1992), para avaliação da fadiga muscular, movimentos espasmódicos e alterações posturais. O processamento e análise de EMG são também relevantes na área esportiva (Fox, 1991; Al-Imari, 2003) e ergonômica. Eletromiografia de músculos faciais vem sendo utilizada para facilitar o reconhecimento da fala (Jou, 2006) em ambientes ruidosos (Betts, 2005), com aplicações que variam do auxílio a deficientes à exploração espacial (Binsted, 2006; Bradley, 2006). Em particular, o uso de sinais registrados com eletrodos superficiais (ver Seção 2.1) encontra aplicações clínicas (Kleissen, 1998; Hogrel, 2005), e na avaliação do desempenho motor em diversos ambientes (Winter, 1990; Veneziano, 2006).

A compressão de sinais procura, em geral, atender duas necessidades principais: armazenamento e transmissão. No caso específico de sinais eletrofisiológicos (Nait-Ali, 2008), que podem apresentar grande volume, em razão de sua extensão e resolução, a necessidade de otimização dos recursos de armazenamento se faz presente tanto na utilização de dispositivos portáteis, com capacidade limitada de memória, tal como é comumente efetuado com registros de longa duração de ECG, com a utilização de dispositivos portáteis (Lin, 1997), quanto em grandes bases clínicas, em que pode ser necessário o armazenamento de dados de grande número de indivíduos, por diversos anos. A transmissão de sinais eletrofisiológicos, por sua vez, vai de encontro a questões de largura de banda e velocidade de transmissão. O espectro de frequências é um bem limitado e o meio de transmissão é comumente um recurso compartilhado; a Internet, em especial, gera preocupações ligadas à perda de pacotes e atrasos, que aumentam com o volume a ser trafegado. Assim, várias técnicas têm sido propostas para a compressão destes sinais (Salomoni, 2007).

A compressão de sinais de EMG, em particular, se mostra relevante com registros de eletromiografia de longa duração (conforme mencionado anteriormente), para transmissão sem fio (Hughes, 2007), ou fazendo parte de tecnologias de telereabilitação (Holden, 2007).

Diversas técnicas têm sido propostas para a compressão de sinais eletromiográficos. Guerrero e Mailhes (1997) separaram estas técnicas em dois grupos: métodos preditivos (também referenciados como diretos) e métodos por transformação. Entre os métodos por transformação já avaliados, estão as transformadas Karhunen-Loève, de Coseno Discreta e *Wavelet* Discreta. Entre os métodos preditivos estão a Modulação por Codificação de Pulso Diferencial (DPCM), comumente utilizada em sinais de voz, a Codificação Multi-Pulso (MPC) e a Codificação Preditiva Linear com Excitação por Código (CELP).

Compressão de eletromiografia por meio de *wavelet* foi estudada, dentre outros, por Berger (2006b), que empregou redes neurais para otimizar a alocação dinâmica de bits, e Brechet, que aplicaram esta técnica para outros sinais eletrofisiológicos como o eletrocardiograma (Brechet, 2007; Berger, 2006a, 2006c, 2008).

Algoritmos utilizados na codificação de imagens e vídeo já foram adaptados para compressão de EMG, tais como, respectivamente, o EZW, utilizado no JPEG2000 (Carotti) e o H.264/AVC (Costa, 2008).

Grönfors investigou a codificação vetorial na compressão de EMG (Grönfors, 2006a), tendo efetuado a comparação entre diversos métodos de quantização vetorial deste sinal (Grönfors, 2005) e verificado o efeito do comprimento do vetor e do nível do ganho de quantização (Grönfors, 2006b).

A compressão de eletromiografia utilizando técnicas preditivas vem sendo atualmente pesquisada por grupos como o da Politécnica de Turim (Itália). Usando codificação linear preditiva (Carotti, 2005), apenas o contorno espectral do sinal de EMG reconstruído foi mantido. As características espectrais (frequência média e mediana, variância, assimetria, curtose) foram preservadas com erro menor que 3%, com um fator de compressão de 97,1%, representando uma redução da taxa de bits de 24 kbit/s para 352 bit/s. No caso do uso do ACELP (Carotti, 2006), um algoritmo de compressão de voz foi adaptado às características do sinal de EMG. A taxa de compressão foi de 87,3%. O erro RMS para todos os sinais esteve abaixo de 10%. O método preservou o sinal no domínio do tempo e da frequência. Por fim, para a compressão de EMG multicanal, (Carotti, 2007) foi explorada a correlação entre os pares de linhas do espectrais (LSF) do codificador ACELP.

1.2 – CODIFICAÇÃO DE VOZ

Algoritmos de codificação de sinais podem ser classificados em codificadores de forma de onda, codificadores paramétricos e codificadores híbridos (Chu, 2003). Codificadores de forma de onda procuram preservar a forma original do sinal e podem assim ser aplicados a qualquer tipo de sinal. Tais codificadores são mais adequados a taxas de bit mais altas, pois a qualidade do sinal cai abruptamente com o decréscimo da taxa de bit. Já os codificadores paramétricos utilizam um modelo controlado por parâmetros, calculados a partir do sinal de entrada, para reconstruir o sinal. Não há preocupação em preservar a forma original do sinal, cuja qualidade subjetivamente percebida está ligada às características do modelo empregado, que em geral é específico para sinais de voz. O modelo de produção vocal introduzido por Rabiner e Schafer (1978) é utilizado frequentemente. Codificadores paramétricos podem operar em taxas de bits mais baixas que os codificadores de forma de onda para uma dada qualidade percebida. PCM (*pulse code modulation*) e LPC (*linear*

pulse coding) são exemplos conhecidos de codificadores de forma de onda e paramétricos, respectivamente.

Já os codificadores híbridos combinam características de codificadores paramétricos (tal como um modelo de produção de voz) e de forma de onda (procurando preservar a forma original do sinal). Ao contrário de um codificador paramétrico, o codificador híbrido inclui no modelo de produção de voz um sinal de excitação, que é transmitido como parte do fluxo de bits codificado. O codificador CELP (*code-excited linear prediction*) é um exemplo representativo dos codificadores híbridos (Hanzo et al, 2007), sendo largamente utilizado em telefonia celular.

A relação entre qualidade (subjetivamente avaliada) e taxa de bits para as mencionadas técnicas de codificação de voz é mostrada na Figura 1.1. Destaque-se que, com os codificadores híbridos, consegue-se manter melhor qualidade de voz, com a diminuição da taxa, em relação aos demais.

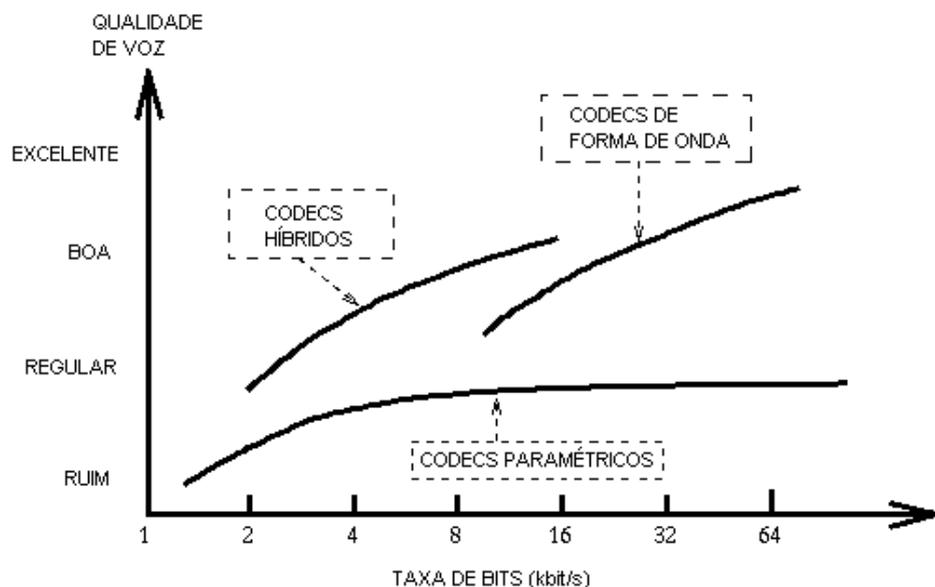


Figura 1.1 - Qualidade x Taxa de Bits para diferentes técnicas de codificação de voz (adaptado de Hanzo et al, 2007).

Em alguns codificadores, os parâmetros são simplesmente extraídos do sinal de entrada e quantizados para serem posteriormente utilizados na produção do sinal sintetizado. Outros utilizam um procedimento de otimização em laço fechado, em que os parâmetros são utilizados para produzir o sinal sintetizado durante a codificação; os parâmetros são então

sucessivamente ajustados até que sejam determinados aqueles que resultam na produção da voz sintetizada que mais se aproxime da original. Uma vez que para isso o sinal deve ser sintetizado durante a codificação, para análise, este método é conhecido como “análise-por-síntese”.

1.3 – OBJETIVO

O objetivo deste trabalho é, a partir do conhecimento de eletromiografia e do funcionamento do codec AMR-ACELP, propor alterações neste codec que o tornem adequado à compressão de sinais de eletromiografia, sob um ponto de vista de indicadores objetivos de qualidade do sinal reconstruído, e consecutivamente avaliar estas alterações realizando simulações e comparações com outras técnicas.

1.4 – ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em seis capítulos.

Este capítulo 1 conceitua e justifica a compressão de sinais, com ênfase na compressão de sinais eletrofisiológicos e, em particular, eletromiográficos, e apresenta os princípios dos codificadores em geral.

No capítulo 2 é efetuada a discussão sobre eletromiografia de superfície, incluindo fisiologia muscular que leva à produção do sinal, formas de captação e as características do sinal no tempo e em frequência. São também analisadas as semelhanças e diferenças deste sinal para com os de voz, para os quais os codificadores em análise foram originalmente destinados.

No capítulo 3 é apresentado o codificador CELP e realizado o estudo aprofundado do codec ACELP multitaxa “Adaptive Multi Rate” (AMR) conforme padronizado pelo 3rd Generation Partnership Project (3GPP).

O capítulo 4 apresenta as modificações efetuadas no codec AMR, de forma a adequá-lo à compressão de sinais EMG.

O capítulo 5 descreve os sinais e medidas de desempenho utilizados nas simulações realizadas com as versões modificadas do codec AMR e apresenta os resultados obtidos e comparações com trabalhos prévios.

Por fim, no capítulo 6 são apresentadas as conclusões sobre o estudo e as simulações realizadas. São também propostas sugestões para trabalhos futuros.

Ao final, encontram-se as referências bibliográficas que compreendem a literatura tradicional e recente a respeito dos tópicos abordados neste trabalho, como eletromiografia, processamento digital de sinais (em especial os biológicos), processamento de voz, compressão de sinais.

2 – CONTRAÇÃO MUSCULAR E ELETROMIOGRAFIA

Neste capítulo são apresentados fundamentos do processo de contração muscular e da correspondente geração de potenciais elétricos e de seu registro por meio da eletromiografia (EMG). São feitas ainda comparações entre registros eletromiográficos e de voz.

A eletromiografia é um instrumento de fundamental importância na análise biomecânica do movimento humano. Diversas técnicas de processamento digital de sinais têm sido empregadas no tratamento deste sinal.

2.1 – FISIOLOGIA DA CONTRAÇÃO MUSCULAR

A unidade anatômica e funcional do tecido muscular é a fibra muscular. Esta fibra tem um comprimento de até 30 cm, tendo, porém, menos que 100 micrômetros de espessura (Basmajian, 1975). Fisiologicamente, cada fibra é uma célula cuja principal característica é sua capacidade de contração, fenômeno que ocorre em resposta a determinados estímulos eletroquímicos. Tais estímulos são em geral produzidos e transmitidos por outras estruturas musculares ou por células nervosas, muitas vezes em associação com estruturas hierarquicamente superiores de controle de movimentos voluntários ou involuntários.

As fibras geralmente não se contraem de modo isolado. Ao contrário, pequenos grupos delas contraem-se ao mesmo tempo, quando são acionadas pelos ramos terminais de uma mesma célula nervosa. Uma mesma fibra nervosa pode se conectar a diversas fibras musculares (entre 10 e 2000) e estimulá-las simultânea e repetitivamente. O neurônio produz impulsos em intervalos variáveis. Este mesmo comportamento é observado na codificação da informação entre neurônios (Dayan, 2001).

O corpo da célula nervosa se encontra na substância cinzenta espinhal ou no cérebro. Coletivamente, o corpo da célula nervosa, seu axônio, seus ramos terminais e todas as fibras musculares acionadas por aqueles ramos constituem uma unidade motora, isto é, a unidade funcional de um músculo (Figura 2.1). Um impulso descendo do axônio faz com que todas as fibras musculares se contraíam quase simultaneamente. Impulsos são gerados em várias frequências, geralmente abaixo de 50 Hz (De Luca, 1979).

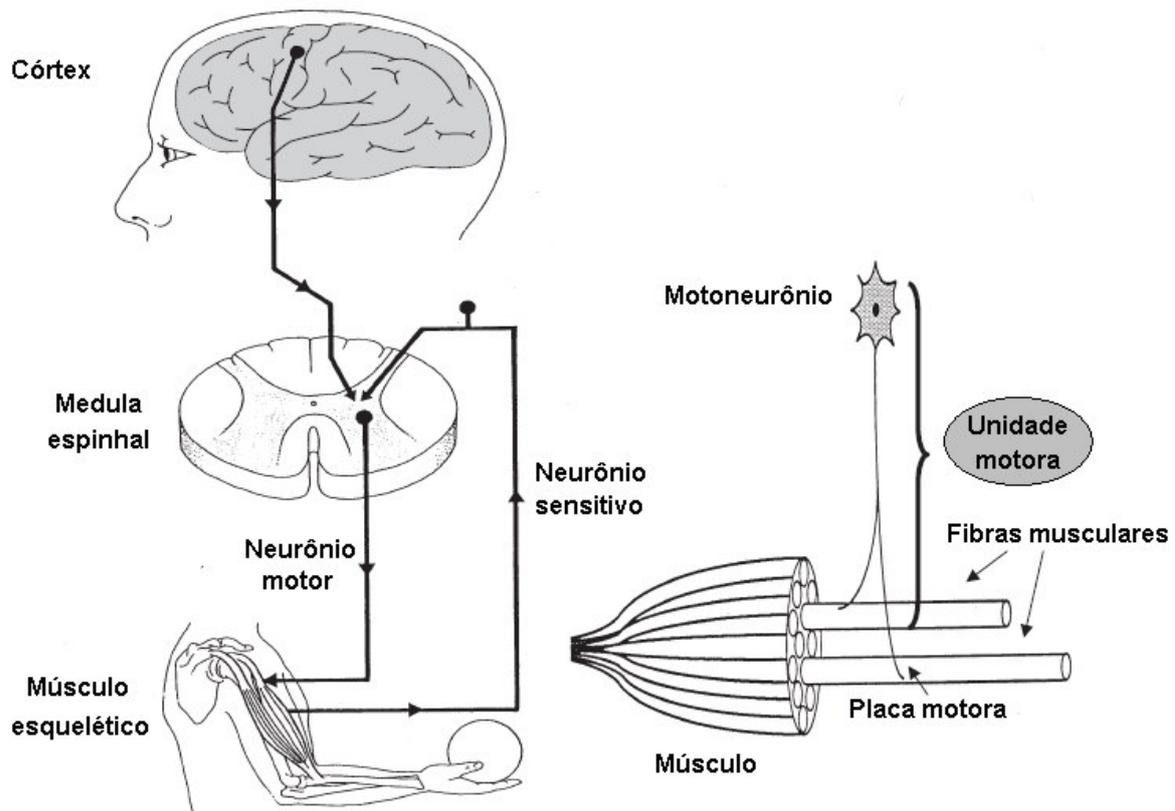


Figura 2.1 - Representação esquemática dos mecanismos básicos de controle motor e da unidade motora e seus componentes (adaptado de Merletti, 2004).

Os estímulos nervosos são transmitidos à superfície da fibra muscular por meio de uma junção denominada “placa motora”. Em tal junção, as células nervosas lançam moléculas do neurotransmissor acetilcolina (Kandel, 2000) que, ao atingirem a membrana plasmática da célula muscular, produzem um desequilíbrio eletroquímico (despolarização, hiperpolarização e repolarização), relacionado ao fluxo de íons pela membrana, gerando uma onda elétrica que se propaga pela superfície da fibra muscular e consecutivamente afeta estruturas protéicas internas à fibra, denominadas sarcômeros. O sarcômetro é formado por feixes superpostos das proteínas actina e miosina que, sob o impulso elétrico, alteram sua estrutura de forma a mudar para uma forma mais compacta.

O efeito global desta alteração é a contração, após alguns milissegundos, da fibra muscular (Figura 2.2). O "pico de tensão" associado à contração da fibra é habitualmente designado "potencial de ação".

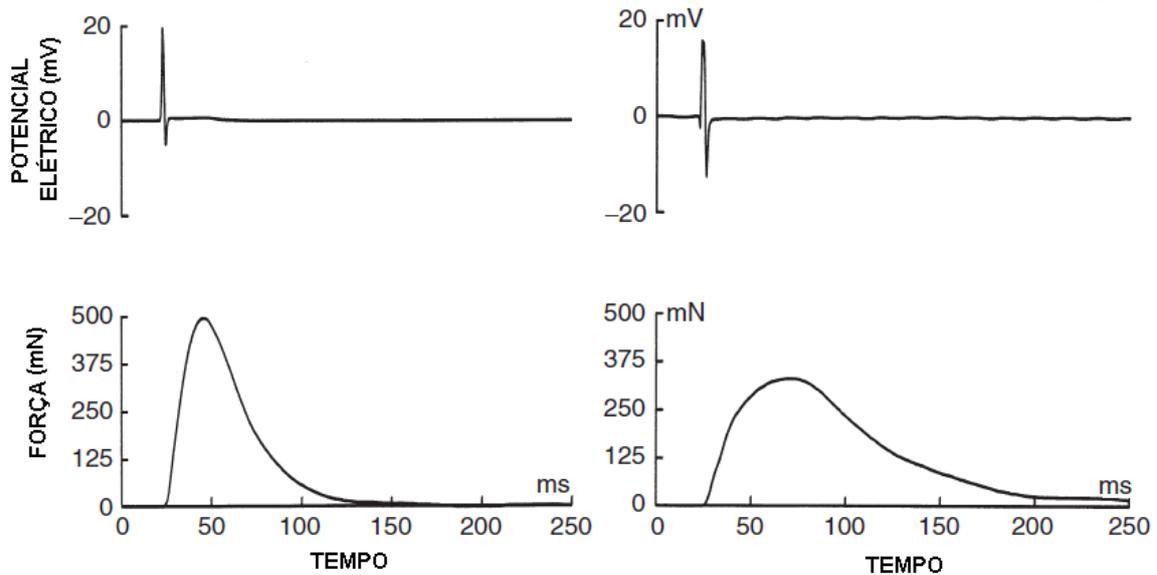


Figura 2.2 - Características contráteis de fibras rápidas (esquerda) e lentas (direita), mostrando a estimulação elétrica na junção neuromuscular (acima) e a resposta biomecânica (abaixo) (adaptado de Merletti, 2004).

A contração muscular obedece ao chamado “princípio tudo ou nada” (Hamill, 1999): cada fibra muscular, individualmente, ou se encontra plenamente contraída ou não o está. A variação da quantidade de fibras contraídas em um músculo é que determina seu nível global de contração. Em relação ao nível de contração muscular global, cada unidade motora possui um limiar de disparo, e uma “curva de redisparo” caracterizada por um período de redisparo que se reduz progressivamente conforme cresce o grau de contração. Em condições normais, unidades motoras menores são recrutadas de antemão. Conforme o esforço vai aumentando, automática ou conscientemente, unidades motoras maiores são recrutadas; ao mesmo tempo, aumenta a frequência de contração das unidades motoras já ativas.

A tensão em um músculo normal é espalhada através de uma rápida e abrangente série de contrações que ocorrem assincronamente entre as fibras. A contração global e continuada um determinado músculo é, na verdade, o resultado do somatório espacial e temporal dos pulsos assíncronos de contração individual das diversas unidades motoras.

2.2 – AQUISIÇÃO E CARACTERÍSTICAS DO SINAL ELETROMIOGRÁFICO

A despolarização e repolarização da membrana plasmática das fibras musculares são a fonte das mudanças de potencial elétrico de um músculo, cuja captação e registro se denomina eletromiografia.

A captação dos potenciais mioelétricos é realizada com a utilização de eletrodos, que podem ser internos ou de superfície. Eletrodos internos, de agulha ou de fio, prestam-se melhor à captação de potenciais de apenas uma ou poucas fibras musculares, enquanto que os eletrodos de superfície podem, em geral, captar os potenciais gerados na contração de um grande número de fibras.

Sob pequeno grau de contração muscular, o EMG apresenta uma sequência de picos bem individualizados, indicando os disparos individuais das unidades motoras. Entretanto, a partir de um nível suficientemente alto de contração, o sinal de EMG de superfície começa a assemelhar-se a um ruído com função de distribuição de probabilidade gaussiana ou laplaciana, denominado “padrão de interferência” resultante da superposição de inúmeros potenciais de ação, em razão dos diferentes momentos e limiares de disparo e curvas de redisparo das unidades motoras. A literatura menciona que um padrão de interferência típico possui amplitude de cerca de 3 mV e a maior parte de sua energia situa-se abaixo de 1-2 kHz (Basmajian, 1975) ou 400 Hz (Merletti, 2004), discrepâncias que refletem diferentes configurações e equipamentos utilizados para a captação do sinal.

A Figura 2.3 apresenta o EMG de uma contração muscular isométrica — em que o comprimento do músculo não se altera durante a contração —, mostrando que (a) o nível de tensão da EMG de superfície está diretamente relacionado ao grau de contração do músculo, em consequência do grau de recrutamento e da taxa de disparo de unidades motoras; (b) a frequência média do espectro de potência (neste caso situado por volta de 100 Hz) cai durante uma contração prolongada, em razão da fadiga muscular. A porcentagem do percentual de contração voluntária máxima pressupõe um registro de contração voluntária no máximo de esforço, a ser usado como referência.

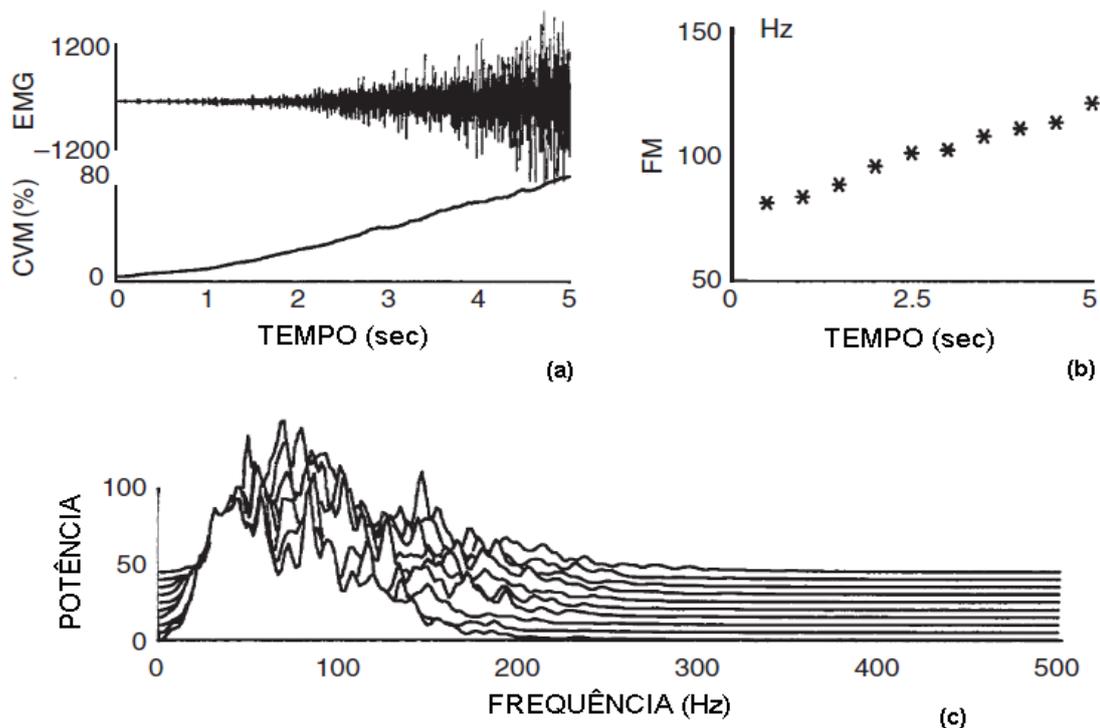


Figura 2.3 - Relação entre percentual de contração voluntária máxima (CVM), frequência média (FM) e espectro de potência de EMG (adaptado de Merletti, 2004).

2.3 – COMPARAÇÃO COM SINAIS DE VOZ

Um sinal de voz típico é composto de segmentos que podem ser classificados ou como sonoros ou como surdos. Na produção de sons sonoros, o ar emitido pelos pulmões atravessa livremente o aparelho fonador, enquanto que na produção de sons surdos o ar encontra obstáculos resultantes do posicionamento da língua, dentes e lábios (Titze, 2008).

Sons sonoros são foneticamente associados com vogais e consoantes sonoras (Câmara Júnior, 2007). Apresentam periodicidade, o que no domínio da frequência reflete-se na concentração da energia em torno de algumas poucas frequências, que são características de cada falante e resultado da estrutura particular de seu aparelho fonador (Figura 2.4).

Em contraste, sons surdos são foneticamente associados com consoantes surdas. Não apresentam periodicidade e seu espectro assemelha-se ao de um ruído (Figura 2.5).

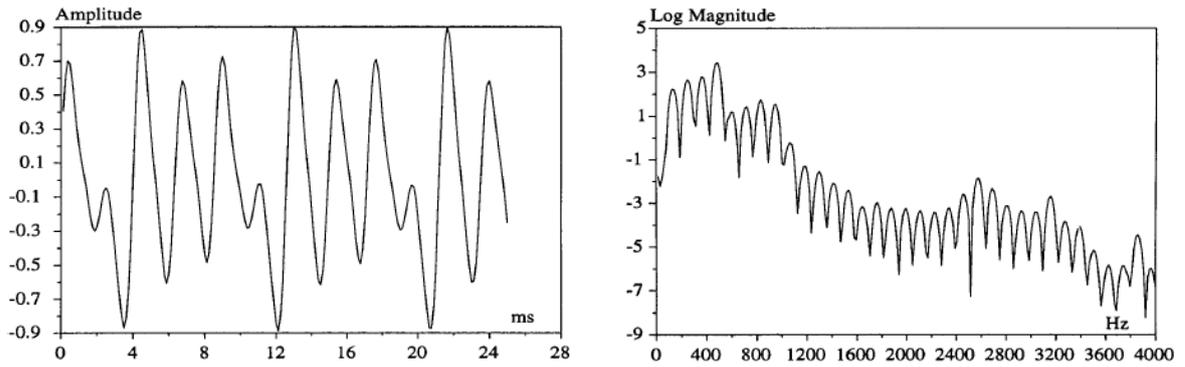


Figura 2.4 - Forma de onda e espectro do som sonoro /U/ (adaptado de Goldberg, 2000).

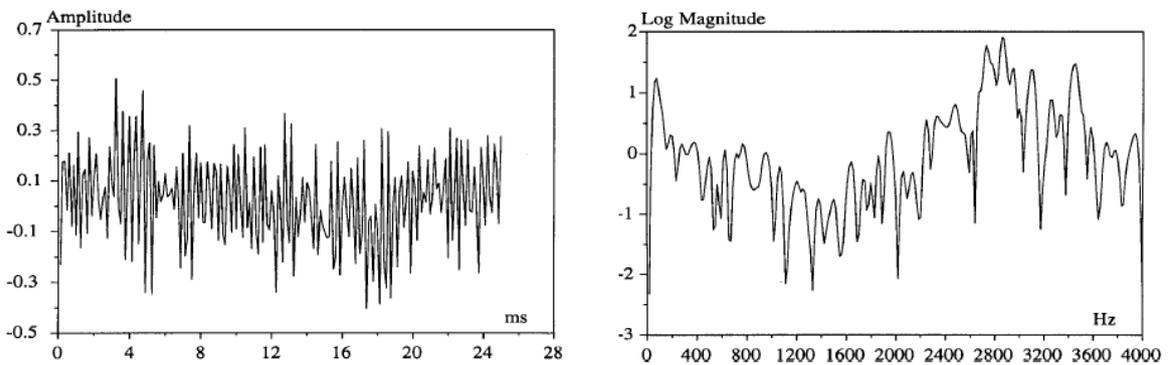


Figura 2.5 - Forma de onda e espectro do som surdo /sh/ (adaptado de Goldberg, 2000).

Sinais de eletromiografia de fibra única apresentam uma maior similaridade a sons sonoros, uma vez que é possível registrar as repetidas contrações de fibras individuais, enquanto que sinais de eletromiografia de superfície, em virtude de sua característica ruidosa acima de um nível suficientemente alto de contração, a sons surdos.

3 – O CODIFICADOR DE VOZ AMR-ACELP

O consórcio GSM adotou a codificação paramétrica como forma de compressão dos sinais de voz trafegados na interface aérea. O desenvolvimento mais moderno neste sentido é o codificador AMR que, por meio da análise da capacidade do canal e da qualidade do sinal efetua o balanceamento destes dois parâmetros, utilizando taxas de compressão variáveis.

O codificador ACELP multitaxa AMR (*Adaptive Multi-Rate*) foi originalmente desenvolvido para o padrão GSM, tendo sido posteriormente (1998) adotado pelo 3GPP (Projeto de Parceria de Terceira Geração) como o codec obrigatório para os sistemas de telefonia celular de terceira geração baseados na evolução do núcleo da rede GSM – WCDMA, EDGE, GPRS – (Hillebrand, 2001) e, em 1999, pelo Instituto Europeu de Padrões de Telecomunicação (ETSI). Sua documentação e código-fonte podem ser obtidos livremente nos sítios destas entidades.

Este capítulo apresenta os princípios da codificação CELP, da codificação multitaxa e do codec AMR, avançando para o fluxo do sinal e os detalhes da implementação do codificador e do decodificador. Conforme detalhado nas referências bibliográficas, os tópicos aqui apresentados podem ser aprofundados, em especial, nas especificações técnicas do 3GPP e em Kondoz, 2004; Chu, 2003; e Hanzo, 2007.

3.1 – PRINCÍPIOS DE CODIFICAÇÃO CELP

O codificador CELP foi introduzido em sua forma original por Schroeder e Atal (1985). No modelo CELP a voz é sintetizada alimentando-se um filtro de síntese de curto termo com um sinal de excitação construído pela adição de dois vetores, apropriadamente escolhidos, de um dicionário adaptativo e um dicionário fixo (também chamado de inovativo). O CELP é baseado no princípio da análise por síntese, sendo que as sequências de excitação são determinadas em laço fechado e os demais parâmetros, em laço aberto.

A estrutura geral do codificador CELP, conforme a configuração empregada no restante deste trabalho, é mostrada na Figura 3.1.

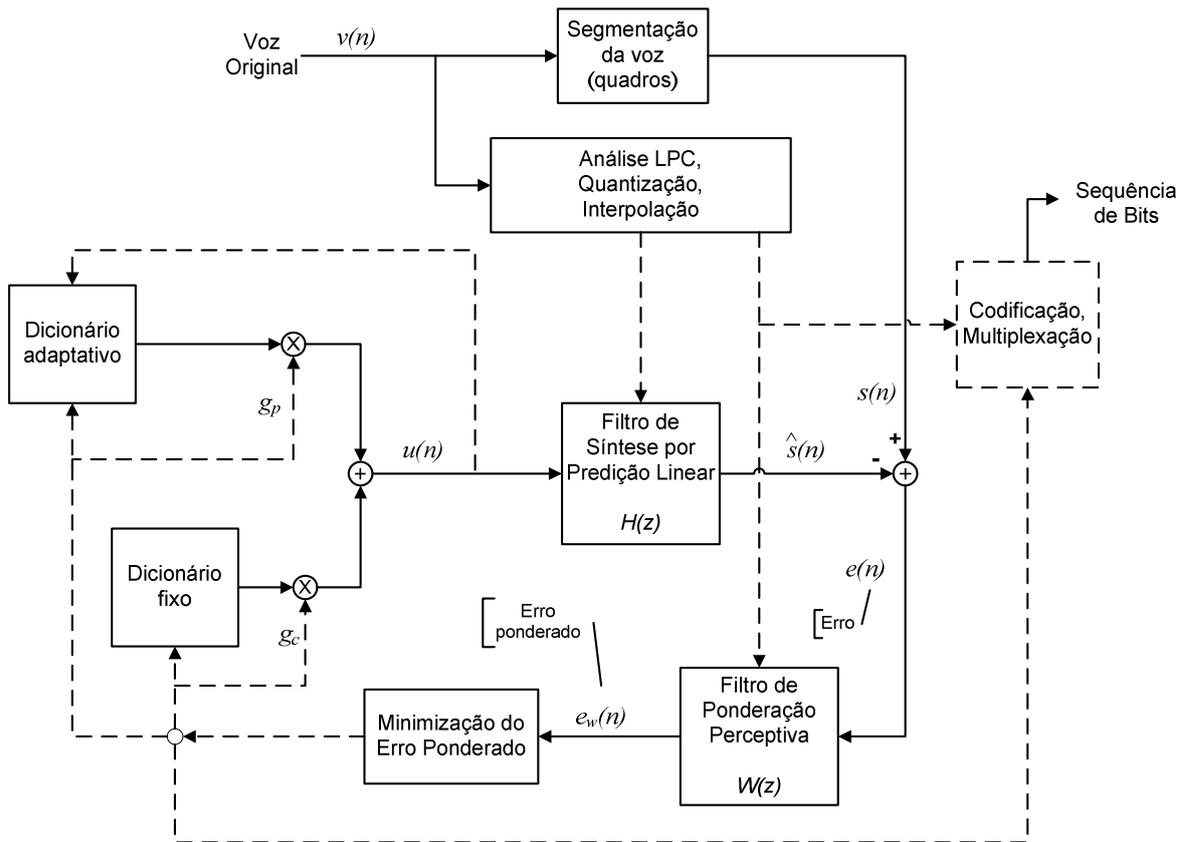


Figura 3.1 - Estrutura do codificador CELP (adaptado de Silva, 1996).

O codificador opera em quadros de voz de 20 ms, correspondendo a 160 amostras na frequência de amostragem de 8000 amostras/s. A cada 160 amostras o sinal de voz é analisado para extração dos parâmetros do modelo CELP (coeficientes do filtro preditivo linear, índices e ganhos dos dicionários fixo e adaptativo), que são codificados e transmitidos. No decodificador, os parâmetros são decodificados e a voz é reconstruída aplicando-se o sinal de excitação pelo filtro de síntese preditivo linear.

O filtro de síntese por predição linear, ou de curto termo, modela a envoltória espectral de cada segmento e é dado por:

$$H(z) = \frac{1}{\hat{A}(z)} = \frac{1}{1 + \sum_{i=1}^m \hat{a}_i z^{-i}} \quad (3.1)$$

onde \hat{a}_i , $i = 1, \dots, m$ são os parâmetros (quantizados) de predição linear, e $m = 10$ é a ordem do filtro.

O dicionário adaptativo consiste de versões deslocadas no tempo das sequências de excitação passadas e modela a correlação de longo prazo do sinal de voz, sendo implementado em um filtro de síntese de longo termo, dado por

$$\frac{1}{B(z)} = \frac{1}{1 - g_p z^{-T}} \quad (3.2)$$

onde T é o atraso ou período tonal (*pitch*) e g_p , o ganho tonal. Valores comuns de T estão na ordem de 10^{-3} a 10^{-2} s.

A sequência de excitação ótima em um dicionário é escolhida usando um procedimento de busca de análise-por-síntese, na qual é minimizado o erro quadrático médio, sujeito a uma ponderação perceptiva, entre a voz original e a sintetizada. Tendo em vista redução de complexidade, os dicionários adaptativo e fixo são varridos sequencialmente em busca das melhores entradas de código. Em um codificador ótimo, tanto o índice e o ganho do dicionário fixo como os parâmetros do dicionário adaptativo são otimizados conjuntamente para minimizar o sinal de erro, e_w . Entretanto, devido à alta complexidade e esforço computacional associados, uma abordagem sub-ótima é utilizada. Primeiramente, os parâmetros do dicionário adaptativo são determinados, assumindo que a excitação do dicionário fixo é zero, isto é, $g_c = 0$, uma vez que neste estágio da otimização a entrada do dicionário fixo não foi ainda determinada. Então, dado que os parâmetros do dicionário adaptativo foram encontrados (consistindo no atraso e ganho do dicionário adaptativo), os parâmetros do dicionário fixo são determinados.

O filtro de ponderação perceptiva utilizado na técnica de análise por síntese é derivado do filtro de síntese LPC e leva em conta o efeito psicoacústico de mascaramento do ruído de quantização, que é menos perceptível nas vizinhanças espectrais das frequências formantes, sendo dado por:

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (3.3)$$

Onde $A(z)$ é o filtro de predição linear não quantizado e $0 < \gamma_2 < \gamma_1 \leq 1$ são os fatores de ponderação perceptual. O filtro utiliza os valores $\gamma_1 = 0,9$ (para os modos de 12,2 e 10,2

kbit/s) ou $\gamma_1 = 0,94$ (demais modos) e $\gamma_2 = 0,6$. O filtro utiliza os parâmetros de predição linear não quantizados.

O codificador CELP algébrico ou ACELP (*algebraic code excited linear prediction*), introduzido por Adoul (1987), possui a mesma estrutura básica do CELP. Entretanto, seu dicionário fixo é algébrico, de modo a atender à preocupação em reduzir o custo computacional e propiciar maior eficiência em relação ao codificador CELP original, uma vez que a busca no dicionário de excitação é a mais intensa das operações do CELP (Chu, 2003).

3.2 – CODIFICAÇÃO MULTITAXA

A maior parte dos codecs empregados em sistemas de comunicação – tal como os codecs de GSM previamente existentes de taxa integral, meia taxa e taxa integral melhorado (Alencar, 2004) – opera a uma taxa de bits fixa, com um compromisso entre as codificações de fonte e de canal. Entretanto, estimar a qualidade do canal e ajustar a taxa de bits adaptativamente de acordo com as condições do canal tem o potencial de melhorar a resiliência a erros e a qualidade da fala em canais de comunicação sem fio (Hanzo, 2007).

A codificação multitaxa provê uma variação na alocação total de bits para um quadro de voz, adaptando a taxa às características fonéticas locais do sinal de voz, à qualidade do canal ou às condições de tráfego da rede. Isto é particularmente útil em comunicações celulares digitais, em que um dos maiores desafios é projetar um codec que seja capaz de prover fala de alta qualidade em uma ampla variedade de condições do canal. Idealmente, uma boa solução deve prover a mais alta qualidade possível sob perfeitas condições do canal, sendo capaz também de manter boa qualidade em ambientes hostis. A adaptação do modo do codec é uma característica chave do padrão AMR, que não havia sido usada previamente em nenhum outro padrão. Dada uma taxa de bits global fixa, este mecanismo de adaptação da codificação de fonte tem o potencial de alterar o particionamento entre a taxa de bits da fonte de voz e a redundância adicionada para proteção de erro (Makinen, 2004).

Quando a qualidade instantânea do canal está baixa, o codificador de voz opera em uma baixa taxa de bits, facilitando, deste modo, o emprego de controles de erro mais potentes,

dentro de uma disponibilidade fixa de taxa de bits, ou usando um modo de transceptor mais robusto, mas de baixas taxas. Em contraste, sob condições de canal favoráveis o codificador pode usar taxas de bits mais altas, resultando em alta qualidade de voz, uma vez que neste caso uma proteção de erro mais fraca é suficiente ou pode ser invocado um modo menos robusto, porém com taxas de bit mais altas.

Em redes GSM, o AMR permite um aumento na capacidade do canal de até 150%, em comparação ao codec *enhanced full rate* (EFR), considerando-se uma mesma qualidade subjetiva de voz. Em redes WCDMA, a operadora, ao ajustar a taxa de bits do AMR, pode variar a quantidade máxima de usuários por célula em cerca de 100%. Futuramente, a adaptação da taxa baseada na fonte poderá aumentar a capacidade do AMR em cerca de 25% (Holma, 2003). Dependendo das características da fonte, alguns quadros podem ser codificados com taxas menores com um decréscimo nulo ou muito pequeno na qualidade subjetiva da voz (Barreto, 2005).

O AMR possui ainda o recurso DTX, de transmissão descontínua, em que é feita a detecção de presença ou não de atividade vocal. Neste último caso, ocorre redução automática da taxa de codificação, por meio da mudança para o modo AMR_SID, em que são transmitidos apenas parâmetros indicadores de silêncio.

O AMR gera blocos codificados com 95, 103, 118, 134, 148, 159, 204 e 244 bits a cada 20 ms. Isto leva a taxas de bits de 4,75, 5,15, 5,9, 6,7, 7,4, 7,95, 10,2 e 12,2 kbps, respectivamente. Explicitamente, o codec AMR provê oito diferentes modos (além de um modo de indicação de silêncio).

A Tabela 3.1 identifica cada um destes modos e sua correspondência com codecs previamente desenvolvidos.

Três dos modos de codificação do AMR correspondem a padrões já existentes, fazendo com que sistemas de comunicação empregando este codec sejam interoperáveis com outros sistemas. Especificamente, o modo AMR_122 é idêntico ao padrão GSM EFR (Järvinen, 1997), os modos AMR_122 e AMR_740 correspondem aos codecs US1 e EFR (IS-641) do sistema TDMA, e o modo AMR_670 é equivalente ao codec EFR do sistema japonês PDC. Para cada um dos modos, existem codecs de canal correspondentes que executam o mapeamento entre os bits da fonte de voz e número fixo de bits dos canais codificados.

Tabela 3.1 – Modos utilizados no AMR

Modo	Bits/quadro	Taxa de bits (kbit/s)	Notas
AMR_122	244	12,20	Idêntico aos codecs GSM-EFR e TIA TDMA-US1.
AMR_102	204	10,20	
AMR_795	159	7,95	
AMR_740	148	7,40	Idêntico ao codec TDMA TIA/EIA IS-641.
AMR_670	134	6,70	Idêntico ao codec PDC-EFR ARIB.
AMR_590	118	5,90	
AMR_515	103	5,15	
AMR_475	95	4,75	
AMR_SID	36	1,80	Quadros de silêncio. Assumindo quadros SID transmitindo continuamente.

A adaptação de modo permite invocar um subconjunto de no máximo quatro modos, dos oito modos disponíveis. Este subconjunto é chamado de conjunto ativo do codec (ACS, da denominação em inglês). A adaptação de modo pode estar baseada na qualidade do canal, caso em que a probabilidade de se alternar entre modos é maior que de permanecer em um modo específico. Alterações muito frequentes de modo podem ser indesejáveis devido às resultantes variações de qualidade percebida. Um mecanismo mais desejável consiste no controle da taxa efetuado primariamente pela fonte (3GPP TS 26.093), apoiado por um controle de qualidade do canal.

O AMR opera sobre quadros de voz de 20 ms, correspondendo a 160 amostras a uma taxa de amostragem de 8000 amostras/s, com uma resolução de 13 bits justificados à esquerda. Assim, em um arquivo de entrada com codificação PCM a 16 bits, os três bits menos significativos são deixados em '0'. A saída do decodificador tem o mesmo formato. O AMR emprega o modelo ACELP de codificação de voz.

3.3 – FLUXO DO SINAL NO CODIFICADOR

O fluxo do sinal no codificador é mostrado na Figura 3.2.

Um filtro de análise LPC de décima ordem é empregado para modelar a correlação de curto prazo do sinal de voz. Predição de curto prazo ou, equivalentemente, análise

preditiva linear, é realizada para cada quadro de voz de 20 ms utilizando um algoritmo de Levinson-Durbin.

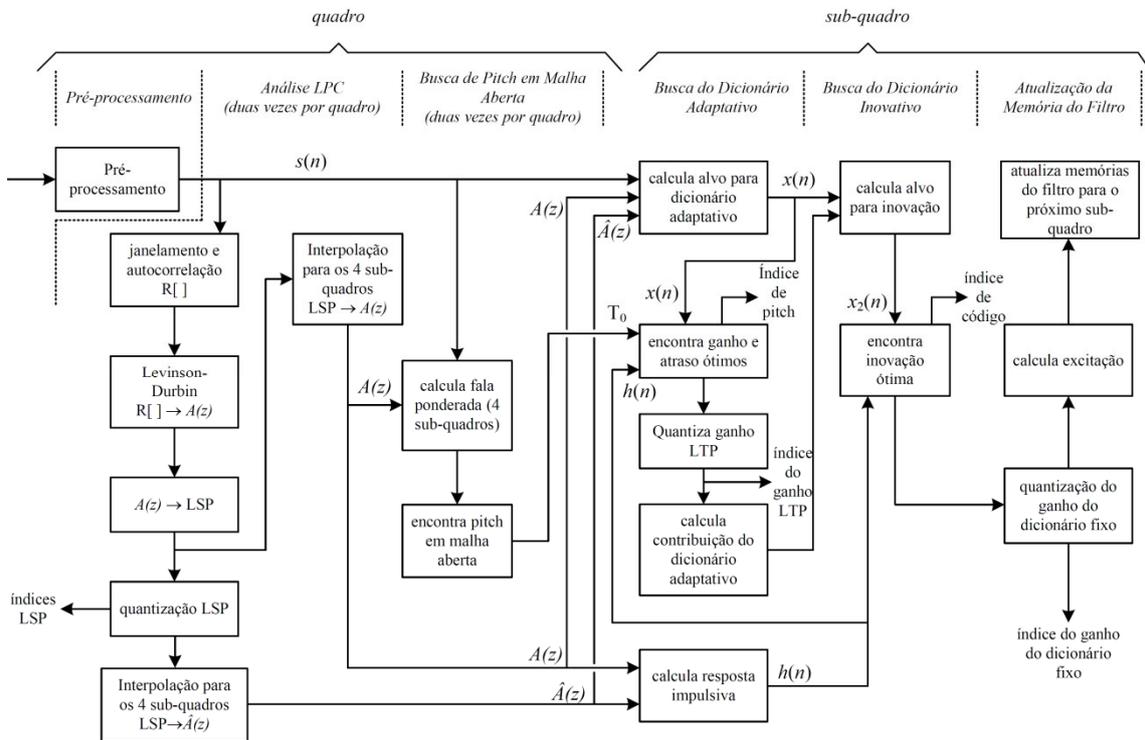


Figura 3.2 - Diagrama de blocos simplificado do codificador adaptativo multitaxa (adaptado de Santos, 2006).

A análise de predição linear é efetuada duas vezes por quadro para o modo de AMR_122 kbit/s e uma vez para os demais modos. Para o modo AMR_122, os dois conjuntos de parâmetros são convertidos em pares de linhas espectrais (LSP - *Line Spectral Pairs*) e quantizados conjuntamente usando quantização matricial partida (SMQ - *Split Matrix Quantization*) com 38 bits. Para os outros modos, o conjunto único de parâmetros LP é convertido para pares de linhas espectrais e quantizado usando quantização matricial partida.

O quadro de voz é dividido em quatro subquadros, cada um com 5 ms de duração (40 amostras). Os parâmetros dos dicionários adaptativo e fixo são transmitidos a cada subquadro. Conforme o subquadro, os parâmetros LP quantizados e não quantizados ou suas versões interpoladas são usados. Com base no sinal de voz ponderado perceptivamente, um período tonal em malha aberta é estimado em cada subquadro (exceto nos modos AMR_515 e AMR_475, em que a estimação é feita em cada quadro).

Consecutivamente, as operações seguintes são repetidas para cada sub-quadro:

- O sinal-alvo, $x(n)$, é computado, por meio da filtragem do resíduo da predição linear pelo filtro de síntese ponderado, $W(z)H(z)$. Previamente, os estados iniciais dos filtros são atualizados, por meio da filtragem do erro entre o resíduo da predição linear e a excitação (o que equivale à subtração, do sinal de voz ponderado, da resposta impulsional nula do filtro de síntese ponderado).
- A resposta impulsional, $h(n)$, do filtro de síntese ponderado, é calculada.
- A análise tonal em malha fechada é executada para determinação de período e ganho, usando o sinal-alvo $x(n)$ e a resposta impulsional $h(n)$, buscando-se em torno do período tonal de malha aberta. Conforme o modo, é usado período tonal fracionário com resolução de 1/6 ou 1/3 de amostra.
- O sinal-alvo $x(n)$ é atualizado, removendo-se a contribuição do dicionário adaptativo (vetor de código adaptativo filtrado). O novo alvo, $x_2(n)$, é usado na busca do dicionário algébrico fixo, para encontrar a inovação ótima.
- Os ganhos dos dicionários adaptativo e fixo são submetidos à quantização escalar com, respectivamente, 4 e 5 bits, ou vetorial, com 6-7 bits. Predição por média móvel é aplicada ao ganho do dicionário fixo.
- As memórias do filtro são atualizadas, usando o sinal de excitação encontrado, a fim de determinar o sinal-alvo no próximo subquadro.

3.4 – A ALOCAÇÃO DE BITS DOS MODOS DO CODEC AMR

A alocação de bits do AMR para cada modo é mostrada na Tabela 3.2.

Em cada quadro de voz de 20 ms são produzidos 95, 103, 118, 134, 148, 159, 204 ou 244 bits, o que corresponde às taxas de 4,75, 5,15, 5,90, 6,70, 7,40, 7,95, 10,2 ou 12,2 kbit/s.

Tabela 3.2 – AMR - Alocação de bits por modo, por parâmetro.

Modo	Parâmetro	Sub-quadro 1	Sub-quadro 2	Sub-quadro 3	Sub-quadro 4	Total por quadro
12,2 kbit/s	2 conjuntos LSP					38
	Atraso tonal	9	6	9	6	30
	Ganho tonal	4	4	4	4	16
	Código algébrico	35	35	35	35	140
	Ganho do dicionário fixo	5	5	5	5	20
	Total					244
10,2 kbit/s	Conjunto LSP					26
	Atraso tonal	8	5	8	5	26
	Código algébrico	31	31	31	31	124
	Ganhos	7	7	7	7	28
	Total					204
7,95 kbit/s	Conjunto LSP					27
	Atraso tonal	8	6	8	6	28
	Ganho tonal	4	4	4	4	16
	Código algébrico	17	17	17	17	68
	Ganho do dicionário fixo	5	5	5	5	20
	Total					159
7,40 kbit/s	Conjunto LSP					26
	Atraso tonal	8	5	8	5	26
	Código algébrico	17	17	17	17	68
	Ganhos	7	7	7	7	28
	Total					148
6,70 kbit/s	Conjunto LSP					26
	Atraso tonal	8	4	8	4	24
	Código algébrico	14	14	14	14	56
	Ganhos	7	7	7	7	28
	Total					134
5,90 kbit/s	Conjunto LSP					26
	Atraso tonal	8	4	8	4	24
	Código algébrico	11	11	11	11	44
	Ganhos	6	6	6	6	24
	Total					118
5,15 kbit/s	Conjunto LSP					23
	Atraso tonal	8	4	4	4	20
	Código algébrico	9	9	9	9	36
	Ganhos	6	6	6	6	24
	Total					103
4,75 kbit/s	Conjunto LSP					23
	Atraso tonal	8	4	4	4	20
	Código algébrico	9	9	9	9	36
	Ganhos		8		8	16
	Total					95

3.5 – DESCRIÇÃO FUNCIONAL DO CODIFICADOR AMR

As diversas funções do codificador são detalhadas a seguir. São indicadas, quando necessário, as diferenças de tratamento empregado pelos diversos modos do AMR.

3.5.1 - Pré-processamento

Duas funções de pré-processamento são aplicadas antes do processo de codificação: a redução da escala do sinal, consistindo na divisão da entrada por dois, para reduzir a possibilidade de saturação, e filtragem passa-altas, que serve como precaução contra componentes indesejáveis de baixa frequência. É usado filtro com frequência de corte de 80 Hz. As duas funções são combinadas dividindo-se por dois os coeficientes do numerador do filtro passa-altas.

3.5.2 - Predição de curto termo

Predição linear (LP) é efetuada, no modo 12,2 kbit/s, duas vezes por quadro de voz, usando técnica de auto-correlação, com janelas assimétricas de 30 ms. Ambas as análises LP são executadas no mesmo conjunto de amostras de voz. As janelas são aplicadas em 80 amostras do quadro de voz passado, em adição às 160 amostras do quadro presente. O cálculo da auto-correlação não usa amostras de quadros futuros.

Um diagrama das duas janelas LP é mostrado na Figura 3.3. A primeira janela tem seus pesos concentrados no segundo subquadro e consiste em duas metades de janelas de Hamming com diferentes tamanhos. A segunda janela tem seus pesos concentrados no quarto subquadro e consiste em duas partes: a primeira é a metade de uma janela de Hamming, e a segunda é um quarto de ciclo de uma função cosseno.

Nos demais modos, LP é efetuada uma vez por quadro de voz, usando aproximação de auto-correlação, com uma janela assimétrica de 30 ms, que possui seu peso concentrado no quarto sub-quadro e consiste de duas partes: a primeira é a metade de uma janela de Hamming e a segunda é um quarto de ciclo de uma função cosseno. O cálculo da auto-correlação usa 40 amostras futuras (5 ms).

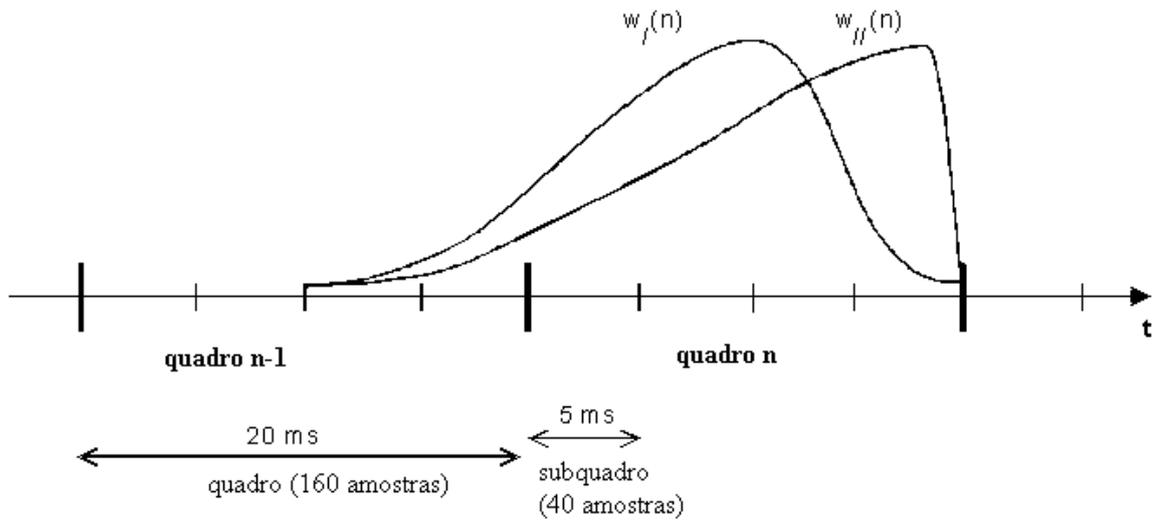


Figura 3.3 - Janelamento no modo 12.2.

As auto-correlações da voz janelada, $s'(n)$, $n=0, \dots, 239$, são calculadas por:

$$r_{ac}(k) = \sum_{n=k}^{239} s'(n)s'(n-k), \quad k = 0, \dots, 10, \quad (3.4)$$

e uma expansão de largura de banda de 60 Hz é usada por meio do janelamento periódico das auto-correlações, por meio da janela:

$$w_{lag}(i) = \exp\left[-\frac{1}{2}\left(\frac{2\pi f_0 i}{f_s}\right)^2\right], \quad i = 1, \dots, 10 \quad (3.5)$$

onde $f_0 = 60$ Hz é a expansão da largura de banda e $f_s = 8000$ Hz é a frequência de amostragem. Adicionalmente, $r_{ac}(0)$ é multiplicado pelo fator de correção de ruído branco 1.0001, o que é equivalente à adição de um limiar inferior de ruído em -40 dB. As auto-correlações são convertidas em um conjunto de dez coeficientes LP usando o algoritmo de Levinson-Durbin (Lim e Oppenheim, 1987).

Os coeficientes LP são transformados para o domínio de Pares de Linhas Espectrais (LSP) para fins de quantização e interpolação (Kabal, 1986; Itakura, 1995). O emprego desta representação é motivado por suas propriedades estatísticas, uma vez que, dentro de cada quadro de voz, há uma forte correlação em razão da propriedade de ordenamento de valores LSP adjacentes (Alencar, 2004). Isto essencialmente motiva o emprego de

quantização vetorial. Para um filtro LP de décima ordem, dados os coeficientes a_k , $k = 1, \dots, 10$, os pares de linhas espectrais (LSPs) são definidos como as raízes dos polinômios de soma e diferença:

$$F_1'(z) = A(z) + z^{-11}A(z^{-1}) \quad (3.6)$$

e

$$F_2'(z) = A(z) - z^{-11}A(z^{-1}) \quad (3.7)$$

respectivamente, sendo que $F_1'(z)$ é simétrico e $F_2'(z)$, anti-simétrico, sendo que todas as raízes destes polinômios estão sobre o círculo unitário e que se alternam. $F_1'(z)$ possui uma raiz $z = -1$ ($\omega = \pi$) e $F_2'(z)$ possui uma raiz $z = 1$ ($\omega = 0$). A fim de eliminar estas duas raízes, são definidos os novos polinômios:

$$F_1(z) = F_1'(z)/(1 + z^{-1}) \quad (3.8)$$

e

$$F_2(z) = F_2'(z)/(1 - z^{-1}) \quad (3.9)$$

cada um dos quais possui cinco raízes conjugadas sobre o círculo unitário ($e^{\pm j\omega_i}$) e podem, assim, ser escritos na forma

$$F_1(z) = \prod_{i=1,3,\dots,9} (1 - 2q_i z^{-1} + z^{-2}) \quad (3.10)$$

e

$$F_2(z) = \prod_{i=2,4,\dots,10} (1 - 2q_i z^{-1} + z^{-2}) \quad (3.11)$$

onde $q_i = \cos(\omega_i)$. Os q_i são chamados de LSPs no domínio cosseno, enquanto os ω_i representam as frequências das linhas espectrais (LSFs), e satisfazem a propriedade de ordenamento $0 < \omega_1 < \omega_2 < \dots < \omega_{10} < \pi$.

Ressonâncias no filtro LPC são monitoradas para detecção de possíveis áreas problemáticas. Divergências entre as memórias dos dicionários adaptativos do codificador e do decodificador poderiam produzir filtros instáveis em áreas com sinais contínuos altamente correlacionados. Tipicamente, erros de canal são os causadores desta divergência. O monitoramento de sinais ressonantes é realizado usando LSPs não quantizados, disponíveis após a conversão LP para LSP. O algoritmo utilizado faz uso da grande proximidade dos LSPs de um pico no espectro. Inicialmente, duas distâncias são calculadas em duas diferentes regiões. O atendimento de pelo menos uma das condições de distância classifica o quadro como de ressonância e incrementa o contador de ressonância. Doze quadros de ressonância consecutivos são necessários para indicar a possibilidade de condições problemáticas. Um flag (*LSP_flag*) é ativado ou não de acordo com esta condição.

A quantização dos LSPs (Paliwal, 1993; Xydeas, 1995) usa as técnicas de predição de média móvel (MA) de primeira ordem e quantização matricial partida (SVQ), na qual se efetua a divisão do vetor LSP original em vários subvetores de dimensões reduzidas, simplificando a complexidade da busca associada. A configuração empregada minimiza o desvio-padrão médio dentro de uma dada complexidade total. A quantização utiliza a representação LSP no domínio da frequência, ou seja:

$$f_i = \frac{f_s}{2\pi} \arccos(q_i), \quad i=1, \dots, 10, \quad (3.12)$$

onde:

f_i = frequências de linhas espectrais (LSF) em Hz [0, 4000].

f_s = 8000 Hz = frequência de amostragem.

O vetor LSF é dado por $\mathbf{f}^t = [f_1 f_2 \dots f_{10}]$, onde t indica transposição.

No modo AMR_122 há dois conjuntos de coeficientes do filtro LP por quadro. Os dois vetores LSF residuais são quantizados conjuntamente. Para os demais modos, há um conjunto de coeficientes do filtro LP por quadro. Em todos os casos, o processo de quantização usa uma medida ponderada de distorção de LSP. Assim, em geral, para um vetor LSP de entrada \mathbf{f} e um vetor quantificado no índice k , $\hat{\mathbf{f}}^k$, a quantização é executada por meio da busca do índice k que minimiza:

$$E_{LSP} = \sum_{i=1}^{10} \left[f_i w_i - \hat{f}_i^k w_i \right]^2. \quad (3.13)$$

Os pesos w_i , $i=1, \dots, 10$ são dados por:

$$\begin{aligned} w_i &= 3,347 - (1,547/450)d_i && \text{para } d_i < 450 \\ &= 1,8 - (0,8/1050)(d_i - 450) && \text{c.c.} \end{aligned} \quad (3.14)$$

onde $d_i = f_{i+1} - f_{i-1}$; $f_0 = 0$ e $f_{11} = 4000$.

No modo AMR_122, dois conjuntos de coeficientes de ponderação são computados para os dois vetores LSF. Na quantização de cada submatriz, dois coeficientes de ponderação de cada conjunto são usados com seus LSFs correspondentes.

A interpolação é executada nos LSPs, no domínio \mathbf{q} . Os vetores LSP interpolados são usados para calcular um filtro LP distinto para cada subquadro, tanto com coeficientes quantizados quanto não-quantizados, usando o método de conversão de LSP para LP previamente descrito. No modo AMR_122, os dois conjuntos de parâmetros LP, quantizados e não quantizados, são usados no segundo e quarto subquadros; o primeiro e terceiro subquadros usam interpolação linear dos parâmetros dos subquadros adjacentes. Nos demais modos, o conjunto de parâmetros LP, quantizados e não quantizados, é usado no quarto subquadro; o primeiro, segundo e terceiro subquadros usam interpolação linear dos parâmetros dos subquadros adjacentes.

Após a quantificação e interpolação dos LSPs, eles são reconvertidos para o domínio dos coeficientes LP, $\{a_k\}$, para serem utilizados na construção dos filtros de síntese e ponderação em cada subquadro. O filtro de síntese é usado no decodificador para produzir o sinal de voz reconstruído a partir do sinal de excitação recebido.

Os coeficientes dos polinômios $F_1(z)$ e $F_2(z)$ são encontrados expandindo-se as equações correspondentes, conhecendo-se os LSPs quantificados e interpolados q_i , $i = 1, \dots, 10$.

3.5.3 - Análise tonal em laço aberto

A estimação do período tonal em laço aberto baseia-se no sinal de voz ponderado, $s_w(n)$, que é obtido pela filtragem do sinal de voz de entrada pelo filtro de ponderação:

$$W(z) = A(z/\gamma_1)/A(z/\gamma_2) \quad (3.15)$$

Ou seja, em um subquadro de tamanho L , a voz ponderada é dada por:

$$s_w(n) = s(n) + \sum_{i=1}^{10} a_i \gamma_1^i s(n-i) - \sum_{i=1}^{10} a_i \gamma_2^i s_w(n-i), \quad n = 0, \dots, L-1 \quad (3.16)$$

A análise tonal usando o dicionário adaptativo modela a periodicidade de longo prazo (ou *pitch*) do sinal de voz, produzindo uma saída que vem a ser uma versão em escala do dicionário adaptativo, baseada em excitações prévias. O sinal de excitação, $u(n) = G_p u(n-\alpha) + G_c c_k(n)$, é determinado a partir de seu histórico, escalado por G_p , adicionado a c_k , que é o vetor do dicionário algébrico fixo, escalado por G_c , para cada subquadro de 5 ms. A excitação ótima é escolhida com base na minimização do erro quadrático médio e_w em cada subquadro.

O AMR emprega tanto estimação de laço aberto quanto fechado dos parâmetros de atraso do dicionário adaptativo. A estimativa em laço aberto do período tonal é utilizada para estreitar a faixa possível de valores de atraso do dicionário adaptativo e então o procedimento de análise-por-síntese, de laço fechado, é usado para encontrar um atraso de alta resolução em torno da posição aproximada de laço aberto, confinando a busca em laço fechado a um menor número de intervalos em torno dos intervalos previamente estimados. Nos modos AMR_515 e AMR_475 a análise em laço aberto é executada uma vez por quadro e determina uma estimativa do período tonal em cada quadro. Nos demais modos, é executada duas vezes por quadro e determina duas estimativas do período tonal em cada quadro.

3.5.4 - Resposta impulsional

Seja o filtro de síntese ponderado

$$H(z)W(z) = A(z/\gamma_1)/[\hat{A}(z)A(z/\gamma_2)] \quad (3.17)$$

A resposta ao impulso, $h(n)$, necessária para a busca dos dicionários adaptativo e fixo, é computada a cada subquadro filtrando-se o vetor de coeficientes do filtro $A(z/\gamma_1)$, completado com zeros, pelos dois filtros $1/\hat{A}(z)$ e $1/A(z/\gamma_2)$.

3.5.5 - Determinação do sinal-alvo

O sinal-alvo para a busca do dicionário adaptativo é usualmente computado subtraindo-se a resposta à entrada nula do filtro de síntese ponderado $H(z)W(z)$ do sinal de voz ponderado $s_w(n)$. Tal procedimento se dá a cada subquadro.

Um procedimento equivalente para o cálculo do sinal-alvo, usado no AMR, utiliza o sinal LP residual (ou resíduo LP), $res_{LP}(n)$, dado por

$$res_{LP}(n) = s(n) + \sum_{i=1}^{10} \hat{a}_i s(n-i). \quad (3.18)$$

O resíduo LP é filtrado pela combinação do filtro de síntese $1/\hat{A}(z)$ e do filtro de ponderação $A(z/\gamma_1)/A(z/\gamma_2)$. Após a determinação da excitação do subquadro, os estados iniciais destes filtros são atualizados filtrando-se a diferença entre o resíduo LP e a excitação. O sinal residual $res_{LP}(n)$, além de necessário para encontrar-se o vetor-alvo, é também usado na busca do dicionário adaptativo para alongar o *buffer* de excitação. Isto simplifica o procedimento de busca para atrasos menores que o tamanho de 40 amostras do subquadro.

3.5.6 - Dicionário adaptativo

A busca no dicionário adaptativo é efetuada a cada subquadro, consistindo na busca do período tonal em laço fechado e no consecutivo cálculo do vetor adaptativo, interpolando-se a excitação anterior no período tonal fracionário selecionado. Os parâmetros do dicionário adaptativo (ou parâmetros tonais) são o período e o ganho do filtro tonal. Na aproximação do dicionário adaptativo para implementar o filtro tonal, a excitação é repetida para atrasos menores que o comprimento do subquadro. A cada estágio da busca, a excitação é estendida pelo resíduo LP para simplificar a busca em laço fechado.

O vetor do dicionário adaptativo é calculado interpolando-se o sinal da excitação passada, $u(n)$, no atraso inteiro fornecido, k , e na fase t (fracionária):

$$v(n) = \sum_{i=0}^9 u(n-k-i) b_{60}(t+i \cdot 6) + \sum_{i=0}^9 u(n-k+1+i) b_{60}(6-t+i \cdot 6), \quad n=0, \dots, 39, \quad t=0, \dots, 5. \quad (3.19)$$

b_{60} : é o filtro de interpolação baseado em uma função $\text{sen}(x)/x$, usando janela de Hamming truncada em ± 59 e completada com zeros em ± 60 ($b_{60}(60)=0$). Tem frequência de corte (-3 dB) em 3600 Hz.

O ganho do dicionário adaptativo, g_p é dado por

$$g_p = \frac{\sum_{n=0}^{39} x(n)y(n)}{\sum_{n=0}^{39} y(n)y(n)} \quad (3.20)$$

Onde $y(n) = v(n)*h(n)$ é o vetor filtrado do dicionário adaptativo, sendo a resposta no estado zero de $H(z)W(z)$ a $v(n)$. O ganho é limitado ao intervalo $0 \leq g_p \leq 1,2$ e quantizado usando quantização escalar não uniforme de 4 bits. O ganho médio do dicionário adaptativo é calculado se *LSP_flag* estiver ativado e o ganho não quantizado do dicionário adaptativo exceder o limiar de ganho $GP_{th} = 0,95$. É calculado a partir do ganho não quantizado atual e do ganho quantizado dos sete subquadros anteriores. Se o ganho médio do dicionário adaptativo exceder GP_{th} , o ganho não quantizado é limitado ao valor de limiar e o *flag GpC_flag* é ativado para indicar a limitação. Este *flag* é usado na quantização de ganho.

3.5.7 - Dicionário fixo

Uma vez que os parâmetros do dicionário adaptativo tenham sido encontrados, o dicionário fixo é varrido levando-se em conta o agora conhecido vetor do dicionário adaptativo. Esta busca sequencial constitui um compromisso entre o desempenho ótimo e uma complexidade computacional aceitável. O dicionário fixo é varrido utilizando uma técnica de análise-por-síntese, eficiente, mas não-exaustiva, minimizando o erro quadrático médio entre a entrada de voz ponderada e a voz sintetizada ponderada.

A estrutura dos dicionários fixos é baseada no design de código conhecida como “permutação de pulsos simples intercalados” (ou ISPP, do inglês). A complexidade

computacional da busca do dicionário fixo é substancialmente reduzida quando os vetores deste dicionário, $c_k(n)$, possuem elementos, em sua maioria, iguais a zero. A estrutura algébrica da excitação que possui apenas alguns pulsos diferentes de zero permite um rápido procedimento de busca. As 40 posições em um subquadro são divididas em trilhas, cada qual podendo conter somente alguns pulsos não-nulos, de valor +1 ou -1, restritos a um número limitado de posições (Howard, 1993).

A estrutura dos dicionários fixos é especificada na Tabela 3.3 a Tabela 3.8 para cada um dos modos do AMR, que indicam os pulsos não-nulos e suas posições possíveis, em cada trilha.

No modo 12,2 kbit/s (Tabela 3.3) o dicionário de excitação contém dez posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em cinco trilhas. Cada trilha contém dois pulsos. Em cada trilha, a posição de cada pulso é codificada com três bits. O sinal do primeiro pulso é codificado com um bit. O sinal do segundo pulso é determinado por sua posição relativa ao primeiro: menor, sinal oposto; maior, mesmo sinal. Assim, são utilizados sete bits em uma trilha, ou 35 bits por subquadro.

Tabela 3.3 - Posições potenciais de pulsos no dicionário fixo - modo 12,2 kbit/s.

Trilha	Pulso	Posições
1	i_0, i_5	0, 5, 10, 15, 20, 25, 30, 35
2	i_1, i_6	1, 6, 11, 16, 21, 26, 31, 36
3	i_2, i_7	2, 7, 12, 17, 22, 27, 32, 37
4	i_3, i_8	3, 8, 13, 18, 23, 28, 33, 38
5	i_4, i_9	4, 9, 14, 19, 24, 29, 34, 39

No modo 10,2 kbit/s (Tabela 3.4), o dicionário de excitação contém oito posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em quatro trilhas. Os pulsos são agrupados em três, três e dois pulsos e suas posições são codificadas com dez, dez e sete bits, respectivamente. Em cada trilha, o sinal do primeiro pulso é codificado com um bit. Os sinais dos pulsos restantes são determinados por sua posição relativa ao primeiro, como explicado acima. Assim, são utilizados 31 bits por subquadro.

Tabela 3.4 - Posições potenciais de pulsos no dicionário fixo - modo 10,2 kbit/s.

Trilha	Pulso	Posições
1	i0, i4	0, 4, 8, 12, 16, 20, 24, 28, 32, 36
2	i1, i5	1, 5, 9, 13, 17, 21, 25, 29, 33, 37
3	i2, i6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38
4	i3, i7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39

Nos modos 7,95 kbit/s e 7,40 kbit/s (Tabela 3.5), o dicionário de excitação contém quatro posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em quatro trilhas. Cada trilha contém um pulso. As posições dos pulsos são codificadas com três, três, três e quatro bits. O sinal de cada pulso é codificado com um bit. Assim, são utilizados quatro, quatro, quatro e cinco bits nas trilhas, ou 17 bits por subquadro.

Tabela 3.5 - Posições potenciais de pulsos no dicionário fixo - modos 7.95 e 7.40 kbit/s.

Trilha	Pulso	Posições
1	i0	0, 5, 10, 15, 20, 25, 30, 35
2	i1	1, 6, 11, 16, 21, 26, 31, 36
3	i2	2, 7, 12, 17, 22, 27, 32, 37
4	i3	3, 8, 13, 18, 23, 28, 33, 38, 4, 9, 14, 19, 24, 29, 34, 39

No modo 6,70 kbit/s (Tabela 3.6), o dicionário de excitação contém três posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em três trilhas. Cada trilha contém um pulso. As posições dos pulsos são codificadas com três, quatro e quatro bits. O sinal de cada pulso é codificado com um bit. Assim, são utilizados quatro, cinco e cinco bits nas trilhas, ou 14 bits por subquadro.

Tabela 3.6 - Posições potenciais de pulsos no dicionário fixo - modo 6,70 kbit/s.

Trilha	Pulso	Posições
1	i0	0, 5, 10, 15, 20, 25, 30, 35
2	i1	1, 6, 11, 16, 21, 26, 31, 36, 3, 8, 13, 18, 23, 28, 33, 38
3	i2	2, 7, 12, 17, 22, 27, 32, 37, 4, 9, 14, 19, 24, 29, 34, 39

No modo 5,90 kbit/s (Tabela 3.7), o dicionário de excitação contém duas posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em duas trilhas. Cada trilha contém um pulso. As posições dos pulsos são codificadas com quatro e cinco bits. O sinal de cada pulso é codificado com um bit. Assim, são utilizados cinco e seis bits nas trilhas, ou 11 bits por subquadro.

Tabela 3.7 - Posições potenciais de pulsos no dicionário fixo - modo 5,90 kbit/s.

Trilha	Pulso	Posições
1	i_0	1, 6, 11, 16, 21, 26, 31, 36, 3, 8, 13, 18, 23, 28, 33, 38
2	i_1	0, 5, 10, 15, 20, 25, 30, 35, 1, 6, 11, 16, 21, 26, 31, 36, 2, 7, 12, 17, 22, 27, 32, 37, 4, 9, 14, 19, 24, 29, 34, 39

Nos modos 5,15 e 4,70 kbit/s (Tabela 3.8), o dicionário de excitação contém duas posições de pulsos não-nulos. As 40 posições em um subquadro são divididas em cinco trilhas. Cada trilha contém um pulso. Dois diferentes subconjuntos de duas trilhas são utilizadas para cada subquadro. Um bit é utilizado para codificar o subconjunto. A posição de cada pulso é codificada com três bits. O sinal de cada pulso é codificado com um bit. Assim, são utilizados nove bits por subquadro.

Tabela 3.8 - Posições potenciais de pulsos no dicionário fixo - modos 5,15 e 4,70 kbit/s.

Subquadro	Subconjunto	Pulso	Posições
1	1	i_0	0, 5, 10, 15, 20, 25, 30, 35
		i_1	2, 7, 12, 17, 22, 27, 32, 37
	2	i_0	1, 6, 11, 16, 21, 26, 31, 36
		i_1	3, 8, 13, 18, 23, 28, 33, 38
2	1	i_0	0, 5, 10, 15, 20, 25, 30, 35
		i_1	3, 8, 13, 18, 23, 28, 33, 38
	2	i_0	2, 7, 12, 17, 22, 27, 32, 37
		i_1	4, 9, 14, 19, 24, 29, 34, 39
3	1	i_0	0, 5, 10, 15, 20, 25, 30, 35
		i_1	2, 7, 12, 17, 22, 27, 32, 37
	2	i_0	1, 6, 11, 16, 21, 26, 31, 36
		i_1	4, 9, 14, 19, 24, 29, 34, 39
4	1	i_0	0, 5, 10, 15, 20, 25, 30, 35
		i_1	3, 8, 13, 18, 23, 28, 33, 38
	2	i_0	1, 6, 11, 16, 21, 26, 31, 36
		i_1	4, 9, 14, 19, 24, 29, 34, 39

Para a determinação do sinal-alvo do dicionário fixo o seguinte procedimento é executado: sendo o vetor do dicionário adaptativo filtrado:

$$y(n) = v(n)*h(n) \quad (3.21)$$

e o ganho quantizado do dicionário adaptativo \hat{g}_p , o sinal-alvo usado na busca do período tonal em malha fechada é atualizado subtraindo-se a contribuição do dicionário adaptativo:

$$x_2(n) = x(n) - \hat{g}_p y(n), \quad n = 0, \dots, 39 \quad (3.22)$$

O dicionário algébrico é pesquisado minimizando-se o erro quadrático médio entre a voz de entrada ponderada e a voz sintetizada ponderada. Se \mathbf{c}_k é o vetor de código algébrico no índice k , então o dicionário algébrico é pesquisado maximizando-se o termo:

$$A_k = \frac{(C_k)^2}{E_{Dk}} = \frac{(\mathbf{d}^t \mathbf{c}_k)^2}{\mathbf{c}_k^t \Phi \mathbf{c}_k} \quad (3.23)$$

onde:

O vetor $\mathbf{d} = \mathbf{H}^t \mathbf{x}_2$ é a correlação entre o sinal-alvo $x_2(n)$ e a resposta impulsional $h(n)$. É chamado de “alvo reversamente filtrado”.

\mathbf{H} é uma matriz triangular inferior de convolução de Toepliz, com diagonal $h(0)$ e diagonais inferiores $h(1), \dots, h(39)$.

$\Phi = \mathbf{H}^t \mathbf{H}$ é a matriz simétrica de correlações de $h(n)$.

\mathbf{d} e Φ são calculados antes da busca, sendo seus elementos dados por:

$$d(n) = \sum_{i=n}^{39} x_2(i) h(i-n), \quad n = 0, \dots, 39 \quad (3.24)$$

$$\phi(i, j) = \sum_{n=j}^{39} h(n-i) h(n-j), \quad (j \geq i) \quad (3.25)$$

A estrutura algébrica dos dicionários permite busca muito rápida, uma vez que o vetor de inovação \mathbf{c}_k contém apenas alguns pulsos não-nulos e unitários.

A correlação no numerador de A_k é dada por:

$$C = \sum_{i=0}^{N_p-1} \vartheta_i d(m_i) \quad (3.26)$$

onde:

m_i = posição do m -ésimo pulso;

ϑ_i = amplitude;

$N_p = 10$ = número de pulsos.

A energia no denominador de A_k será dada por

$$E_D = \sum_{i=0}^{N_p-1} \phi(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \vartheta_i \vartheta_j \phi(m_i, m_j). \quad (3.27)$$

Por simplificação, as amplitudes dos pulsos são ajustadas por simples quantização de um sinal apropriado. O processo de simplificação é efetuado antes da busca do dicionário.

A determinação das posições ótimas dos pulsos é então efetuada por uma técnica de busca eficiente, de análise-por-síntese, não-exaustiva, em que $A(k)$ é testado para uma pequena porcentagem das combinações de posição.

O vetor selecionado do dicionário fixo é filtrado por um pré-filtro adaptativo, $F_E(z)$, que reforça componentes espectrais especiais com a finalidade de aumentar a qualidade subjetiva da voz sintetizada. O filtro é dado por:

$$F_E(z) = 1 / (1 - \beta z^{-T}) \quad (3.28)$$

onde T é o período tonal dado pelo valor inteiro mais próximo do período fracionário em laço fechado, no subquadro. β é o ganho tonal, limitado a $[0.0, 1.0]$.

A resposta impulsional $h(n)$ é modificada para incluir o pré-filtro $F_E(z)$:

$$h(n) = h(n) - \beta h(n - T), \quad n = T, \dots, 39 \quad (3.29)$$

Finalmente, o ganho do dicionário fixo será dado por:

$$g_c = \frac{\mathbf{x}_2^t \mathbf{z}}{\mathbf{z}^t \mathbf{z}} \quad (3.30)$$

onde:

\mathbf{x}_2 é o vetor-alvo da busca do dicionário fixo;

\mathbf{z} é o vetor do dicionário fixo convoluído com $h(n)$:

$$z(n) = \sum_{i=0}^n c(i) h(n-i), \quad n=0, \dots, 39. \quad (3.31)$$

3.5.8 - Quantização do ganho dos dicionários

Para a quantização do ganho do dicionário fixo, um preditor de ganho de energia é usado de modo a explorar a correlação entre os ganhos do dicionário em quadros adjacentes. O ganho é expresso como o produto entre um ganho previsto baseado em energias prévias e um fator de correção. O fator de correção é quantizado e codificado conjuntamente com o ganho do dicionário adaptativo em todos os modos, exceto o AMR_122 e AMR_795.

3.5.9 - Atualização das memórias

Por fim, para que se possam calcular os sinais-alvo em cada subquadro, é efetuada a atualização dos estados dos filtros de síntese e ponderação.

3.6 – FLUXO DO SINAL NO DECODIFICADOR

O fluxo do sinal no decodificador é mostrado na Figura 3.4.

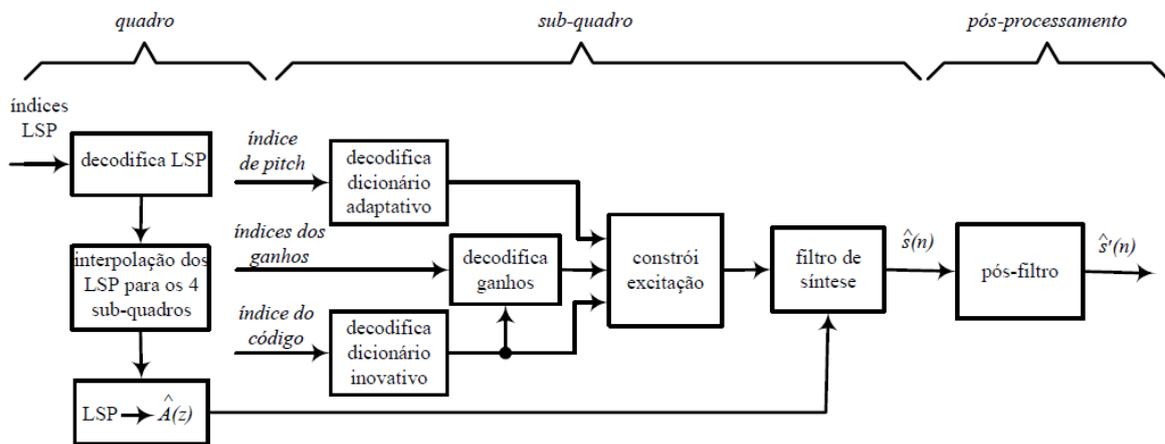


Figura 3.4 - Diagrama de blocos simplificado do decodificador adaptativo multitaxa.

No decodificador, as seguintes operações são realizadas a cada quadro:

- Com base no modo escolhido, os índices transmitidos são extraídos do fluxo de bits.

- Os índices são decodificados para obterem-se os parâmetros do codificador a cada quadro de transmissão:
 - vetores LSP
 - períodos tonais fracionários
 - vetores do dicionário inovativo
 - ganhos tonal e inovativo
- Os vetores LSP são convertidos em coeficientes do filtro LP e interpolados para a obtenção dos filtros LP em cada subquadro.

Em seguida, a cada subquadro de 40 amostras:

- A excitação é construída pela adição dos vetores dos dicionários adaptativo e inovativo, escalados pelos respectivos ganhos.
- A voz é reconstruída pela filtragem da excitação pelo filtro de síntese LP.

Por fim, o sinal de voz reconstruído passa por um pós-filtro adaptativo.

3.7 – DESCRIÇÃO FUNCIONAL DO DECODIFICADOR AMR

O decodificador realiza a decodificação dos parâmetros transmitidos (parâmetros LP e dos dicionários adaptativo e fixo) e síntese para geração da voz reconstruída, que é então pós-filtrada e reescalada.

3.7.1 - Decodificação e Síntese

Para a decodificação dos parâmetros LP, os índices LSP são usados para reconstruir os vetores LSP quantizados. A interpolação descrita previamente para o codificador é executada obtendo-se quatro vetores LSP, um para cada subquadro. Em cada subquadro, o vetor LSP interpolado é convertido para o domínio $\{a_k\}$, dos coeficientes do filtro LP.

Na decodificação do vetor do dicionário adaptativo, o índice do dicionário adaptativo é usado para determinar as partes inteira e fracionária do período tonal. O vetor do dicionário adaptativo é encontrado utilizando-se a excitação passada $u(n)$.

Na decodificação do vetor do dicionário fixo, o índice do dicionário fixo é usado para determinar as posições e sinais dos pulsos de excitação e montar o vetor algébrico $c(n)$. Se

a parte inteira do período tonal T for menor que o tamanho do subquadro (40 amostras), é executado um procedimento de aguçamento tonal, modificando $c(n)$ para

$$c(n) = c(n) + \beta c(n - T) \quad (3.32)$$

onde β é o ganho tonal decodificado, limitado a $[0 \ 1]$ ou $[0 \ 0,8]$, conforme o modo.

Para a decodificação dos ganhos dos dicionários adaptativo e fixo, no caso dos modos 12,2 e 7,95 kbit/s, os índices recebidos permitem a obtenção direta do ganho quantificado do dicionário adaptativo e o fator de correção do ganho do dicionário fixo, γ_{gc} , a partir das tabelas de quantização. Nos demais modos, o índice recebido fornece tanto o ganho quantificado do dicionário adaptativo quanto o fator de correção do ganho do dicionário fixo, γ_{gc} . O ganho quantizado do dicionário fixo é então determinado a partir do ganho estimado do dicionário fixo, g'_c .

Uma suavização adaptativa do ganho do dicionário fixo é executada para evitar flutuações de energia. Este procedimento é efetuado nos modos 10,2, 6,70, 5,90, 5,15, 4,75 kbit/s.

3.7.2 - Pós-processamento anti-esparsidade

A esparsidade dos vetores do dicionário fixo (que contém apenas alguns elementos não-nulos) pode gerar artefatos perceptivos, que podem ser reduzidos por um processamento anti-esparsidade (Hagen, 1998), consistindo na convolução circular do vetor do dicionário fixo com uma resposta impulsional pré-armazenada e selecionada adaptativamente, de três opções (modificação nula, média ou forte), a partir dos ganhos dos dicionários adaptativo e fixo.

Este procedimento é efetuado nos modos 10,2, 6,70, 5,90, 5,15 e 4,75 kbit/s.

3.7.3 - Cálculo da voz reconstruída

A excitação na entrada do filtro de síntese é dada por:

$$\hat{u}(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \quad (3.33)$$

Antes da síntese de voz, é realizado um pós-processamento nos elementos de excitação: a excitação total é modificada enfatizando-se a contribuição do vetor do dicionário adaptativo:

$$\hat{u}(n) = u(n) + \beta_1 \hat{g}_p v(n) \quad (3.34)$$

Controle adaptativo de ganho (AGC) é usado para compensar a diferença de ganho entre as excitações não enfatizada e enfatizada, $u(n)$ e $\hat{u}(n)$. O fator de escalamento de ganho para a excitação enfatizada é dado por:

$$\eta = \begin{cases} \sqrt{\frac{\sum_{n=0}^{39} u^2(n)}{\sum_{n=0}^{39} \hat{u}^2(n)}}, & \hat{g}_p > 0.5, \\ 1.0, & \hat{g}_p \leq 0.5. \end{cases} \quad (3.35)$$

O sinal de excitação enfatizado escalado é dado por:

$$\hat{u}'(n) = \hat{u}(n)\eta \quad (3.36)$$

O sinal de voz reconstruído para o subquadro é dado por:

$$\hat{s}(n) = \hat{u}'(n) - \sum_{i=1}^{10} \hat{a}_i \hat{s}(n-i), \quad n = 0, \dots, 39 \quad (3.37)$$

Onde \hat{a}_i são os coeficientes do filtro LP interpolado.

3.7.4 - Proteção contra instabilidades

Proteção adicional contra instabilidades é conseguida monitorando-se saturações no filtro de síntese. Caso ocorra saturação, a memória do dicionário adaptativo é sub-escalada por um fator de quatro e reutilizada repetitivamente.

3.7.5 - Pós-processamento

Um pós-filtro adaptativo é usado para melhorar a qualidade subjetiva da voz reconstruída. Tal filtro consiste no cascadeamento de um pós-filtro baseado em formantes e um filtro de compensação de assimetria espectral, sendo atualizado a cada subquadro. O filtro de formantes é dado por:

$$H_f(z) = \frac{\hat{A}(z/\gamma_n)}{\hat{A}(z/\gamma_d)} \quad (3.38)$$

onde $\hat{A}(z)$ é o filtro inverso de predição linear (quantificado e interpolado) e os fatores γ_n e γ_d controlam a intensidade da pós-filtragem.

O filtro $H_t(z)$ compensa assimetrias no filtro $H_f(z)$ é dado por:

$$H_t(z) = 1 - \mu z^{-1} \quad (3.39)$$

onde $\mu = \gamma_t k'_1$ é um fator de assimetria, com k'_1 sendo o primeiro coeficiente de reflexão calculado na resposta impulsional truncada ($L_h = 22$), $h_f(n)$, do filtro $H_f(z)$.

k'_1 é dado por:

$$k'_1 = \frac{r_h(1)}{r_h(0)}; \quad r_h(i) = \sum_{j=0}^{L_h-i-1} h_f(j)h_f(j+i) \quad (3.40)$$

O processo de pós-filtragem segue os seguintes passos: Filtragem inversa da voz sintetizada, $\hat{s}(n)$, por $\hat{A}(z/\gamma_n)$, produzindo o sinal residual $\hat{r}(n)$, que é filtrado pelo filtro de síntese, $1/\hat{A}(z/\gamma_d)$. O sinal na saída deste filtro é aplicado no filtro de compensação $H_t(z)$ resultando no sinal de voz pós-filtrado $\hat{s}_f(n)$.

Controle automático de ganho (AGC) é usado para compensar a diferença de ganho entre os sinais de voz sintetizado (saída do filtro de síntese) e pós-filtrado, $\hat{s}(n)$ e $\hat{s}_f(n)$. O fator de escalamento do ganho no subquadro presente é dado por:

$$\gamma_{sc} = \sqrt{\frac{\sum_{n=0}^{39} \hat{s}^2(n)}{\sum_{n=0}^{39} \hat{s}_f^2(n)}} \quad (3.41)$$

O sinal pós-filtrado escalado é dado por:

$$\hat{s}'(n) = \beta_{sc}(n) \hat{s}_f(n) \quad (3.42)$$

onde $\beta_{sc}(n)$ é atualizado a cada amostra, sendo dado por:

$$\beta_{sc}(n) = \alpha\beta_{sc}(n-1) + (1-\alpha)\gamma_{sc} \quad (3.43)$$

e o fator AGC $\alpha = 0,9$.

Por fim, um filtro passa-altas (frequência de corte de 60 Hz) funciona como precaução contra componentes indesejáveis de baixa frequência. Aumento de escala é efetuado multiplicando-se a voz por um fator de 2, para compensar a divisão por dois efetuada na entrada do codificador.

3.8 – IMPLEMENTAÇÃO EM SOFTWARE

O código-fonte do codec AMR é disponibilizado em C ANSI, em duas versões: ponto fixo (TS 26.073) e ponto flutuante (TS 26.104). C ANSI foi escolhido como padrão em função da necessidade de garantia de portabilidade.

Segundo a ETSI, a especificação em ponto fixo é a única implementação autorizada para os serviços mandatórios de voz em 3G; deste modo, o uso do codec de ponto flutuante é estritamente limitado a outros serviços, sendo direcionada para aplicações multimídia ou baseadas em pacote. A implementação em ponto fixo é preferencial para todas as aplicações, embora a implementação em ponto flutuante possa ser utilizada quando a plataforma de for mais adequada para ela.

O codificador de ponto flutuante é uma implementação não exata em nível de bit do codificador de ponto fixo. Em termos de qualidade de saída, entretanto, são indistinguíveis. O decodificador de ponto flutuante é funcionalmente uma implementação exata em nível de bit do codificador de ponto fixo. O código foi otimizado para velocidade e não são usadas as bibliotecas padrão de ponto fixo.

A especificação em ponto fixo é necessária para uma implementação exata do transcodificador AMR e das funções de VAD, ruído de conforto, operação de taxa controlada pela fonte e tratamento de quadros perdidos. O código C de ponto fixo foi previamente testado em sistemas *Unix* e *Linux* e o compilador *GNU gcc*, sendo fornecido com arquivos de *make* compatíveis com cada uma destas configurações.

A partir do código-fonte são gerados dois arquivos executáveis — *encoder* e *decoder* —, correspondentes ao codificador e ao decodificador AMR, respectivamente. Em tempo de compilação é possível escolher entre duas modalidades de VAD (ENS ou *Motorola*). Para a verificação dos executáveis são fornecidos conjunto de arquivos de teste para sistemas *little-endian* (byte LSB primeiro, tal como Linux em um *PC*) e *big-endian* (byte MSB primeiro, tal como servidores *Sun*), respectivamente, para cada uma das opções de VAD, consistindo cada conjunto de um arquivo de entrada do codificador, um arquivo de comparação para a saída do codificador (e entrada do decodificador) e um arquivo de comparação para a saída do decodificador.

Foi verificado que os codecs de ponto fixo e flutuante interoperam sem nenhum artefato. A estrutura básica dos códigos é a mesma.

3.9 – FORMATOS DE ARQUIVO

Para a testagem do codec AMR em sua configuração original é necessária a utilização de sinais de voz compatíveis com seu formato de entrada – sinais amostrados à taxa de 8 kHz, armazenados formato PCM *signed* de 16 bits.

O consórcio 3GPP oferece arquivos com tais características para download a partir de seu *website* (www.3gpp.com), com a finalidade de utilização como sequências de teste para o codec AMR. Mais especificamente, para a verificação da exatidão da implementação do *transcoder* (TS 26.090), detecção de atividade vocal (TS 26.094), ruído de conforto (TS 26.092), e operação controlada pela fonte (TS 26.093), bem como da instalação do código ANSI C (TS 26.073). O código ANSI C (TS 26.073) vem também acompanhado de um conjunto de arquivos de teste. No caso presente, os sinais de voz acima referidos serão também utilizados na verificação do codec AMR de ponto flutuante.

Os arquivos de sequências de teste são fornecidos para PC – com ordem de byte *little-endian*.

Cada conjunto de arquivos compõe-se dos tipos dados pela Tabela 3.9.

A seguir, é feita a descrição dos formatos de arquivo usados pelo codificador e pelo decodificador. As sequências de teste igualmente usam os formatos de arquivo descritos aqui.

Tabela 3.9 – tipos de arquivo do AMR.

<i>Tipo de Arquivo</i>	<i>Extensão</i>
Entrada do codificador de voz	INP
Saída do codificador e entrada para o decodificador de voz	COD
Saída do decodificador de voz	OUT
Controle de chaveamento dos modos (opcional)	MOD

Arquivo de voz (entrada do codificador e saída do decodificador): Arquivos de voz lidos pelo codificador e escritos pelo decodificador consistem em palavras de 16 bits, cada uma contendo uma amostra de voz de 13 bits de resolução, alinhada a esquerda. A ordem dos bytes depende da arquitetura empregada (p. ex.: MSB primeiro em estações *Sun*, LSB primeiro em PCs). Tanto o codificador quanto o decodificador processam apenas quadros completos de 160 amostras. Deste modo, o codificador irá processar apenas n quadros se o comprimento do arquivo de entrada for de $n*160 + k$ palavras, e os arquivos produzidos pelo decodificador sempre terão um comprimento de $n*160$ palavras.

Arquivo de controle de modo (entrada do codificador): O codificador pode opcionalmente ler um arquivo de controle de modo que especifica o modo de codificação para cada quadro de voz processado. Este é um arquivo texto contendo uma linha por quadro de voz. Cada linha contém o nome de um dos modos da lista $\{MR475, MR515, MR59, MR67, MR74, MR795, MR102, MR122\}$.

Arquivo de fluxo de bits de parâmetros (saída do codificador e entrada do decodificador): O arquivo de parâmetros pode ser gerado em vários formatos, conforme a opção utilizada na compilação do codificador.

Formato ETSI:

Nesse formato, cada bit do fluxo de bits é representado por meio de um byte completo. Este formato é adotado como padrão pela versão em ponto fixo do codec. Os arquivos

produzidos pelo codificador e esperados pelo decodificador contém um número arbitrário de quadros, no formato dado pela Tabela 3.10.

Tabela 3.10 - Formato do quadro do arquivo de saída do codificador.

FRAME_TYPE	B1	B2	...	B244	MODE_INFO	unused1	...	unused4
------------	----	----	-----	------	-----------	---------	-----	---------

Cada célula corresponde a um valor de 16 bits (Word16) no arquivo de fluxo de bits, em um total de 250 palavras ou 500 bytes por quadro. O detalhamento de cada campo é dado na Tabela 3.11.

Tabela 3.11 – Quadro codificado no formato ETSI.

Campo	Descrição	Valores literais	Valores numéricos	Observações
FRAME_TYPE	Tipo do quadro transmitido	TX_SPEECH	0x0000	Quadro normal de voz
		TX_SID_FIRST	0x0001	Primeiro quadro SID após voz
		TX_SID_UPDATE	0x0002	Quadro de atualização de SID no DTX
		TX_NO_DATA	0x0003	Quadro vazio no DTX
B1...B244	Bits de parâmetros do codificador de voz (isto é, o fluxo de bits propriamente dito)		0x0000	Somente o modo AMR_122 efetivamente usa todos os 244 bits. Para os demais modos, apenas os primeiros n bits são usados ($35 \leq n \leq 204$). Os demais são escritos como 0x0000.
			0x0001	
MODE_INFO	Informação do modo de codificação	MR475	0x0000	4.75 kbit/s
		MR515	0x0001	5.15 kbit/s
		MR59	0x0002	5.90 kbit/s
		MR67	0x0003	6.70 kbit/s
		MR74	0x0004	7.40 kbit/s
		MR795	0x0005	7.95 kbit/s
		MR102	0x0006	10.20 kbit/s
		MR122	0x0007	12.20 kbit/s
unused1...4	Não usado		0x0000	

Como observado anteriormente, a ordem dos bytes depende da arquitetura empregada.

Se invocado sem a opção `-dtx`, o codificador produz apenas quadros do tipo `0x0000` (`TX_SPEECH`).

Formato MIME:

O arquivo de parâmetros pode ser formatado de acordo com o formato MIME usado pelo serviço MMS, conforme especificado na RFC 3267 – “*Real-Time Transport Protocol (RTP) Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs*”. O uso deste formato não tem efeitos nos algoritmos de codificação e decodificação. Apenas o formato do fluxo de bits é afetado. Entretanto, não é possível utilizar os arquivos de teste com este formato. Este formato é adotado como padrão pela versão em ponto flutuante do codec.

Formato IF2:

Disponível apenas para o codec de ponto flutuante. Compatível com o formato AMR IF2 constante da especificação 3GPP TS 26.101 “*Mandatory speech processing functions; AMR speech codec frame structure*”.

4 – VERSÃO DO AMR-ACELP ADAPTADA PARA SINAIS EMG

Utilizando-se o código-fonte do codec AMR disponibilizado pelo 3GPP, várias alterações foram sucessivamente aplicadas ao codec original, com o objetivo de melhorar objetivamente seu desempenho com sinais de EMG. Tais modificações deram origem a diversos novos modos de operação (empregando-se o termo “modo” com uma acepção equivalente à da documentação do AMR). Tais alterações e as características destes modos são apresentadas a seguir, assim como considerações sobre procedimentos adicionais que poderão vir a ser implementados. Nas alterações, foi utilizado a versão de ponto flutuante do código-fonte e o formato ETSI do arquivo codificado, descrito anteriormente.

4.1 – ESTRATÉGIA PARA IMPLEMENTAÇÃO DE NOVOS MODOS

A partir do codec AMR original, diversos novos modos de operação foram criados por meio de combinações de dicionários fixos e de conjuntos de parâmetros do preditor de curto prazo e da utilização ou não do dicionário adaptativo.

Dentre os diversos modos do AMR (ver Tabela 3.1), os modos AMR_122 e AMR_795 foram selecionados como base a criação dos novos modos, tendo em vista que o modo AMR_122 propicia a maior qualidade para o sinal reconstruído, seguido pelos modos AMR_102 e AMR_795. Entretanto, a estrutura deste último é a que mais se aproxima do modo AMR_122, tornando mais apropriado o processo de combinação de módulos. Acrescente-se que estes dois modos, AMR_122 e AMR_795, são os únicos que codificam separadamente os ganhos dos dicionários fixo e adaptativo, característica que facilita a sua modificação.

4.2 – MODIFICAÇÕES GERAIS

Foram efetuadas algumas alterações gerais, válidas em todos os novos modos, incluindo:

- aumento dos *buffers* correspondentes ao número máximo de parâmetros codificados e ao número de bytes por quadro do arquivo codificado em formato ETSI, de forma a poder acomodar um número maior de parâmetros codificados, decorrentes das modificações efetuadas;
- eliminação da rotina de anulação dos três bits menos significativos do arquivo de entrada contendo o sinal original;

- eliminação da filtragem passa-altas (subtópico 3.5.1) no pré-processamento.
- a divisão por 2 no codificador, originalmente associada ao filtro passa-altas, acima mencionado, e a correspondente multiplicação no decodificador foram mantidas.

4.3 – ALTERAÇÃO DO CRITÉRIO BÁSICO DE OTIMIZAÇÃO

Conforme apresentado anteriormente (eq. 3.3), o critério original de otimização de escolha dos parâmetros dos dicionários, utiliza um filtro de enriquecimento da qualidade subjetivamente percebida da voz, qual seja:

$$W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (4.1)$$

No codificador, o filtro $W(z)$ é utilizado na determinação do período tonal em laço aberto, efetuada a cada dois subquadros (subtópico 3.5.3) e na busca em laço fechado dos parâmetros dos dicionários adaptativo e fixo a cada subquadro (subtópicos 3.5.6 e 3.5.7). No decodificador, $W(z)$ é aplicada no pós-processamento do sinal reconstruído.

A fim de eliminar o efeito deste filtro, $W(z)$, os valores de γ_1 e γ_2 foram alterados para

$$\gamma_1 = \gamma_2 = 0 \quad (4.2)$$

Pela equação (3.16), a voz ponderada é dada por:

$$s_w(n) = s(n) + \sum_{i=1}^{10} a_i \gamma_1^i s(n-i) - \sum_{i=1}^{10} a_i \gamma_2^i s_w(n-i), \quad n = 0, \dots, L-1 \quad (4.3)$$

Assim, se $\gamma_1 = \gamma_2 = 0$, resulta que:

$$s_w(n) = s(n) \quad (4.4)$$

Com isso, o critério de otimização no processo de análise-por-síntese passa a basear-se diretamente na minimização do erro $e(n)$ (Figura 3.1):

$$e(n) = s(n) - \hat{s}(n) \quad (4.5)$$

Esta alteração é efetiva para todos os novos modos.

4.4 – RETIRADA DE PROCEDIMENTOS PARA MELHORIA SUBJETIVA DA VOZ

Foram eliminadas do codec original, além de $W(z)$, outras rotinas que tem por objetivo aumentar a qualidade subjetivamente avaliada de sinais de voz à custa da qualidade objetivamente mensurada do sinal. Estas alterações foram efetivas para todos os novos modos.

Especificamente, foi eliminado no codificador o pré-filtro adaptativo (subtópico 3.5.7, equações 3.28 e 3.29), que introduz alterações no dicionário fixo. No decodificador, foram eliminados:

- o procedimento de aguçamento tonal (subtópico 3.7.1, equação 3.32), que reforça componentes de frequência na vizinhança das formantes;
- a rotina de suavização do ganho do dicionário fixo (subtópico 3.7.1), que evita flutuações bruscas da intensidade do sinal entre quadros de voz;
- a rotina de dispersão de fase (subtópico 3.7.2), processamento anti-esparsidade dos vetores do dicionário fixo;
- o pós-processamento nos elementos de excitação (subtópico 3.7.3, equação 3.34), que enfatiza a contribuição do dicionário adaptativo na voz sintetizada;
- a filtragem passa-altas (subtópico 3.7.5).

4.5 – RETIRADA DO DICIONÁRIO ADAPTATIVO

O dicionário adaptativo é particularmente eficiente na modelagem de características periódicas do sinal, tais como as encontradas em trechos sonoros da voz, mas normalmente ausentes no sinal eletromiográfico, o qual pode ser mais eficientemente modelado com o emprego de dicionários fixos, que se mostram particularmente eficientes na parametrização de trechos surdos da voz. Conforme discutido no Capítulo 2, sinais de EMG possuem características em comum com esta modalidade de sinais de voz.

A estratégia utilizada na eliminação do dicionário adaptativo consistiu em tornar nulos o ganho e a sequência de excitação na saída deste estágio.

Em dois dos novos modos o dicionário adaptativo foi mantido. Conforme detalhado no capítulo seguinte, nos novos modos em que foi retirado este dicionário, verificou-se uma

melhoria do desempenho, ou seja, a diminuição das medidas objetivas de erro e, obviamente, o aumento da taxa de compressão.

4.6 – UTILIZAÇÃO DE DOIS DICIONÁRIOS FIXOS

Em alguns dos casos em que foi efetuada a retirada do dicionário adaptativo, foi mantido o primeiro dicionário fixo e um segundo dicionário fixo foi acrescentado. Em outros casos houve a substituição do primeiro concomitantemente à inclusão do novo dicionário.

Em todos estes novos modos foi mantido o mesmo esquema de busca sub-ótima dos vetores de cada dicionário, conforme descrito no tópico 3.1. Como sinal-alvo para este segundo dicionário fixo utilizou-se o sinal-alvo para o primeiro dicionário após a subtração da contribuição desse dicionário.

4.7 – RESUMO DOS NOVOS MODOS

A Tabela 4.1 resume as características dos novos modos desenvolvidos.

Tabela 4.1 - Propriedades dos novos modos.

Modo	2LP/ DA/ DF10	1LP/ DA/ DF10	2LP/ DF10	1LP/ DF10	1LP/ DF10/ DF4	1LP/ DF10/ DF10	1LP/ DF4/ DF4	2LP/ DF10/ DF10
bits por quadro	244	231	198	187	275	347	203	358
taxa de bits (kbit/s)	12,2	11,55	9,9	9,35	13,75	17,35	10,15	17,9
taxa de compressão (%)	87,29	87,97	89,69	90,26	85,68	81,93	89,43	81,35

A designação dos novos modos é feita pela identificação das características de seus principais módulos. Assim, o primeiro termo refere-se à utilização de um ou dois conjuntos de parâmetros LP (2LP, conforme originalmente empregado no modo AMR_122, ou 1LP, nos demais modos). A utilização do dicionário adaptativo (quando houver) é indicada em seguida por DA. Finalmente, os dicionários fixos são identificados pela quantidade de pulsos não-nulos: DF10, no caso do dicionário empregado originalmente no modo AMR_122, ou DF4, no modo AMR_795 (a forma de identificação dos modos originais,

por outro lado, ao indicar a taxa de transmissão, reflete a necessidade de utilização controlada da banda disponível).

A Figura 4.1 esquematiza as relações entre os modos originais do AMR e os novos modos desenvolvidos.

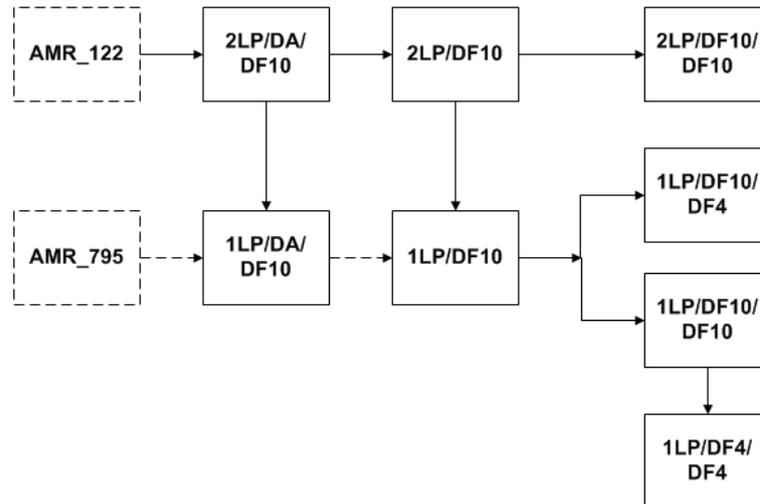


Figura 4.1 – Relações entre os modos originais do AMR e os novos modos.

Assim, o novo modo 2LP/DA/DF10 equivale ao modo original AMR_122, inclusive na quantidade de bits e, conseqüentemente, na taxa de compressão, diferenciando-se pela eliminação das rotinas de enriquecimento da qualidade subjetiva, enquanto que, no modo 2LP/DF10, o dicionário adaptativo foi também excluído.

O modo 1LP/DA/DF10 difere do modo 2LP/DA/DF10 por utilizar apenas um conjunto de parâmetros LP, mantendo o mesmo dicionário fixo. O modo 1LP/DF10, por sua vez, difere do modo 2LP/DF10 por utilizar apenas um conjunto de parâmetros LP, mantendo o mesmo dicionário fixo.

Conforme mencionado, foram criados novos modos com a inclusão de um segundo dicionário fixo, resultando em uma estrutura alternativa a dos modos originais do AMR, formada pela associação de um dicionário adaptativo e um fixo. Assim, a partir do modo 1LP/DF10 foram criados dois outros — 1LP/DF10/DF4 e 1LP/DF10/DF10 — e, a partir do modo 2LP/DF10, foi criado o modo 2LP/DF10/DF10. Por fim, o modo 1LP/DF4/DF4 origina-se da substituição, no modo 1LP/DF10/DF10, de ambos os dicionários fixos por outros de menor extensão.

4.8 – CONSIDERAÇÕES SOBRE OUTROS PROCEDIMENTOS

Melhorias adicionais poderão ser possivelmente alcançadas com o emprego de outras alterações ou de recursos específicos do codec AMR, tais como a escolha dinâmica do modo (quadro a quadro), e a inserção de quadros SID de taxa mínima de bits resultante da detecção (com o recurso DTX) de períodos de inatividade eletrofisiológica (equivalente a do silêncio na fala).

Inovações mais profundas nos algoritmos empregados envolvem a criação de dicionários fixos de maior extensão, e conseqüentemente com maior número de pulsos não-nulos, uma vez que foi verificado que dicionários de maior extensão produziram menores erros. Além disso, um dicionário de maior extensão utilizando apenas um ganho economizará bits em comparação a dois dicionários menores empregando dois ganhos distintos. Novas combinações (ordenação) de dicionários podem também ser exploradas.

Algumas características do AMR permaneceram otimizadas para voz: o dicionário de quantificação de LSP, a ordem do filtro de predição linear (10) e o tamanho, em amostras, das janelas de processamento (160 amostras). Alterações destas características exigem uma maior massa de dados de EMG para geração do dicionário, e modificações mais profundas na estrutura do código-fonte para utilização de diferentes dimensões do filtro e das janelas.

Deste modo, poderá ser efetuada a criação de dicionário de quantização de LSP específico para sinais de EMG, a alteração da ordem do filtro mencionado, e a mudança do tamanho das janelas.

5 – SIMULAÇÕES E RESULTADOS

As novas configurações obtidas a partir das modificações implementadas nos codecs, descritas no capítulo anterior, foram testadas com um grupo de sinais de eletromiografia, bem como com um sinal de referência de voz. São apresentados, a seguir, além de detalhes da implementação do *software*, a descrição dos sinais e as medidas objetivas utilizadas para avaliar o desempenho dos codificadores simulados. Finalmente, são apresentados os resultados obtidos com estas simulações e são efetuadas comparações com trabalhos prévios que utilizaram ou o CELP ou outras técnicas de compressão de sinais.

5.1 – IMPLEMENTAÇÃO

No desenvolvimento deste trabalho foi utilizado o código-fonte do codec AMR com variáveis de ponto flutuante (especificação técnica 3GPP TS 26.104), versão 7.0.0 (junho de 2007). A escolha da versão de ponto flutuante do codec se deu em função do aumento da faixa dinâmica dos sinais que podem ser representados neste formato, reduzindo a possibilidade de ocorrência de saturações, a que a opção de código com ponto fixo estaria mais sujeita.

A compilação do código-fonte foi efetuada no sistema operacional *Windows XP*, no ambiente integrado de desenvolvimento *Microsoft Visual C++ 2008*. Em tempo de compilação foi determinada a geração do arquivo codificado em formato ETSI, em que cada bit é representado em dois bytes, conforme detalhado na Tabela 3.11. Em tempo de execução foi determinada a não ativação do recurso DTX, de transmissão descontínua.

5.2 – SINAIS UTILIZADOS NAS SIMULAÇÕES

Os dados em teste consistiram em um grupo de doze sinais eletromiográficos, empregados em estudo descrito no artigo de Berger *et al* (2006), produzidos durante a contração isométrica do músculo *biceps brachii*, a 60% da contração máxima voluntária. Tais sinais foram coletados com eletrodos de superfície pré-amplificados, sendo amostrados a uma taxa de 2000 amostras/s e quantizados com resolução de 12 bits. A eliminação de trechos sem atividade muscular no início e no final de cada captura resultou em sinais com duração entre 61 e 110 segundos. Cada um dos sinais eletromiográficos foi também submetido a

filtragem passa-faixas com filtro digital FIR de fase linear, com frequências de corte inferior e superior de 20 Hz e 500 Hz, respectivamente.

Por sua vez, o sinal de voz, utilizado para comparação do desempenho dos codecs, com duração de três segundos, foi amostrado à taxa de 8000 Hz e quantizado com resolução de 16 bits.

Em todos os casos, os sinais foram previamente normalizados pelo valor máximo absoluto, de modo a ocuparem ao máximo a faixa de valores $[-32767 \text{ } +32767]$ e gravados em formato PCM de 16 bits, compatível com o formato de arquivo de entrada dos codecs.

5.3 – MEDIDAS DE DESEMPENHO

O desempenho dos codecs foi mensurado objetivamente por meio de três critérios usualmente empregados na avaliação da compressão de sinais eletrofisiológicos (Berger, 2006, Carotti, 2007): fator de compressão (*compression factor, CF*), diferença percentual residual (Cetin, 2000) (*percent residual difference, PRD*) e erro quadrático médio normalizado (*mean square error, MSE*).

O fator de compressão é definido por:

$$CF = \frac{N_{orig} - N_{comp}}{N_{orig}} * 100\% \quad (5.1)$$

onde N_{orig} é a quantidade de bits necessária para armazenar os dados originais e N_{comp} é a quantidade de bits necessária para armazenar os dados comprimidos.

Considerando-se quadros de 160 amostras com resolução original de 12 bits, $N_{orig} = 1920$ bits/quadro.

A diferença residual percentual e o erro quadrático médio percentual são definidos, respectivamente, nas equações seguintes:

$$PRD = \sqrt{\frac{\sum_{n=0}^{N-1} (x[n] - \hat{x}[n])^2}{\sum_{n=0}^{N-1} (x[n])^2}} * 100\% \quad (5.2)$$

$$MSE = \frac{\sum_{n=0}^{N-1} (x[n] - \hat{x}[n])^2}{\sum_{n=0}^{N-1} (x[n])^2} * 100\% \quad (5.3)$$

onde x é o sinal original, \hat{x} é o sinal reconstruído e N o número de amostras do sinal.

Uma vez que MSE e PRD não são independentes, as análises de desempenho se darão mais extensivamente apenas sobre este último.

5.4 – DESEMPENHO DO CODIFICADOR AMR-ACELP ORIGINAL

Os sinais de EMG e voz foram também processados utilizando-se o modo AMR_122 original do AMR, conforme resultados mostrados na Tabela 5.1. O erro atinge valores significativamente altos para o sinal de voz e ainda mais para os de eletromiografia, refletindo as distorções sofridas pela forma de onda do sinal reconstruído em razão do tratamento efetuado pelo codec original para garantir o aumento da qualidade subjetiva de voz.

Tabela 5.1 – Desempenho do modo original AMR_122 (CF = 87,3%) para sinais de eletromiografia e de voz.

Sinais	PRD (%)
EMG1	98,39
EMG2	87,91
EMG3	94,96
EMG4	105,78
EMG5	99,00
EMG6	105,96
EMG7	90,69
EMG8	108,59
EMG9	87,36
EMG10	95,41
EMG11	95,21
EMG12	91,23
<i>Média</i>	<i>96,71</i>
<i>Desvio Padrão</i>	<i>7,10</i>
Voz1	78,32

A Figura 5.1 e a Figura 5.2, respectivamente, apresentam comparações de trechos de sinais de EMG e voz, originais e reconstruídos após codificação com o modo AMR_122.

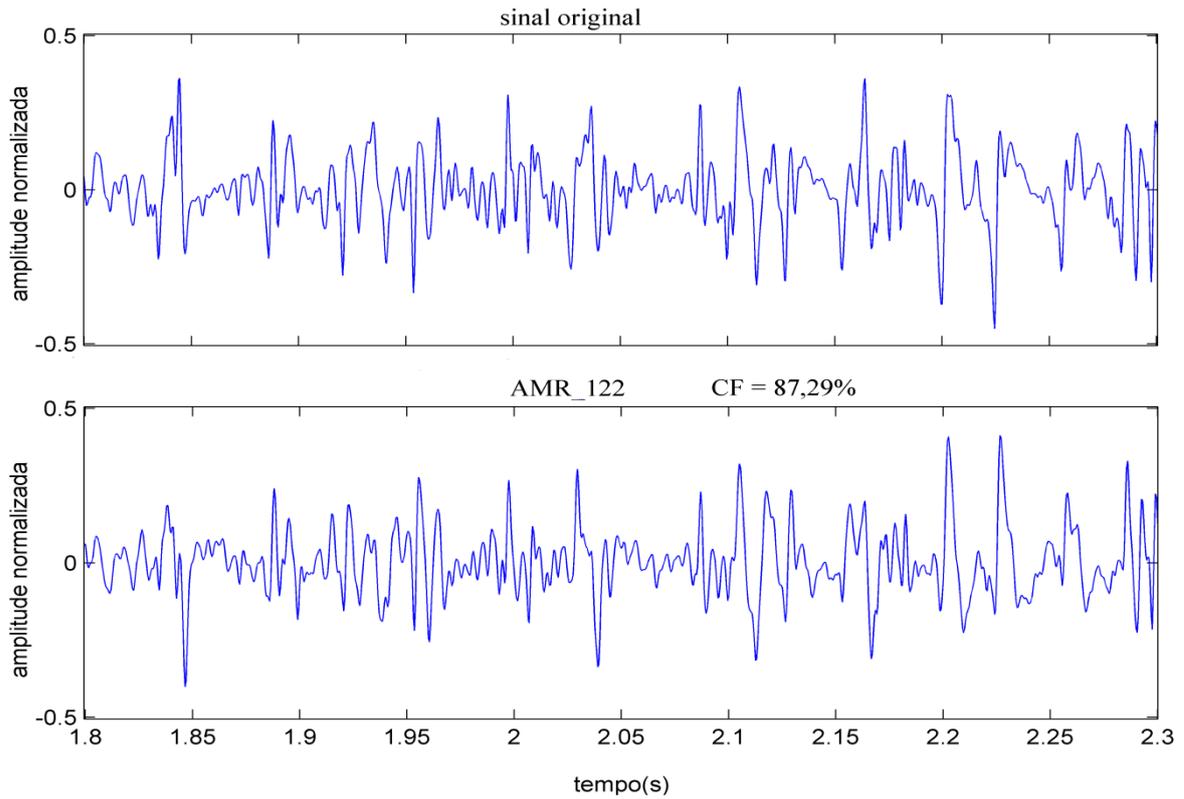


Figura 5.1 – Sinal de EMG, original e reconstruído após compressão no modo AMR_122.

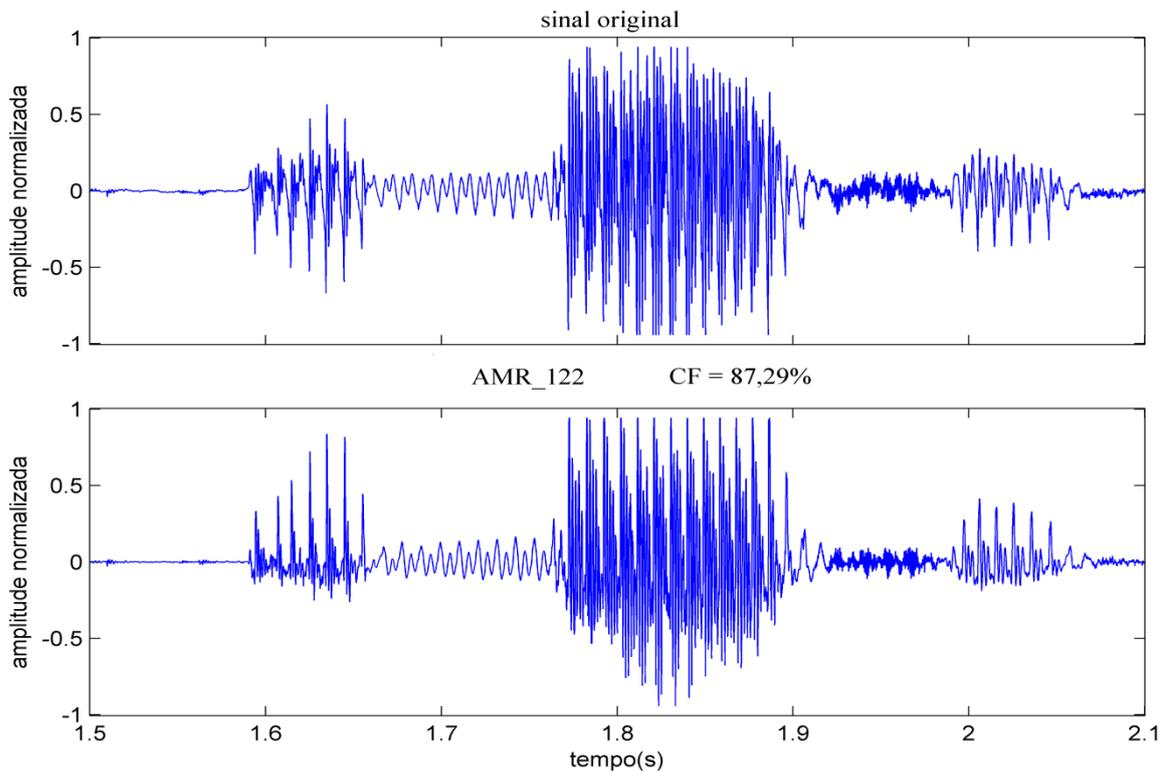


Figura 5.2 - Sinal de voz, original e reconstruído após compressão no modo AMR_122.

5.5 – DESEMPENHO DO CODIFICADOR AMR-ACELP MODIFICADO

Os sinais eletromiográficos foram processados utilizando-se cada um dos novos modos de codificação descritos anteriormente, sendo realizadas, sequencialmente, a compressão (codificação) e a reconstrução (decodificação) de cada sinal. Os sinais reconstruídos foram comparados aos originais por meio da determinação de PRD e MSE.

Os resultados da compressão e reconstrução dos sinais de EMG, tanto para cada sinal quanto consolidados (média e desvio padrão), para cada modo, são mostrados na Tabela 5.2 (PRD) e na Tabela 5.3 (MSE).

Procedimento semelhante foi empregado com o sinal de voz, tendo sido obtidos os resultados mostrados na Tabela 5.4 (PRD) e na Tabela 5.5 (MSE).

Tabela 5.2 – Desempenho (valores de PRD) dos novos modos de codificação para sinais de EMG.

Sinal	Modo							
	2LP/ DA/ DF10	1LP/ DA/ DF10	2LP/ DF10	1LP/ DF10	1LP/ DF10/ DF4	1LP/ DF10/ DF10	1LP/ DF4/ DF4	2LP/ DF10/ DF10
EMG1	23,92	25,86	21,05	21,66	17,30	14,96	25,31	13,72
EMG2	24,27	26,12	17,38	18,27	16,04	12,72	27,64	12,60
EMG3	24,48	26,36	21,02	21,85	16,31	15,11	24,44	13,73
EMG4	21,83	23,99	16,96	17,72	18,73	12,72	24,21	13,38
EMG5	21,67	23,55	15,39	15,89	18,45	11,34	27,02	11,96
EMG6	20,10	22,47	14,19	14,86	13,04	10,67	25,00	11,66
EMG7	24,18	25,38	21,84	22,13	19,54	15,50	26,05	14,21
EMG8	24,87	26,63	23,51	23,20	18,98	16,42	27,97	15,04
EMG9	22,78	24,09	19,05	19,46	15,85	13,49	23,07	12,60
EMG10	22,84	24,44	18,09	18,63	14,38	13,03	22,29	12,53
EMG11	21,82	23,40	16,19	16,58	15,11	11,62	25,82	12,08
EMG12	23,52	24,63	17,17	17,80	17,21	12,54	23,19	12,27
<i>Média</i>	<i>23,02</i>	<i>24,74</i>	<i>18,49</i>	<i>19,00</i>	<i>16,75</i>	<i>13,34</i>	<i>25,17</i>	<i>12,98</i>
<i>Desvio Padrão</i>	<i>1,44</i>	<i>1,32</i>	<i>2,84</i>	<i>2,68</i>	<i>1,99</i>	<i>1,80</i>	<i>1,83</i>	<i>1,03</i>
CF (%)	87,29	87,97	89,69	90,26	85,68	81,93	89,43	81,35

Tabela 5.3 - Desempenho (valores de MSE) dos novos modos de codificação para sinais de EMG.

Sinal	Modo							
	2LP/ DA/ DF10	1LP/ DA/ DF10	2LP/ DF10	1LP/ DF10	1LP/ DF10/ DF4	1LP/ DF10/ DF10	1LP/ DF4/ DF4	2LP/ DF10/ DF10
EMG1	5,72	6,69	4,43	4,69	2,99	2,24	6,40	1,88
EMG2	5,89	6,82	3,02	3,34	2,57	1,62	7,64	1,59
EMG3	5,99	6,95	4,42	4,77	2,66	2,28	5,97	1,88
EMG4	4,76	5,76	2,88	3,14	3,51	1,62	5,86	1,79
EMG5	4,70	5,54	2,37	2,53	3,41	1,28	7,30	1,43
EMG6	4,04	5,05	2,01	2,21	1,70	1,14	6,25	1,36
EMG7	5,85	6,44	4,77	4,90	3,82	2,40	6,78	2,02
EMG8	6,19	7,09	5,53	5,38	3,60	2,70	7,83	2,26
EMG9	5,19	5,80	3,63	3,79	2,51	1,82	5,32	1,59
EMG10	5,22	5,97	3,27	3,47	2,07	1,70	4,97	1,57
EMG11	4,76	5,48	2,62	2,75	2,28	1,35	6,67	1,46
EMG12	5,53	6,07	2,95	3,17	2,96	1,57	5,38	1,51
Média	5,32	6,14	3,49	3,68	2,84	1,81	6,36	1,70
Desvio Padrão	0,65	0,65	1,07	1,03	0,66	0,49	0,92	0,27
CF (%)	87,29	87,97	89,69	90,26	85,68	81,93	89,43	81,35

Tabela 5.4 –Desempenho (valores de PRD) dos novos modos de codificação para sinal de VOZ.

Sinal	Modo							
	2LP/ DA/ DF10	1LP/ DA/ DF10	2LP/ DF10	1LP/ DF10	1LP/ DF10/ DF4	1LP/ DF10/ DF10	1LP/ DF4/ DF4	2LP/ DF10/ DF10
Voz1	14,61	15,07	20,82	21,56	17,71	17,78	36,37	18,94
CF (%)	87,29	87,97	89,69	90,26	85,68	81,93	89,43	81,35

Tabela 5.5 - Desempenho (valores de MSE) dos novos modos de codificação para sinal de VOZ.

Sinal	Modo							
	2LP/ DA/ DF10	1LP/ DA/ DF10	2LP/ DF10	1LP/ DF10	1LP/ DF10/ DF4	1LP/ DF10/ DF10	1LP/ DF4/ DF4	2LP/ DF10/ DF10
Voz1	2,14	2,27	4,33	4,65	3,14	3,16	13,23	3,59
CF (%)	87,29	87,97	89,69	90,26	85,68	81,93	89,43	81,35

As formas de onda relativas ao sinal de EMG original e aos reconstruídos a partir da codificação com o modo AMR_122 e os novos modos podem ser vistas na Figura 5.3.

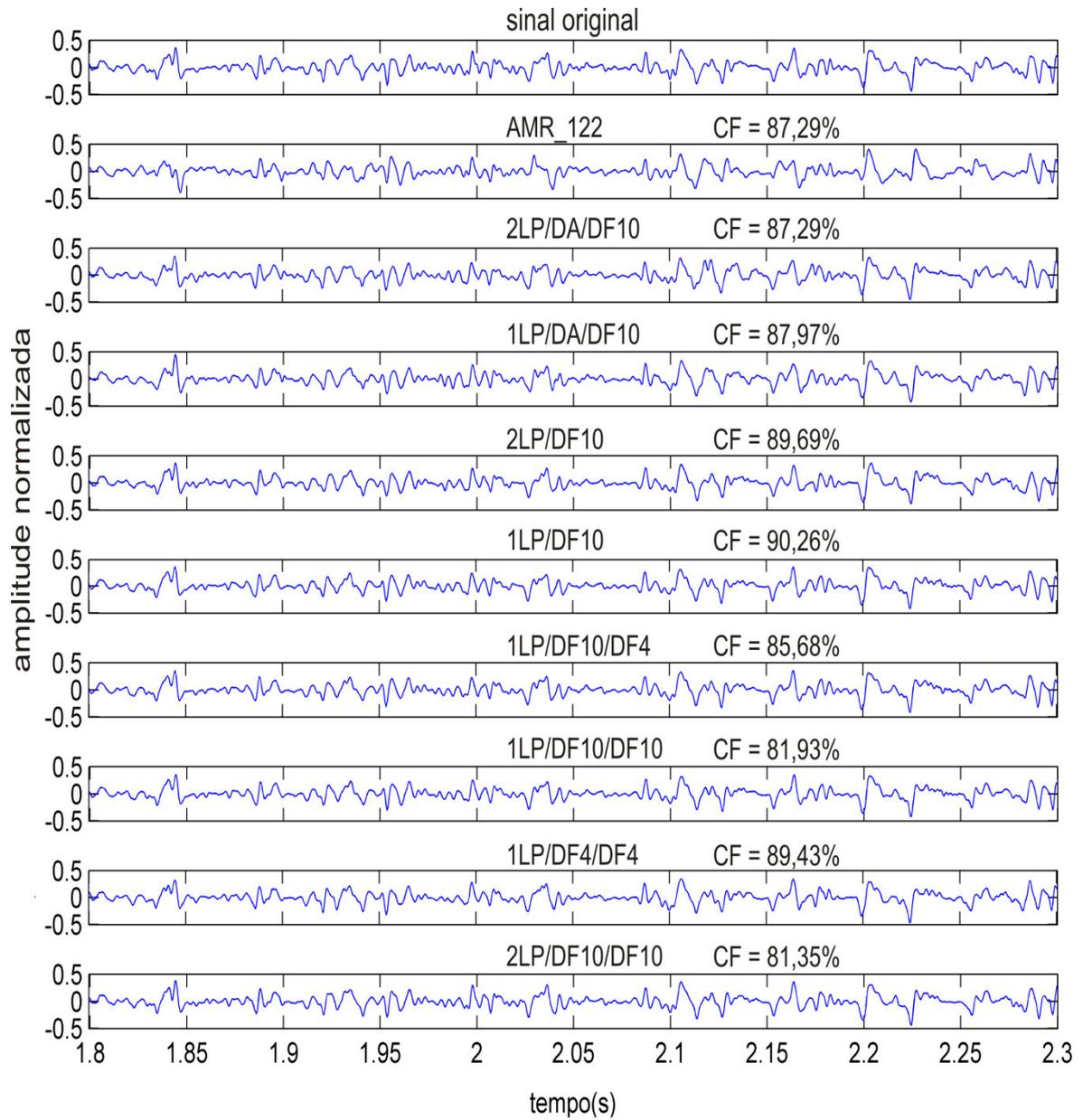


Figura 5.3 – Formas de onda do sinal de EMG original e reconstruídos a partir da codificação usando o modo AMR_122 e os novos modos.

As correspondentes formas de onda de voz podem ser vistas na Figura 5.4.

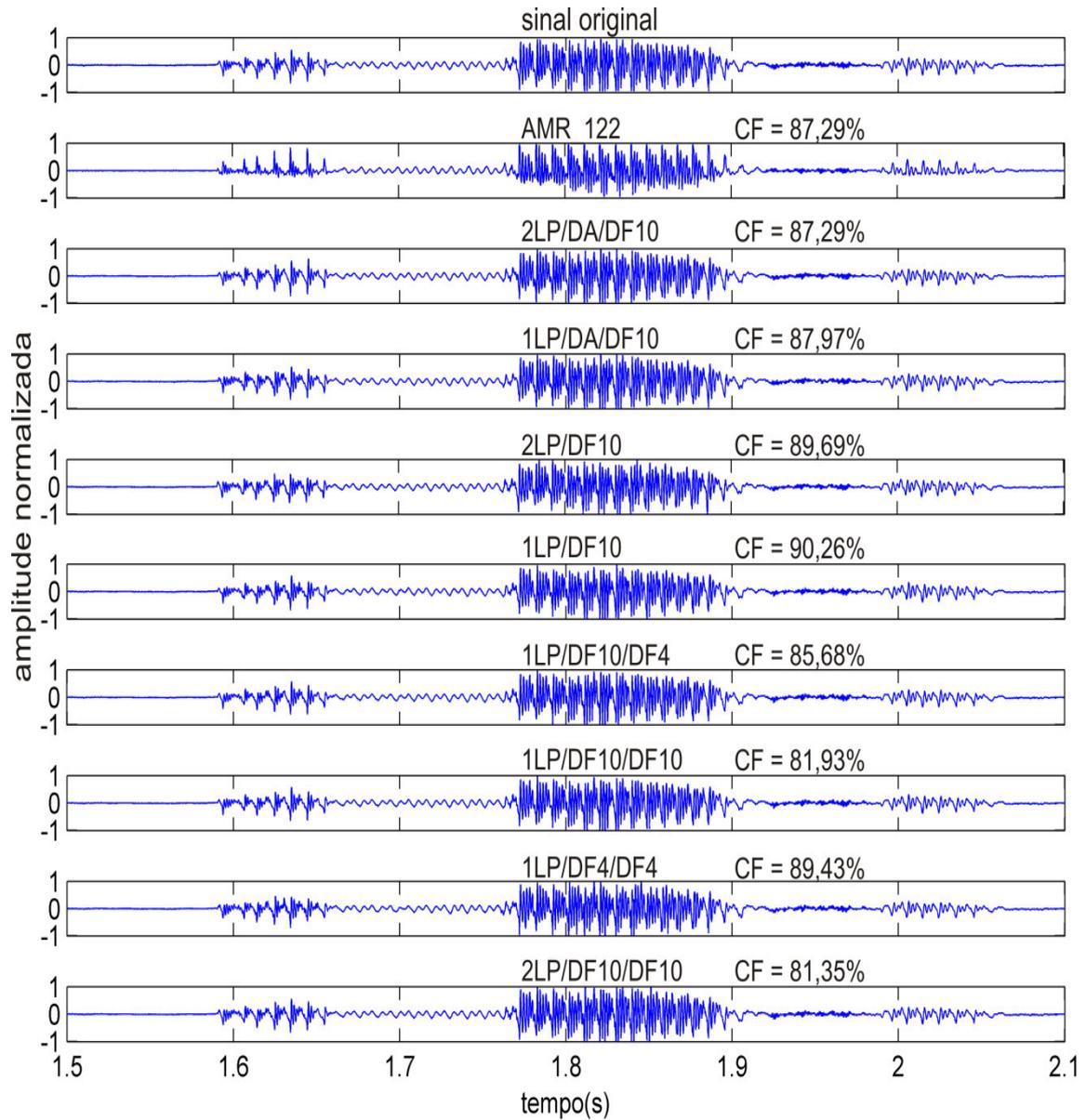


Figura 5.4 - Formas de onda do sinal de voz original e reconstruídos a partir da codificação usando o modo AMR_122 e os novos modos.

Uma comparação entre o desempenho dos novos modos, para os sinais de EMG e de voz, pode ser visto na Figura 5.5.

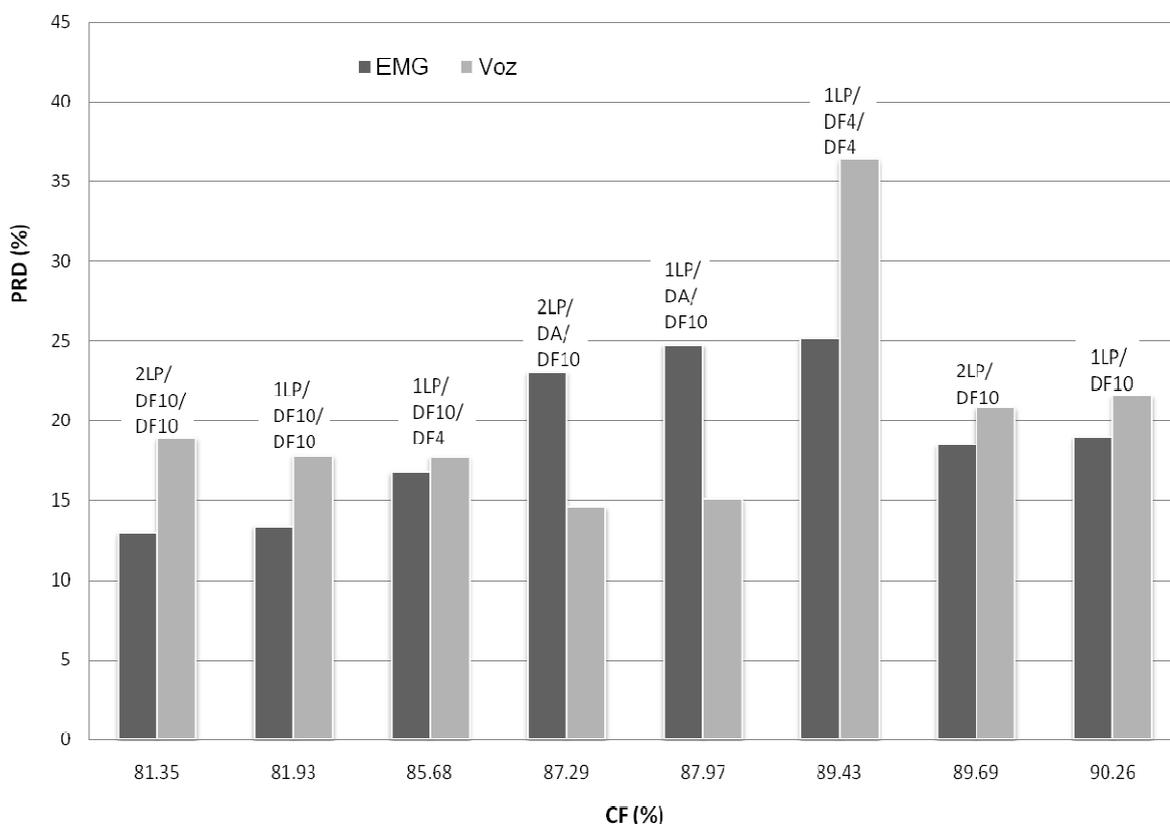


Figura 5.5– Comparação do desempenho dos novos modos de codificação para os sinais de voz e EMG.

Verifica-se que, em termos do erro percentual residual, o desempenho do codec é superior para EMG em todos os modos, exceto naqueles em que o dicionário adaptativo é utilizado (2LP/DA/DF10 e 1LP/DA/DF10), conforme discutido a seguir.

Os valores de PRD mostrados na Tabela 5.2 estão representados graficamente na Figura 5.6, apresentando a variação do erro com a taxa de compressão para os diversos sinais de EMG e modos em análise.

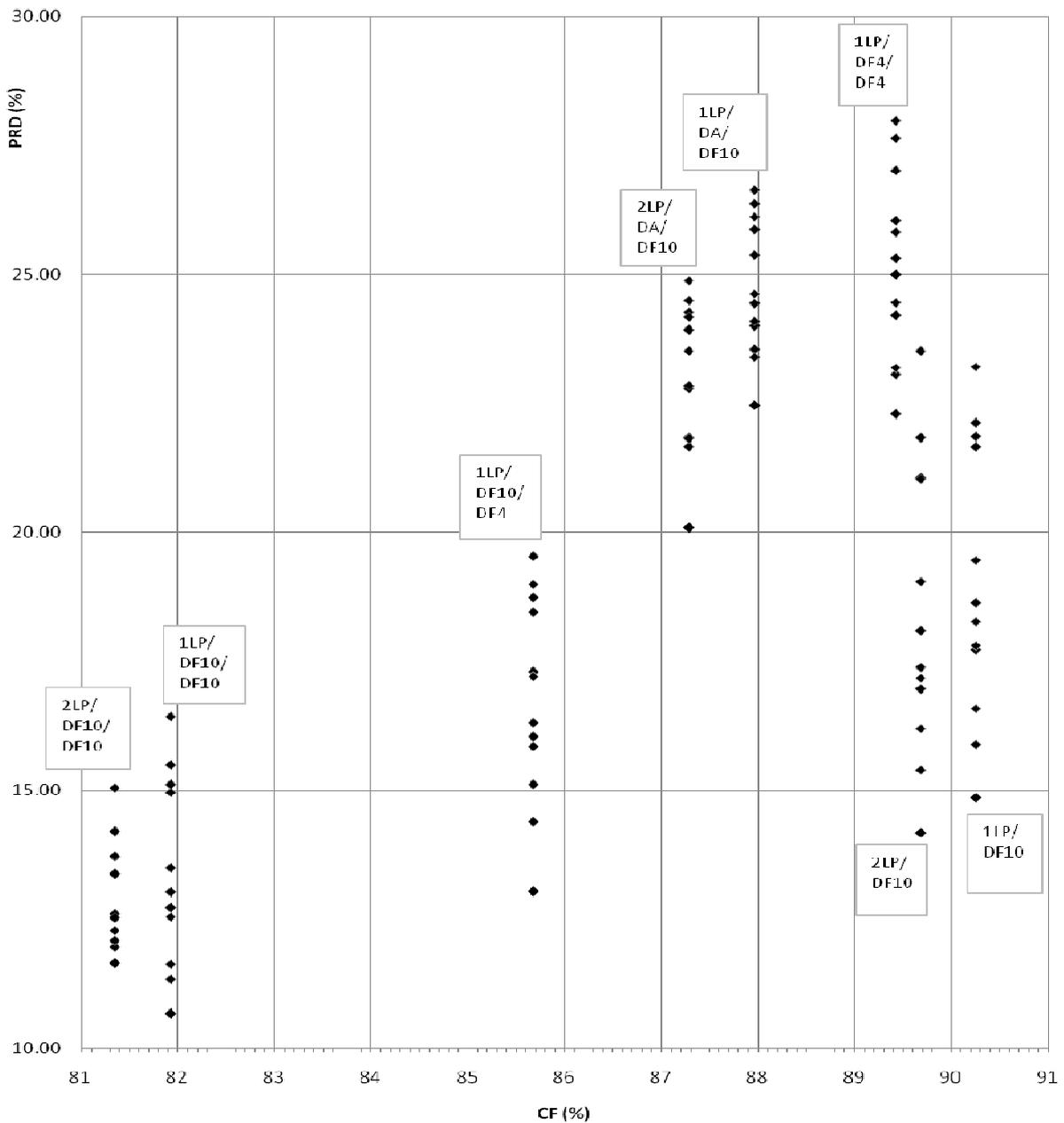


Figura 5.6 – PRD *versus* taxa de compressão para os novos modos de codificação aplicados a sinais EMG.

Ocorre um arranjo linear dos modos 2LP/DF10/DF10, 1LP/DF10/DF10, 1LP/DF10/DF4, 2LP/DF10 e 1LP/DF10, para os quais os erros médios aumentam com a taxa de compressão aproximadamente segundo uma reta.

Por sua vez, os modos 2LP/DA/DF10, 1LP/DA/DF10 e 1LP/DF4/DF4 encontram-se fora desta reta, apresentando valores de erro mais elevados, o que, nos dois primeiros casos, demonstra a ineficiência do uso do dicionário adaptativo para sinais de eletromiografia de

contrações isométricas e, no segundo, a ineficiência do uso apenas de dicionários fixos de menor tamanho.

A Figura 5.7 apresenta a comparação do sinal original de EMG com os sinais reconstruídos a partir da codificação utilizando os modos mais eficientes mencionados acima.

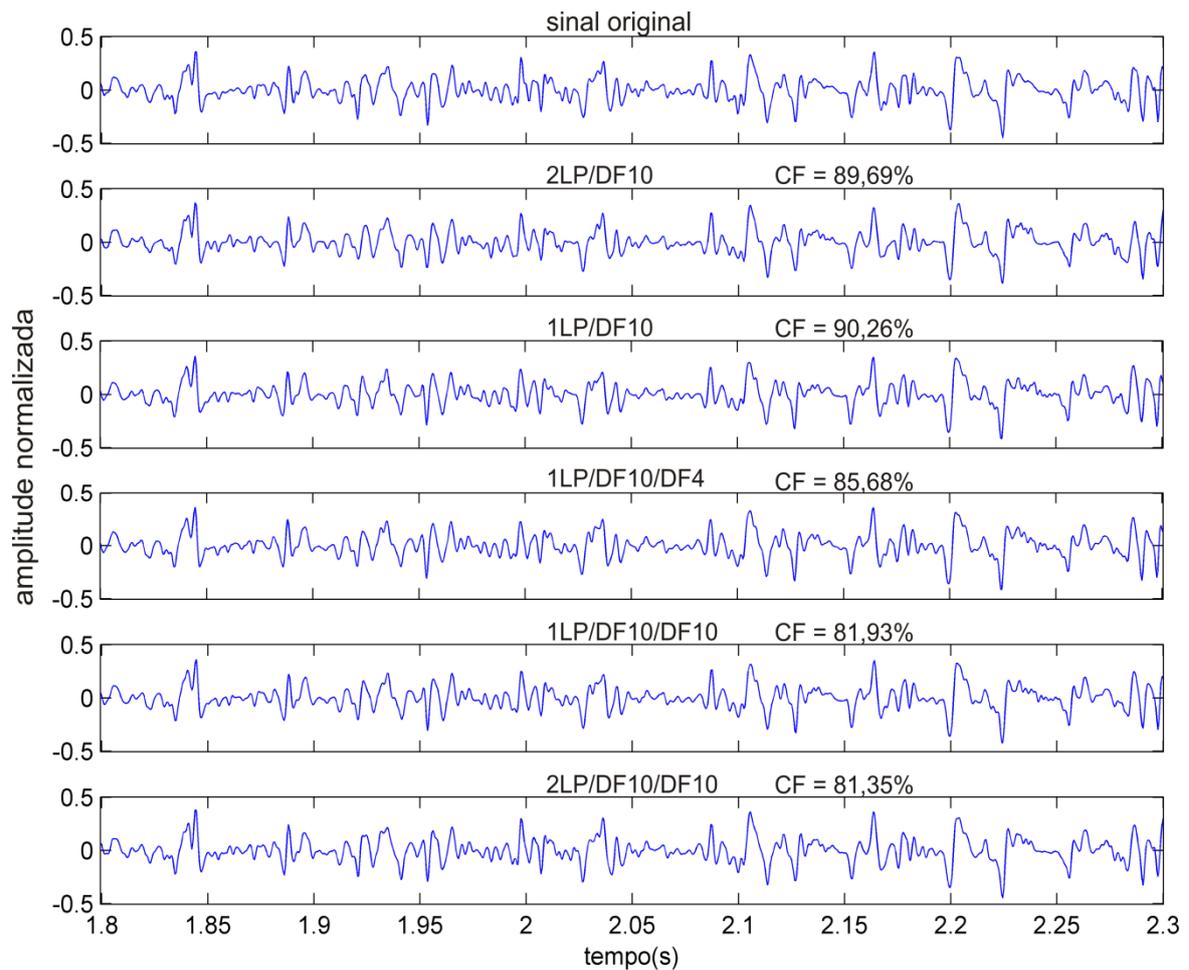


Figura 5.7 - Formas de onda de sinal de EMG original e reconstruído após codificação usando os modos mais eficientes em termos da relação entre o erro e o fator de compressão.

Da mesma maneira, a Figura 5.8 apresenta a comparação do sinal original de voz com os sinais reconstruídos a partir da compressão utilizando os modos mais eficientes mencionados acima.

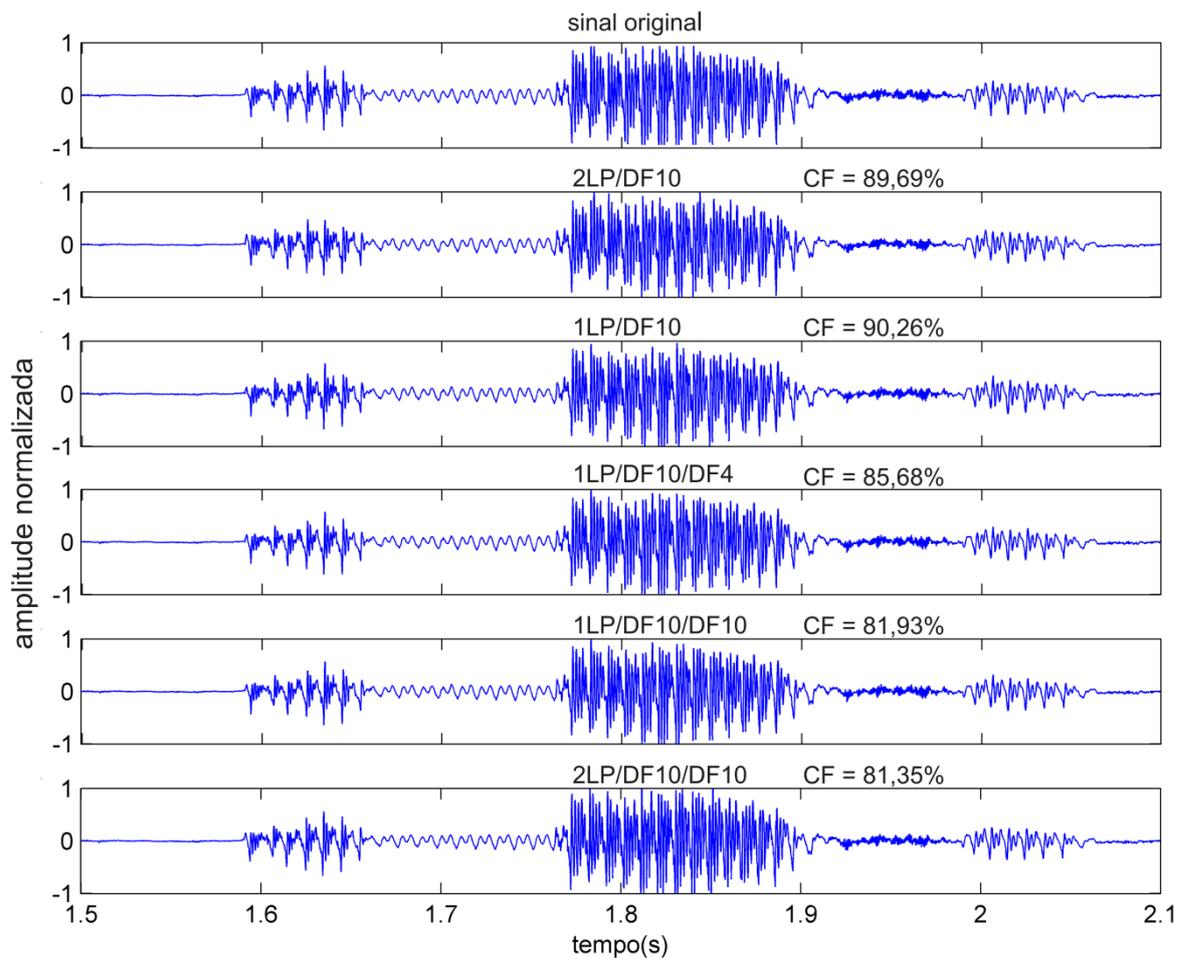


Figura 5.8 - Formas de onda de sinal de voz original e reconstruído após codificação usando os modos mais eficientes em termos da relação entre o erro e o fator de compressão.

5.6 – EFEITO DA RETIRADA DOS PROCEDIMENTOS DE MELHORIA DA QUALIDADE SUBJETIVA DA VOZ

Entre os modos AMR_122 e 2LP/DF10/DF10, ocorreu a redução do PRD tanto no caso do sinal de voz (de 78,32 para 14,61) quanto no caso dos sinais de eletromiografia (de $96,71 \pm 7,10$ para $23,02 \pm 1,44$). Esta variação deixa evidente a influência das rotinas de melhoramento da qualidade subjetiva na perda da qualidade objetiva. Isso é verdadeiro tanto para voz quanto, em maior escala, para EMG, conforme se vê na Figura 5.9.

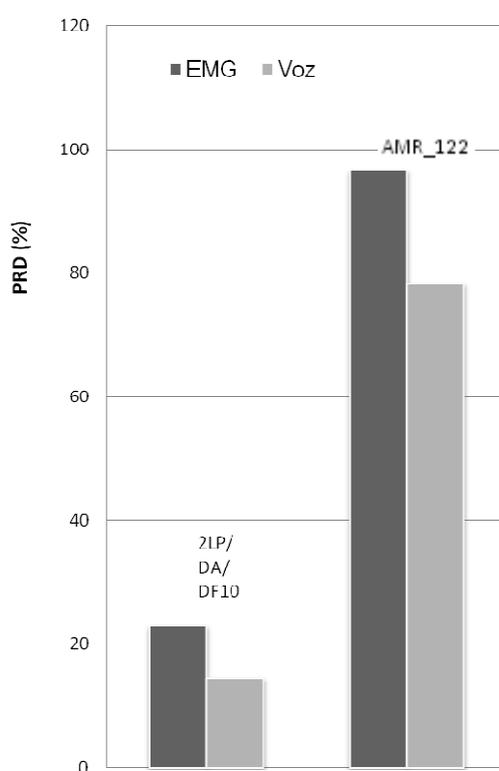


Figura 5.9 – Variação de desempenho devida às rotinas para a melhoria da qualidade subjetiva.

A Figura 5.10 apresenta trechos do sinal de EMG original e dos reconstruídos a partir da codificação usando os modos AMR_122 e 2LP/DA/DF10, indicando a maior preservação da forma de onda proporcionada por este último modo. A Figura 5.11 apresenta resultados equivalentes para o sinal de voz.

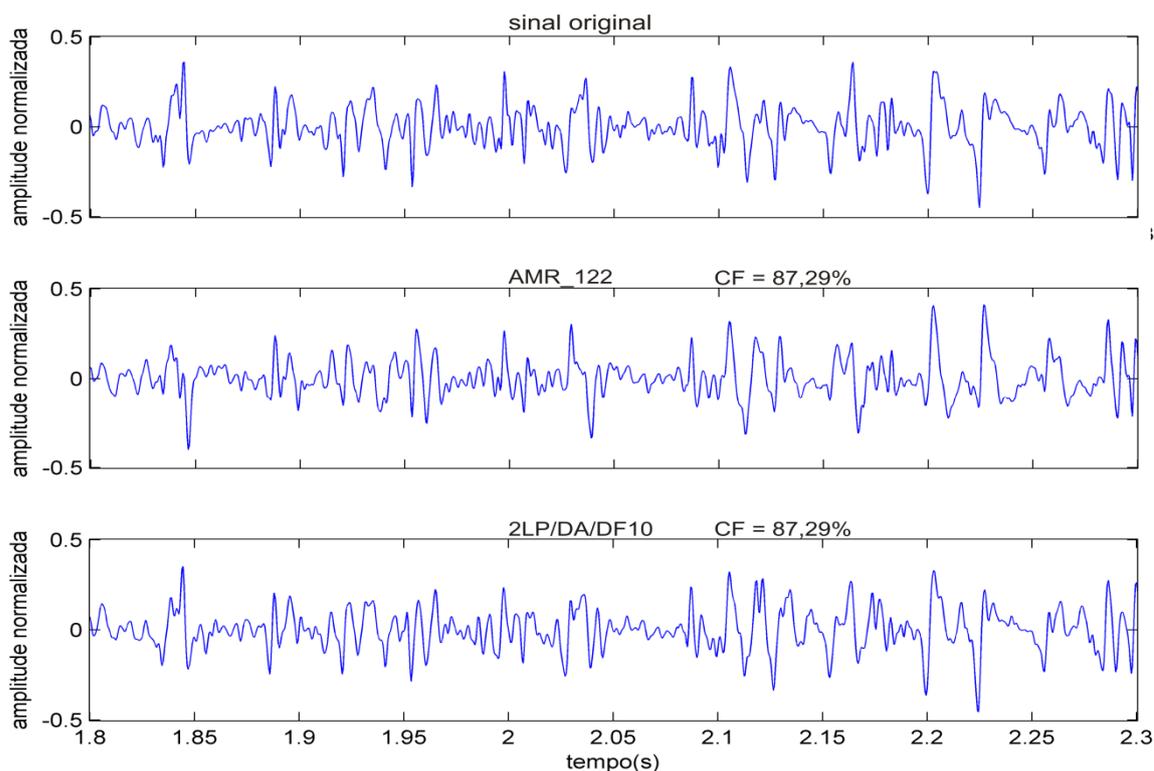


Figura 5.10 – Formas de onda do sinal de EMG original e reconstruído a partir da codificação com os modos AMR_122 e 2LP/DA/DF10, evidenciando o efeito da retirada das rotinas de melhoria da qualidade subjetiva da voz do codec original.

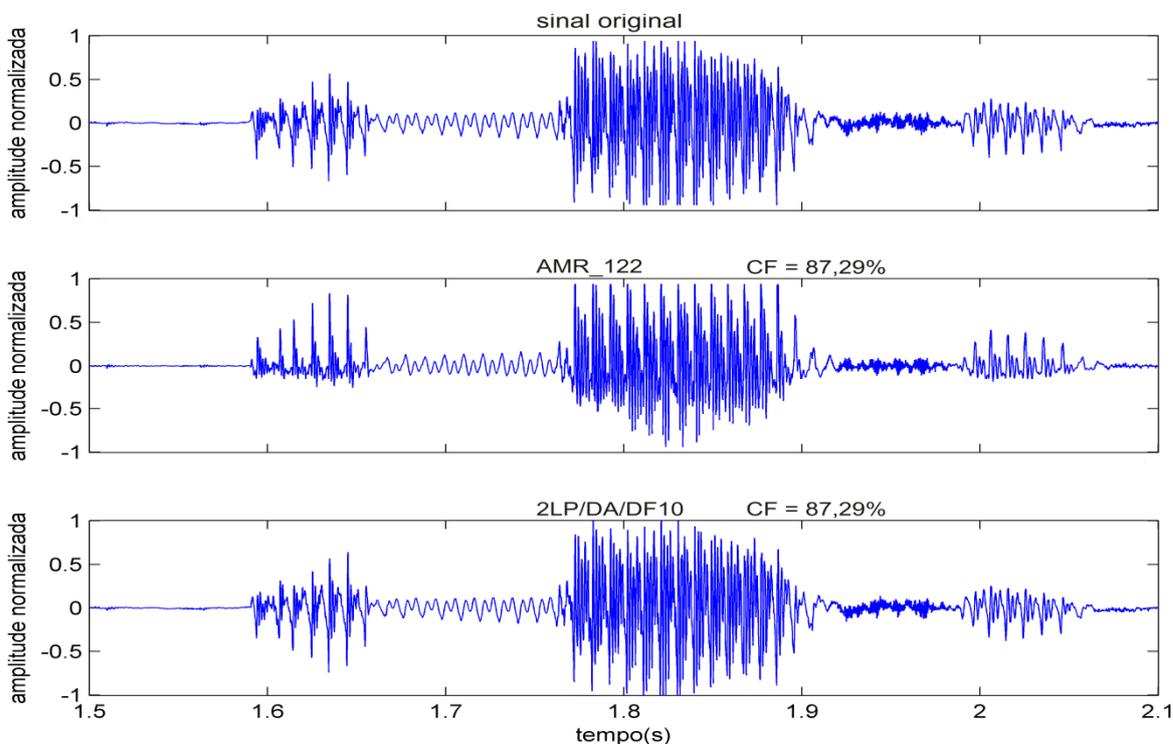


Figura 5.11 - Formas de onda do sinal de voz original e reconstruído a partir da codificação com os modos AMR_122 e 2LP/DA/DF10, evidenciando o efeito da retirada das rotinas de melhoria da qualidade subjetiva da voz do codec original.

5.7 – EFEITO DA RETIRADA DO DICIONÁRIO ADAPTATIVO

Um resultado interessante envolve os modos 2LP/DA/DF10 e 1LP/DA/DF10 e seus correspondentes 2LP/DF10 e 1LP/DF10. Embora os últimos apresentem taxas de compressão mais elevadas (em decorrência da ausência dos bits que codificassem os parâmetros do dicionário adaptativo), produziram, em média, taxas de erro menores.

Em uma primeira avaliação, a simples eliminação do dicionário adaptativo deveria levar a um erro maior. Entretanto, deve ser lembrado que o AMR é um codificador sub-ótimo, em que os parâmetros do dicionário adaptativo são determinados considerando-se nula a contribuição do dicionário fixo. A contribuição do dicionário adaptativo, uma vez determinada, é utilizada na determinação dos parâmetros do dicionário fixo. Além disso, o dicionário adaptativo é particularmente adequado à modelagem de aspectos periódicos da voz. Deste modo, para sinais com características diferentes do sinal de voz, a contribuição do dicionário adaptativo pode não apenas não levar à escolha dos melhores parâmetros do dicionário fixo como influenciar negativamente o desempenho global do codec.

5.8 – EFEITO DA ASSOCIAÇÃO DE DOIS DICIONÁRIOS FIXOS E DO USO DE UM OU DOIS CONJUNTOS DE PARÂMETROS LINEARES

A adição de um segundo dicionário fixo levou, em geral, à redução do erro (assim como da taxa de compressão) em relação ao modo equivalente com apenas um dicionário. Assim, entre os modos 2LP/DF10 e 2LP/DF10/DF10, o PRD reduziu-se de $18,49 \pm 2,84$ para $12,98 \pm 1,03$. No caso do modo 1LP/DF10, que apresentou PRD de $19,00 \pm 2,68$, a adição de um segundo dicionário fixo, efetuada nos modos 1LP/DF10/DF4 e 1LP/DF10/DF10, levou a um PRD de, respectivamente, $16,75 \pm 1,99$ e $13,34 \pm 1,80$, indicando que a redução do erro está condicionada ao tamanho (ou seja, número de pulsos não-nulos) do dicionário escolhido.

Nem sempre uma configuração com dois dicionários fixos mostrou os melhores resultados. Especificamente, o modo 1LP/DF4/DF4 apresentou um erro de $25,17 \pm 1,83$, enquanto que com o modo 1LP/DF10/DF4 conseguiu-se um erro de $16,75 \pm 1,99$, indicando a vantagem da utilização de ao menos um dicionário fixo de maior extensão.

A utilização de um ou dois conjuntos de parâmetros lineares, representando uma diferença de 11 bits, apresentou influência menos acentuada, como pode ser avaliado pela

comparação dos modos 1LP/DF10 e 2LP/DF10 (redução do PRD de $19,00 \pm 2,68$ para $18,49 \pm 2,84$) ou, analogamente, dos modos 1LP/DF10/DF10 e 2LP/DF10/DF10 (redução do PRD de $13,34 \pm 1,80$ para $12,98 \pm 1,03$).

5.9 – COMPARAÇÃO COM TÉCNICA DE COMPRESSÃO USANDO ACELP

A análise dos valores do erro quadrático médio (MSE) obtidos permite comparações com o trabalho desenvolvido por Carotti e colaboradores (2006), mencionado no Capítulo 1, que informam em seu trabalho ter sido empregada uma configuração modificada do modo AMR_122, com taxa de compressão de 87,3%, sendo eliminadas as rotinas de filtragem passa-altas e de pós-processamento para aumento da qualidade subjetiva. Ambos os dicionários (fixo e adaptivo) foram mantidos. Foram utilizados sinais de EMG de superfície, produzidos por contrações isométricas variando entre 10% e 70% da contração voluntária máxima (CVM), amostrados a 1024 Hz e submetidos a filtro passa-faixa com frequências de corte de 10 e 500 Hz. Nos resultados obtidos, o erro para todos os sinais reconstruídos esteve abaixo de 10%.

Em comparação, no presente trabalho foram utilizados sinais de EMG de contrações isométricas a 60% da contração voluntária máxima (CVM).

De modo a minimizar as diferenças entre os trabalhos, tomando de Carotti os valores computados de CVM imediatamente acima e abaixo de 60%, ou seja, 50% e 70%, verifica-se que obteve valores de MSE de, respectivamente, $5,95 \pm 1,40$ e $5,26 \pm 1,20$.

Pela Tabela 5.3, diversos novos modos de codificação apresentaram resultados com erros inferiores aos de Carotti (considerando 70% da CVM), quais sejam, 2LP/DF10, 1LP/DF10, 1LP/DF10/DF4, 1LP/DF10/DF10 e 2LP/DF10/DF10.

Por sua vez, considere-se o novo modo de codificação 2LP/DA/DF10, que é diretamente derivado do AMR (ver Figura 4.1) e, conseqüentemente, aquele que guarda maior correspondência com o trabalho de Carotti, apresentando a mesma taxa de compressão, CF=87,3%. Nesse modo, o valor obtido para o MSE foi de $5,53 \pm 0,65$, que é intermediário aos obtidos por Carotti com CVM de 50% e 70%.

Entretanto, com a eliminação do dicionário adaptativo realizada no modo 2LP/DF10, o desempenho mostrou-se melhor que o obtido por Carotti, tanto em relação ao MSE, que

apresentou redução para $3,49 \pm 1,07$, quanto à taxa de compressão, que aumentou para 89,69%. Finalmente, com o acréscimo do segundo dicionário fixo, resultando no modo 2LP/DF10/DF10, houve nova redução no valor do MSE, para $1,70 \pm 0,27$, embora, com a inclusão dos bits de codificação do dicionário fixo, a taxa de compressão tenha caído para 81,35%.

Uma comparação dos resultados mencionados acima é mostrada na Figura 5.12.

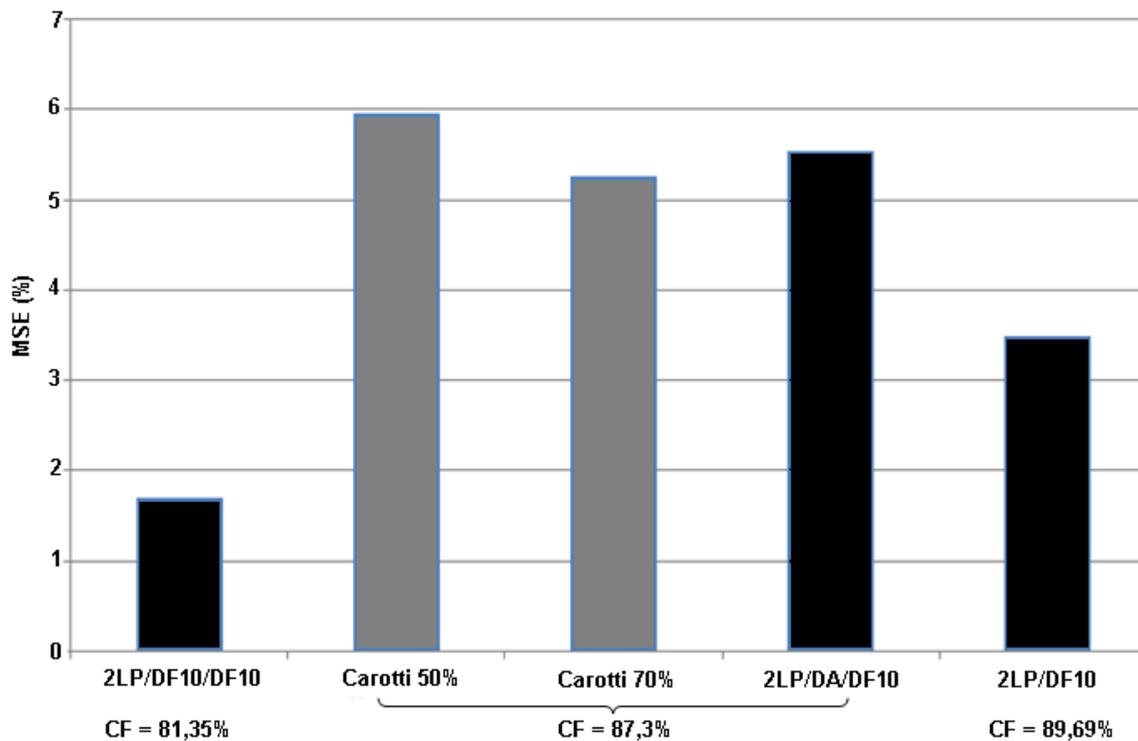


Figura 5.12 - Comparação entre o desempenho da compressão de EMG obtido por Carotti e por alguns dos novos modos de codificação.

Pode-se concluir que tais resultados indicam uma sensível melhoria na eficiência de diversos dos novos modos de codificação implementados neste trabalho em relação ao apresentado por Carotti, como no modo 2LP/DF10, em que houve melhora no desempenho relacionado à taxa de compressão e ao erro, e no modo 2LP/DF10/DF10, em que a redução do erro foi ainda maior que no modo anterior e a máxima de todos os novos modos.

5.10 – COMPARAÇÃO COM TÉCNICAS DE COMPRESSÃO USANDO WAVELETS E H.264

Conforme mencionado no Capítulo 1, diferentes técnicas de compressão de EMG foram utilizadas por Berger (2006), que empregou *wavelets* e redes neurais, e Costa et al (2008), que empregou o algoritmo H.264/AVC de compressão de vídeo. Nestes e no presente trabalho todos os sinais empregados foram obtidos de contrações isométricas.

Uma comparação entre os valores médios obtidos com cada técnica pode ser visto na Figura 5.13.

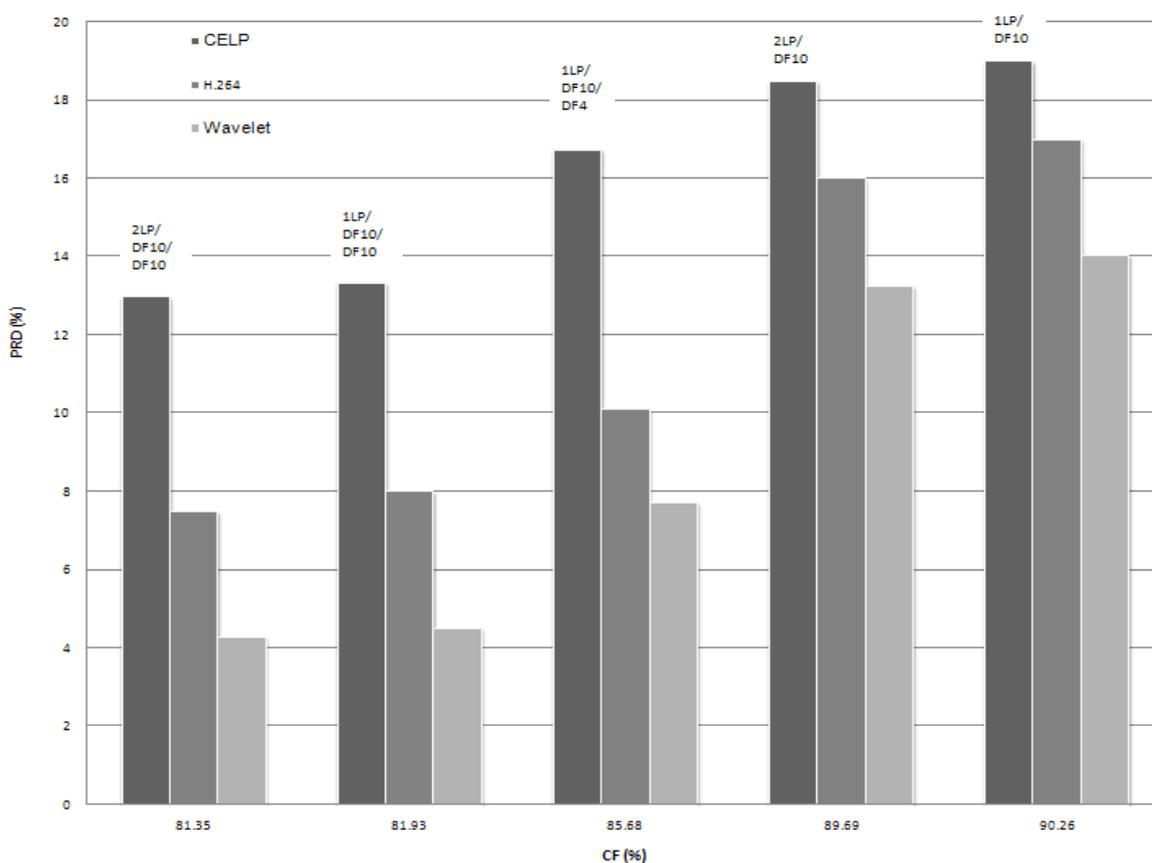


Figura 5.13 - Comparação entre os resultados dos algoritmos empregados neste trabalho (CELP), por Berger et al (Wavelet) e Costa *et al* (H.264/AVC).

Em geral, o erro entre sinais originais e reconstruídos obtido neste trabalho (CELP) está acima do obtido pelas outras técnicas, sendo a diferença menor para a técnica empregando H.264/AVC. Em particular, para as duas maiores taxas de compressão, que correspondem à região em que as técnicas de Berger e de Costa começam a apresentar um aumento de erro mais acentuado, as diferenças são menores.

6 – CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

6.1 – CONCLUSÕES GERAIS

Procurou-se avaliar neste trabalho a viabilidade da utilização de codificadores baseados no modelo CELP, na codificação de sinais de eletromiografia. De fato, os resultados obtidos foram compatíveis e em alguns casos melhores (item 5.9) que com os obtidos por outras pesquisas utilizando seja a mesma ou outras técnicas.

Verificou-se o aumento da qualidade objetivamente medida do sinal resultante da exclusão dos elementos de aumento da qualidade subjetiva e do dicionário adaptativo e da inclusão de um segundo dicionário fixo em cascata com o dicionário fixo original.

Foi avaliada a variação da qualidade obtida em função do tipo e quantidade de dicionários e da quantidade de conjuntos LSP utilizados.

O esquema de codificação baseado em CELP apresenta vantagens como um atraso algorítmico reduzido. O decodificador necessita esperar a chegada de um único quadro completo para efetuar a síntese e reconstrução (Carotti, 2007).

Deve ser dado destaque à complexidade relativamente reduzida e a eficiência obtida com o uso do código derivado do desenvolvido originalmente para o AMR, especialmente devido ao emprego do algoritmo ACELP, em que a busca do dicionário fixo é otimizada pela utilização de vetores em que apenas alguns elementos são não nulos e de valor unitário. Nos algoritmos empregados neste trabalho a simplificação aumenta após a exclusão do uso do dicionário adaptativo.

De fato, o codec AMR original é adequado a aplicações em tempo real e implementação não somente em dispositivos de comunicação celular de terceira geração como em dispositivos portáteis utilizando *chips* DSP (Yang, 2007; Kim, 2003).

6.2 – SUGESTÕES PARA TRABALHOS FUTUROS

Ganhos adicionais podem ser esperados ao serem explorados recursos específicos do codec AMR como a escolha dinâmica do modo (quadro a quadro) e a inserção de quadros SID de taxa mínima de bits resultante da detecção (com o recurso DTX) de períodos de inatividade

eletrofisiológica (equivalentemente a do silêncio na fala) que, empregadas em sinais EMG de superfície, permitiriam otimizar a taxa de compressão mantendo um nível mínimo de qualidade do sinal. A qualidade do sinal pode ser aferida por meio da correlação entre o sinal original e o sinal reconstituído.

Inovações mais profundas nos algoritmos empregados envolvem a criação de dicionários fixos de maior extensão, e conseqüentemente com maior número de pulsos não-nulos, uma vez que foi verificado que dicionários de maior extensão produziram menores erros. Além disso, um dicionário de maior extensão utilizando apenas um ganho economizará bits em comparação a dois dicionários menores empregando dois ganhos distintos. Novas combinações (ordenação) de dicionários podem também ser exploradas.

Algumas características do AMR permaneceram otimizadas para voz. Dentre elas o dicionário de quantificação de LSP e a ordem do filtro de predição linear. Pode ser efetuada a criação de dicionário de quantização de LSP específico para sinais de EMG e a alteração da ordem e coeficientes do filtro. Sinais de EMG de superfície possuem uma frequência máxima entre 1 kHz e 2 kHz, contra os 8 kHz dos sinais de voz processados pelo AMR. Para esta banda, o AMR baseia-se em um filtro de 10 pólos, o que poderia ser reduzido para sinais de EMG, o que em última análise resultaria em aumento da taxa de compressão.

Em termos de desempenho computacional, as rotinas utilizadas podem ser revisadas de modo a eliminar passos que se tornaram desnecessários em razão das modificações efetuadas, tais como aquelas associadas à utilização do filtro perceptivo $W(z)$, cujo efeito foi anulado igualando-se γ_1 e γ_2 a zero.

Por fim, seria interessante a extensão do uso de codificadores derivados do AMR para testes com eletromiografia dinâmica e interna e em contrações isocinéticas e isotônicas e com outros tipos de sinais eletrofisiológicos, tais como eletrocardiografia e eletroencefalografia. Os sinais de ECG, EEG e EMG isométrico possuem aspectos periódicos relevantes, fazendo supor a importância da presença do dicionário adaptativo.

REFERÊNCIAS BIBLIOGRÁFICAS

3GPP TS 26.071 V7.0.1 (2007-07) Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description (Release 7).

3GPP TS 26.073 V7.0.0 (2007-06) Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; ANSI C code for the Adaptive Multi Rate (AMR) speech codec (Release 7).

3GPP TS 26.090 V7.0.0 (2007-06) Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory Speech Codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec; Transcoding functions (Release 7).

3GPP TS 26.093 V7.0.0 (2007-06) Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory speech codec speech processing functions Adaptive Multi-Rate (AMR) speech codec; Source controlled rate operation (Release 7).

3GPP TS 26.104 V 7.0.0 (2007-06) Technical Specification, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; ANSI C code for the floating-point Adaptive Multi Rate (AMR) speech codec (Release 7).

Adoul, J.-P.; Mabillean, P.; Delprat, M.; Morissette, S., Fast CELP coding based on algebraic codes, *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87, Volume 12, Apr 1987* Page(s):1957 – 1960.

Al-Fahoum, A., Quality assessment of ECG compression techniques using a wavelet-based diagnostic measure, *IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 10, NO. 1, JANUARY 2006.*

- Al-Imari, A., et al, Design and implementation of a surface electromyogram system for sport field application, Proceedings. The 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications, 2003. 30 June-2 July 2003 Page(s):357 – 361.
- Alencar, M., *Telefonia Celular Digital*, Editora Érica, São Paulo, 2004.
- Atal, B.; Remde, J., A new model of LPC excitation for producing natural-sounding speech at low bit rates, *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82, Volume 7, May 1982* Page(s):614 – 617.
- Barreto, A.N.; Pirhonen, R.; Capacity Increase in GSM Networks Using Source-Adaptive AMR, *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd, Volume 2, 2006* Page(s):553 – 557.
- Basmajian, J., *Biofeedback - Principles and practice for clinicians*, 3rd ed. Baltimore, William & Wilkins, 1989.
- Basmajian, J., *Computers In Electromyography*. Butterworths, 1975.
- Berger, P., *Otimização de Desempenho de Algoritmos de Compressão de Sinais Biológicos Utilizando Redes Neurais Artificiais*, Tese (Doutorado em Engenharia Elétrica). Departamento de Engenharia Elétrica, Universidade de Brasília: UnB, 2006.
- Berger, P.; Nascimento, F.; Carmo, J.; Rocha, A., Compression of EMG signals with wavelet transform and artificial neural networks, *Physiol. Meas.* **27** 457-465, 2006.
- Berger, P. A.; Nascimento, F. A. O. ;Rocha, A. F. . Algoritmo Para Compressão de Sinais de ECG Utilizando a Transformada Wavelet Discreta e Redes Neurais Artificiais. In: XVI Congresso Brasileiro de Automática - XVI CBA, 2006, 2006, Salvador. Anais do XVI Congresso Brasileiro de Automática, 2006. p. 2684-2688.
- Berger, P. A.; Nascimento, F. A. O.; Rocha, A. F.; Menezes, L. R. A. X.; Carvalho, J. L. A., Biomedical Signal Compression: An Overview. In: Nilmini Wickramasinghe; Eliezer Geisler.. (Org.). *Encyclopedia of Healthcare Information Systems*. 1 ed. Hershey, Pennsylvania: Idea Group Inc, 2008, v. 1, p. 157-165.

- Betts, B.; Jorgensen, C., Small Vocabulary Communication and Control Using Surface Electromyography in an Acoustically Noisy Environment, <http://ti.arc.nasa.gov/publications/pdf/1041.pdf>.
- Betts, B.; Jorgensen, C., Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment, tech. memo TM-2005-213471, NASA, 2005, <http://ti.arc.nasa.gov/projects/nel/Papers/NASA-TM-2005-213471.pdf>.
- Binsted, K.; Jorgensen, C., Sub-Auditory Speech Recognition, HICSS 39 Symposium, 2006, <http://www.itl.nist.gov/iaui/vvrg/hicss39/HICSSPaper.doc>.
- Bradley J. Betts, B.; Jorgensen, C., Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment, HICSS 39 Symposium, 2006, www.itl.nist.gov/iaui/vvrg/hicss39/HICSS06%20Betts1.pdf.
- Brechet, L., et al, Compression of Biomedical Signals with Mother Wavelet Optimization and Best-Basis Wavelet Packet Selection, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 54, NO. 12, DECEMBER 2007.
- Buzo et al. 1980. Speech coding based upon vector quantization. IEEE Transactions on Acoustics, Speech, and Signal Processing 28 5, pp. 562-574.
- Câmara Junior, J. M., Dicionário de linguística e gramática: referente à língua portuguesa, VOZES, 2007.
- Carotti, E.; Martin J.; Merletti, R.; Farina, D., ACELP-based compression of multi-channel surface EMG signals, 32nd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2007. Poster. Abstract: <http://www.icassp2007.org/Papers/viewpapers.asp?papernum=3055>.
- Carotti, E.; Martin J.; Merletti, R.; Farina, D., Compression of surface EMG signals with algebraic code excited linear prediction, Medical Engineering & Physics Volume 29, Issue 2, , March 2007, Pages 253-258.
- Carotti, E.; Martin J.; Farina, D.; Merletti, R., Linear Predictive Coding Of Myoelectric Signals, IEEE ICASSP 2005.

- Cavalcante, J. H. A., Processamento de Sinais de Eletromiografia Dinâmica em Computador para Análise de Ciclos de Marcha. Relatório de Projeto Final, Departamento de Engenharia Elétrica, Universidade de Brasília: UnB, 1991.
- Cavalcante, J. H. A. et al, A System for Computer-Assisted Electromyographic Biofeedback for Neuromuscular Rehabilitation: A Case Study, Abstracts of the World Congress on Medical Physics and Biomedical Engineering 1994.
- Cetin, A. E., Köymen, H., Compression of Digital Biomedical Signals. The Biomedical Engineering Handbook: Second Edition. Ed. Joseph D. Bronzino, Boca Raton, CRC Press LLC, 2000.
- Chu, W., Speech Coding Algorithms: Foundation and Evolution of Standardized Coders, USA, John Wiley and Sons, Inc., 2003.
- Costa, M. et al, COMPRESSÃO DE SINAIS DE ELETROMIOGRAFIA DE SUPERFÍCIE USANDO H.264/AVC, XI Congresso Brasileiro de Informática em Saúde, 2008.
- Dayan, P.; Abbott, L. F., Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems. The MIT Press; 1st edition (2001).
- De Luca, C., Physiology and Mathematics of Myoelectric Signals, Biomedical Engineering, IEEE Transactions on, Volume BME-26, Issue 6, June 1979 Page(s):313 – 325.
- Farina, D.; Merletti, R., Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions J. Appl. Physiol., vol. 10, pp. 337–349, 2000.
- Fox, E. et al, Bases Fisiológicas da Educação Física e dos Desportos – 4ª Ed., Rio de Janeiro – Ed. Guanabara Koogan – 1991.
- Gatzoulis, L.; Iakovidis, I., Wearable and Portable eHealth Systems, IEEE Engineering in Medicine and Biology Magazine, vol 26, number 5, September/October 2007, pp. 51-56.
- Goldberg, R. G., A Practical Handbook of Speech Coders, USA: CRC Press LLC, 2000.

- Gronfors, T.K.; Paivinen, N.S.; The Effect of Vector Length and Gain Quantization Level on Medical Parameters of EMG Signals on Lossy Compression, *Advances in Medical, Signal and Information Processing*, 2006. MEDSIP 2006. IET 3rd International Conference On, 17-19 July 2006 Page(s):1 – 4.
- Gronfors, T.; Sihvonen, T.; Lamsa, V.; Paivinen, N.; Spectral Factors and Medical Parameters of EMG Signals, *Advances in Medical, Signal and Information Processing*, 2006. MEDSIP 2006. IET 3rd International Conference On, 17-19 July 2006 Page(s):1 – 4.
- Gronfors, T.K.; Paivinen, N.S.; Comparison of vector quantization methods for medical fidelity preserving lossy compression of EMG signals, *Computational Intelligence for Modelling, Control and Automation*, 2005 and *International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, International Conference on, Volume 1, 28-30 Nov. 2005 Page(s):1107 – 1113.
- Grönfors, T. et al, Vector quantization as a method for integer EMG signal compression, *Journal of Medical Engineering & Technology*, Vol. 30, No. 1. (February 2006), pp. 41-52.
- Guerrero, A.P.; Mailhes, C., On the choice of an electromyogram data compression method, *Engineering in Medicine and Biology Society*, 1997. *Proceedings of the 19th Annual International Conference of the IEEE*, Volume 4, 30 Oct.-2 Nov. 1997 Page(s):1558 - 1561 vol.4.
- R. Hagen, E. Ekudden, B. Johansson, and W.B. Kleijn, Removal of sparse-excitation artifacts in CELP, in *Proc. ICASSP'98*, pp. I-145-I-148.
- Hamill, J., *Bases Biomecânicas do Movimento Humano*, Editora Manole, 1999.
- Hanzo L.; Somerville, F. C. A.; Woodard, J. P., *Voice and Audio Compression for Wireless Communications Second Edition*, England, John Wiley & Sons, Ltd, 2007.
- Hillebrand, F. (ed.), *GSM and UMTS: The Creation of Global Mobile Communication*, John Wiley & Sons Ltd, 2001.
- Holden, M. et al., Telerehabilitation Using a Virtual Environment Improves Upper Extremity Function in Patients With Stroke, *IEEE TRANSACTIONS ON NEURAL*

SYSTEMS AND REHABILITATION ENGINEERING, VOL. 15, NO. 1, MARCH 2007.

Holma, H.; Melero, J.; Vainio, J.; Halonen, T.; Makinen, J.; Performance of adaptive multirate (AMR) voice in GSM and WCDMA, Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual, Volume 4, 22-25 April 2003 Page(s):2177 - 2181 vol.4.

Hogrel, J., Clinical applications of surface electromyography in neuromuscular disorders, *Neurophysiologie clinique* 35 (2005) 59–71, Elsevier.

Howard, P.; Vitter, J., Arithmetic Coding for Data Compression, (IEEE) 1993(8s).

Hughes, E; Bell, A., A wireless surface electromyography system, SoutheastCon, 2007. Proceedings. IEEE, 22-25 March 2007 Page(s):253 – 253.

Itakura, F., "Line spectral representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, Supplement no. 1, S35, 1975.

Jalaledine, S., et al., ECG Data Compression Techniques-A Unified Approach, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 37. NO. 4, APRIL 1990.

K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "GSM enhanced full rate speech codec", in *Proc. ICASSP'97*, pp. 771-774.

Jou, S. et al, Towards Continuous Speech Recognition Using Surface Electromyography, INTERSPEECH 2006 – ICSLP, http://www.cs.cmu.edu/~scjou/paper/scjou_icslp06.pdf

Kabaland, P.; Ramachandran, R.P., The computation of line spectral frequencies using Chebyshev polynomials, *IEEE Trans. on ASSP*, vol. 34, no. 6, pp. 1419-1426, Dec. 1986.

Kaku, Michio. *Visões do futuro; como a ciência revolucionará o século XXI*. Rio de Janeiro: Rocco, 2001. Título original: *Visions; how science will revolutionize the 21st century*, 1997.

- Kandel E et al. Principles of Neural Science, 4th Edition (2000), McGraw-Hill.
- Kleissen, R.F.M. et al., Electromyography in the biomechanical analysis of human movement and its clinical application, Gait and Posture 8 (1998) 143–158, Elsevier.
- Kondo, A. M., Digital Speech: Coding for Low Bit Rate Communication Systems, Second Edition, England: John Wiley & Sons, Ltd., 2004.
- Kong, X., Watermarking Medical Signals for Telemedicine, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 5, NO. 3, SEPTEMBER 2001.
- Lévy, P., As tecnologias da inteligência: O futuro do pensamento na era da informática. Rio de Janeiro: Nova Fronteira, 1993.
- Levinson, S., Mathematical Models for Speech Technology, USA: Wiley, 2005.
- Lim, J., Advanced Topics in Signal Processing, Wiley, 1988.
- Lin, C., The Design of a DSP Based Portable Physiological Signal Monitor, Proceedings - 19th International Conference - IEEE/EMBS Oct. 30 - Nov. 2, 1997.
- Makinen, J., Vainio, J., Source signal based rate adaptation for GSM AMR speech codec, Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), 2004.
- Maglogiannis, I. et al, Image, Signal, and Distributed Data Processing for Networked eHealth Applications, IEEE Engineering in Medicine and Biology Magazine, vol 26, number 5, September/October 2007, pp. 14-17.
- Merletti, R.; Parker, P. A., Electromyography: Physiology, Engineering and Noninvasive Applications, USA: Wiley-Interscience, 2004.
- Nait-Ali, A.; Cavaro-Menard, C. (Editores), Compression of Biomedical Images and Signals, England: Wiley-ISTE, 2008.
- Paliwal, K.K.; Atal, B.S., "Efficient vector quantization of LPC parameters at 24 bits/frame", IEEE Trans. Speech and Audio Processing, vol. 1, no 1, pp. 3 14, 1993.

- Rabiner, L.; Schafer, R., Digital Processing of Speech Signals, Prentice-Hall Series in Signal Processing, 1978.
- Salomoni, S., Compressão de Sinais Biológicos: Eletrocardiografia, Eletromiografia e Eletroencefalografia, ENE, UnB, 2006.
- Santaella, L., Culturas e Artes do Pós-Humano: da Cultura das Mídias à Cibercultura, São Paulo, Paulus, 2003.
- Santos, Paulo Henrique Marques, Melhoria da qualidade de fala através da interpolação de parâmetros do codec GSM-AMR em redes de pacotes, Dissertação de Mestrado em Engenharia Elétrica, UNICAMP, 2005.
- Schroeder, M.R.; Atal, B.S., "Code Excited Linear Prediction (CELP): High quality speech at very low bit rates," in Proc. ICASSP'85, pp. 937 940, 1985.
- Shimizu, K., Telemedicine by Mobile Communication, IEEE ENGINEERING IN MEDICINE AND BIOLOGY, Jul/August 1999.
- Silva, Lúcio Martins, Contribuições para a Melhoria da Codificação CELP a baixas Taxas de bits, Tese de Doutorado em Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro, PUC/RJ, Brasil, 1996.
- Soares, F. A., Estudo do Comportamento das Variáveis Eletromiográficas ao Longo do Ciclo Menstrual. Dissertação de Mestrado em Engenharia Elétrica, publicação PPGENE.DM-318/07, Departamento de Engenharia Elétrica, Universidade de Brasília: UnB, 2007.
- Tepavac, D. et al., Microcomputer-Based Portable Long-Term Spasticity Recording System. IEEE Transactions on Biomedical Engineering, vol. 39, no. 40. 1992.
- Titze, I., O Refinado Instrumento Humano, Scientific American Brasil, Duetto Editorial, fevereiro de 2008, pp. 62-69.
- Vale, E., Técnicas para a Melhoria de Desempenho da Transmissão de Voz no Sistema UMTS, Tese (Doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2006.

Veneziano W. H., Estudo do Comportamento do Sinal Eletromiográfico de Superfície em Atividades Subaquáticas. Tese (Doutorado em Engenharia Elétrica). Departamento de Engenharia Elétrica, Universidade de Brasília: UnB, 2006.

Xydeas, C. S.; Papanastasiou, C., Efficient coding of LSP parameters using split matrix quantization, Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing Detroit, MI, pp. 740-743, 1995.

Winter, D., Biomechanics and Motor Control of Human Movement, second edition, Wiley-Interscience, 1990.