

3

Codificadores de Voz em Telefonia IP e Móvel Celular

Para o desenvolvimento de sistemas de reconhecimento de voz distribuídos voltados a rede IP e rede móvel celular é necessário conhecer os codificadores de voz utilizados nas mesmas.

Os padrões de codificadores para redes IP e ambientes móveis celulares são enumerados abaixo, colocando-se na seguinte ordem as informações – nome do padrão / tipo do *codec* – ano – taxa – tipo de rede em que é utilizado:

- ITU-T G.723.1 / CELP (*Code-Excited Linear Predictive*) – 1995 – 5,3 ou 6,3 kbit/s – IP
- ITU-T G.729 / CS-ACELP (*Conjugate Structure-Algebraic Code-Excited Linear Prediction*) – 1995 – 8 kbit/s – IP
- TIA (*Telecommunications Industry Association*) – IS-54 / VSELP (*Vector Sum Excited Linear Predictive*) - 1992 – 8 kbit/s – Cel USA TDMA (*Time Division Multiple Access*)
- VSELP – Japão – 1993 – 6,7 kbit/s – Cel Japonês TDMA
- TIA – IS-95 / QCELP (*Qualcomm Code-Excited Linear Predictive*) – 1993 – 1 / 2 / 4 ou 8 kbit/s – Cel USA CDMA (*Code Division Multiple Access*)
- TIA – IS-96-A / QCELP – 1995 – 1,2 / 2,4 / 4,8 / 9,6 kbit/s – Cel USA CDMA
- GSM (*Global System for Mobile Communication*) – HR (*Half Rate*) / VSELP – 1995 – 5,6 kbit/s – Cel Europeu GSM
- GSM – EFR (*Enhanced Full Rate*) / ACELP – 1997 – 12,2 kbit/s – Cel Europeu GSM
- GSM – AMR-NB (*Adaptive Multi-Rate Narrowband*)/ ACELP – 1998 – 4,75 / 5,15 / 5,90 / 6,70 / 7,40 / 7,95 / 10,2 / 12,2 kbit/s – Cel Europeu GSM

- TIA – IS-641 (substitui o IS-54) / ACELP – 1997 – 7,4 kbit/s – Cel USA TDMA
- TIA – IS-733 / QCELP – 1998 – 1,8 / 3,6 / 7,8 / 14,4 kbit/s – Cel USA CDMA
- TIA – IS-127 EVRC (*Enhanced Variable Rate Coder*) / ACELP – 1998 – 1,2 / 4,8 / 9,6 kbit/s – Cel USA CDMA
- AMR-WB (*Adaptive Multi-Rate – Wideband*) / ACELP – 2001 – 6,6 / 8,85 / 12,65 / 14,25 / 15,85 / 18,25 / 19,85 / 23,05 / 23,85 kbit/s – Cel 3G Europa, Japão, USA, Coréia – WCDMA e IP

Destes codificadores apenas 3 serão apresentados neste capítulo, os *codecs* de voz padrão ITU-T G.723.1 [39], AMR-NB [40] e AMR-WB [41], pois são os codificadores mais utilizados pela indústria de telecomunicações nos dias atuais nas redes em análise nesta tese. O codificador ITU-T G.723.1 é um codificador de voz padronizado pela ITU (*International Telecommunication Union*) para utilização em redes IP. Já os codificadores AMR-NB e AMR-WB são codificadores de voz padronizados pelo 3GPP (*3rd Generation Partnership Project*) para utilização em redes móveis celulares, sendo que o AMR-WB foi também padronizado pelo ITU para uso em redes IP sobre a norma ITU-T G722.2 [42].

Na seção 3.1 deste capítulo será feita a apresentação do *codec* ITU-T G.723.1. Na seção 3.2 será descrito o *codec* AMR-NB e na seção 3.3 o *codec* AMR-WB. Finalmente, a seção 3.4 contém uma breve conclusão.

3.1. ITU-T G.723.1

O codificador ITU-T G.723.1 é um dos mais usados padrões para redes IP. O mesmo está presente em diversos produtos de grandes fabricantes e operadoras de telecomunicações. O *codec* ITU-T G.723.1 permite a codificação de voz a taxas de 6,3 kbit/s ou 5,3 kbit/s [39]. A taxa mais elevada fornece uma voz de melhor qualidade, porém a taxa mais baixa também fornece uma boa qualidade de voz. A diferença entre essas taxas resulta do tipo de excitação a ser utilizada e transmitida para o decodificador. Na taxa de 6,3 kbit/s, o codificador utiliza para a

excitação o MP-MLQ (*Multi-pulse Maximum Likelihood Quantization*), enquanto que na taxa de 5,3 kbit/s é empregado o ACELP (*Algebraic Code-Excited Linear Prediction*).

Este codificador é projetado para operar com um sinal digital, obtido primeiramente filtrando o sinal analógico de entrada com um filtro para telefonia (Recomendação ITU-T G.712 [43]), seguido de amostragem à taxa de 8 kHz e conversão para um PCM (*Pulse-Code Modulation*) de 16 bits, o qual será a entrada do codificador. A saída do decodificador deve ser convertida novamente para analógico, de forma similar.

O codificador opera sobre quadros de 240 amostras cada, o que equivale a 30 ms a uma taxa de amostragem de 8 kHz. O sinal sofre uma filtragem passa-altas a fim de remover a componente DC (*Direct Current*) e, em seguida é dividido em 4 sub-quadros de 60 amostras cada. Para todo sub-quadro é realizada uma análise LPC de ordem 10. Os parâmetros LPC do último sub-quadro são quantizados usando um quantizador PSVQ (*Predictive Split Vector Quantizer*), fazendo com que as LSFs sejam codificadas e transmitidas a cada 30 ms. Os demais parâmetros LPC dos outros sub-quadros em conjunto com o LPC do último sub-quadro, serão utilizados apenas para direcionar a busca da excitação de forma a considerar propriedades psicoacústicas [39].

O diagrama esquemático do codificador é apresentado na Fig. 3.1, onde se podem observar seus blocos básicos, bem como sua complexidade estrutural, a qual implica também em um grande consumo de recursos do terminal do usuário.

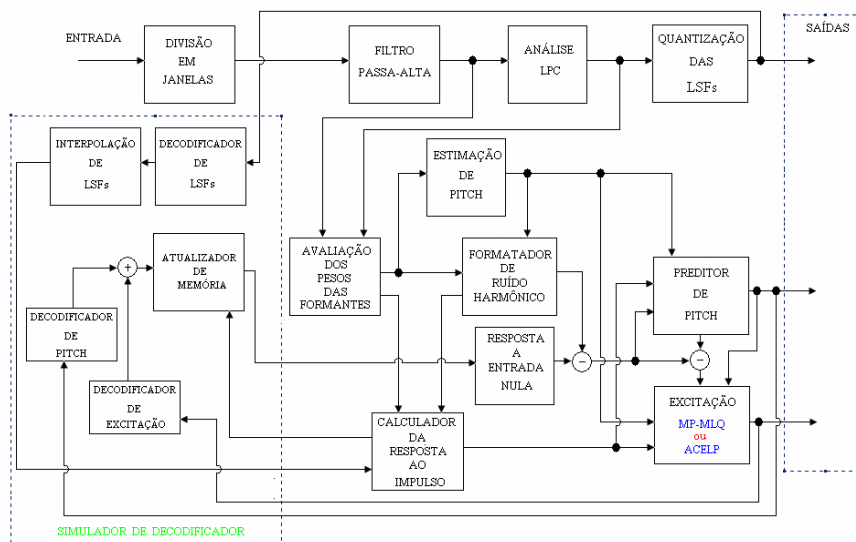


Figura 3.1 – Diagrama de blocos do codificador de voz do ITU-T G.723.1

Outra informação bastante importante sobre o codificador são as máscaras de alocação de bits utilizadas por ambas as taxas (define como serão utilizados os bits disponíveis em cada taxa de codificação) [39], apresentadas na Tab. 3.1, Essa informação dará subsídio a algumas afirmações sobre o sistema de reconhecimento distribuído a ser analisado quando se utiliza o codificador ITU-T G.723.1.

Modo	Parâmetro	1 - subquadro	2 - subquadro	3 - subquadro	4 -subquadro	total por quadro
6,3 kbit/s	LPC					24
	Dicionário Adaptativo	7	2	7	2	18
	Ganhos	12	12	12	12	48
	Posição dos Pulsos	20	18	20	18	73
	Sinais dos Pulsos	6	5	6	5	22
	Índice do Grid	1	1	1	1	4
	Total					189
5,3 kbit/s	LPC					24
	Dicionário Adaptativo	7	2	7	2	18
	Ganhos	12	12	12	12	48
	Posição dos Pulsos	12	12	12	12	48
	Sinais dos Pulsos	4	4	4	4	16
	Índice do Grid	1	1	1	1	4
	Total					158

Tabela 3.1 – Tabela de alocação de bits para o codificador ITU-T G.723.1

Vale ressaltar que como a taxa de transmissão das LSF e a precisão com que as mesmas são transmitidas (número de bits por quadro) são comuns tanto ao codificador funcionando a 5,3 ou 6,3 kbit/s, os resultados para os atributos de reconhecimento que dependam apenas das LSFs quantizadas ou dos parâmetros LPC recuperados das LSFs quantizadas não terão seu resultado de reconhecimento afetado pela variação da taxa de operação do codificador. Já no caso da voz reconstruída, deverá haver uma maior degradação do desempenho dos atributos obtidos através da mesma, para a taxa de 5,3 kbit/s.

A estrutura do decodificador apresentada pela norma ITU-T G.723.1 é aqui ilustrada na Fig. 3.2. A operação de decodificação é realizada quadro a quadro, primeiramente os índices dos parâmetros LSF quantizados são decodificados, sendo consecutivamente interpolados e usados para construir o filtro de síntese. Para cada sub-quadro, a excitação e o pitch são decodificados, sendo os mesmos combinados para gerar a entrada do filtro de síntese, que terá como saída a voz reconstruída a ser ainda filtrada pelo filtro de formantes, que tem como objetivo reduzir o ruído de quantização enfatizando as frequências das formantes e reduzindo os vales presentes no espectro do sinal, sendo sua forma apresentada em [39].

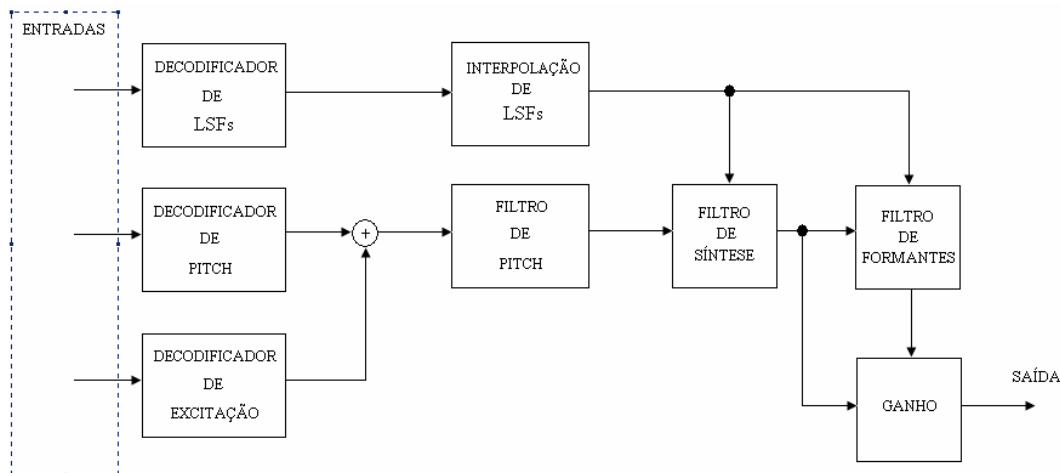


Figura 3.2 – Diagrama de blocos do decodificador de voz do ITU-T G.723.1

O *codec* utiliza também um detector de atividade da voz (VAD - *Voice Activity Detection*). Este detector de atividade decide se cada quadro é composto por voz ou por silêncio com base na energia do sinal amostrado. Os trechos de silêncio no discurso do emissor são codificados a uma taxa denominada SID (*Silence Descriptor*) que reproduz as características do silêncio produzindo o chamado “ruído de conforto” [44].

Nesta tese foi utilizado o código de referência para a aritmética de ponto flutuante definido em [45] para implementar o sistema de reconhecimento de voz distribuído usando o codificador ITU-T G.723.1.

3.2. Adaptive Multi-Rate Narrowband (AMR-NB)

O *codec* AMR-NB [40] é um codificador da família CELP (*Code-Excited Linear Predictive*) que é muito utilizado em sistemas celulares. Como o próprio nome sugere, ele pode operar em diversas taxas, que variam desde 4,75 até 12,2 kbit/s. Uma versão mais recente deste *codec*, o AMR-WB [41], permite trocas de taxas desde 6,6 até 23,85 kbit/s, o que será apresentado com mais detalhes na próxima seção deste capítulo.

Inicialmente o GSM (*Global System for Mobile Communications*) foi utilizado com os codificadores GSM-FR (*Full-Rate*) [46] e GSM-HR [47]. Com o intuito de oferecer qualidade equivalente às redes PSTN (*Public Switching*

Telecommunications Network), foi desenvolvido o GSM-EFR [48] e, em 1998, – a referência pode chamar atenção como sendo de 2008, porém pelo fato destas normas para codificadores estarem em constante revisão, foi considerada a data da revisão que gerou a versão utilizada como base para esta tese – o AMR-NB foi adotado na fase 2+ do GSM (referente à *Release 98 – Rel'98* – de software dos sistemas que adotam o padrão GSM). Atualmente o AMR-WB (*Adaptative Multi-Rate Wideband*) é recomendado pelo 3GPP como padrão a ser utilizado em aplicações e serviços de terceira geração definidos pelo IMT-2000 (*International Mobile Telecommunications - 2000*). Até a adoção do AMR-NB, a maioria dos padrões de codificação de voz utilizava taxas fixas.

O AMR-NB fornece 8 taxas de codificação conforme mostrado na Tab. 3.2. Ele também implementa um detector de atividade da voz (VAD)[49]. Este detector de atividade decide se cada quadro é composto por voz ou por silêncio com base na energia do sinal amostrado. Os trechos de silêncio no discurso do emissor são codificados a uma taxa denominada SID (*Silence Descriptor*) que reproduz as características do silêncio, produzindo o chamado “ruído de conforto” [50]. Além disso, o AMR-NB possui mecanismo de substituição e silenciamento de quadros perdidos que diminui os efeitos da perda de pacotes na rede [51]. A finalidade da substituição é atenuar e ocultar o efeito dos quadros perdidos. A finalidade de silenciar a saída, no caso de muitos quadros perdidos, é indicar a interrupção do canal ao usuário e evitar a geração de possíveis sons inoportunos como um resultado do procedimento de substituição de quadro. Para aumentar a qualidade subjetiva, quadros de fala perdidos são substituídos tanto por repetição como por extrapolação de bom(ns) quadro(s) de fala anterior(es). Esta substituição é feita de forma que o nível de saída diminua gradualmente, resultando em silêncio a partir de um determinado número de quadros perdidos [51].

Modo	Taxa (kbps)	Classes de bits			Total
		A	B	C	
AMR-NB 4.75	4.75	42	53	0	95
AMR-NB 5.15	5.15	49	54	0	103
AMR-NB 5.90	5.90	55	63	0	118
AMR-NB 6.70	6.70	58	76	0	134
AMR-NB 7.40	7.40	61	87	0	148
AMR-NB 7.95	7.95	75	84	0	159
AMR-NB 10.2	10.2	65	99	40	204
AMR-NB 12.2	12.2	81	103	60	244
SID	1.80	39	0	0	39

Tabela 3.2 – Taxa de codificação do AMR-NB e alocação de bits nas classes

Ao receber o sinal de fala, o codificador AMR-NB faz a amostragem do sinal a uma taxa de 8 kHz para gerar quadros de 20 ms (correspondendo a 160 amostras). Cada quadro de 20 ms de voz produz, 95, 103, 118, 134, 148, 159, 204 ou 244 bits de informação dependendo da taxa de codificação utilizada. Após a codificação da voz, os bits são separados em três categorias (A, B e C) conforme a sua importância. Durante a codificação do canal, tais bits são protegidos de acordo com a importância que lhe foi atribuída (codificação de canal mais poderosa ou menos poderosa). No entanto, modernamente a maioria dos equipamentos decodificadores continua a decodificação dos bits menos significantes (Classes B e C), mesmo que tenham sido detectados, previamente, erros severos nos bits de Classe A. Parte-se do princípio de que é mais conveniente tentar-se recuperar a inteligibilidade da voz a partir de quadros com erro do que a partir de quadros que tenham sido suprimidos [52].

O fluxo completo de sinal no codificador é mostrado na Fig. 3.3. A análise LPC é executada duas vezes por quadro para o modo AMR-NB 12.20 e uma vez para os demais modos. Para o modo AMR-NB 12.20, os dois conjuntos de parâmetros LPC são convertidos para LSFs (*Line Spectral Frequencies*) e são conjuntamente quantizados usando-se SMQ (*Split Matrix Quantization*) com 38 bits. Para os outros modos, o único conjunto de parâmetros LPC é convertido para LSFs e é quantizado vetorialmente usando-se SVQ (*Split Vector Quantization*). Os LSFs são uma representação dos parâmetros LPC no domínio da frequência. Detalhes da conversão dos parâmetros LPC para a representação LSF podem ser encontrados em [53]. O quadro de fala é dividido em 4 sub-quadros de 5 ms cada (40 amostras). Os parâmetros do dicionário adaptativo e fixo são transmitidos na cadência de sub-quadro. Os parâmetros LSFs quantizados e não quantizados ou suas versões interpoladas são usados dependendo do sub-quadro. Um período de pitch em malha aberta é estimado em todos os sub-quadros (exceto para os modos AMR 5.15 e AMR 4.75, nos quais isto é feito uma vez por quadro), baseado no sinal de fala ponderado perceptualmente. Esta estimativa é realizada para simplificar a análise do pitch e para realizar a busca em malha fechada utilizando valores próximos ao que foi estimado em malha aberta [54].

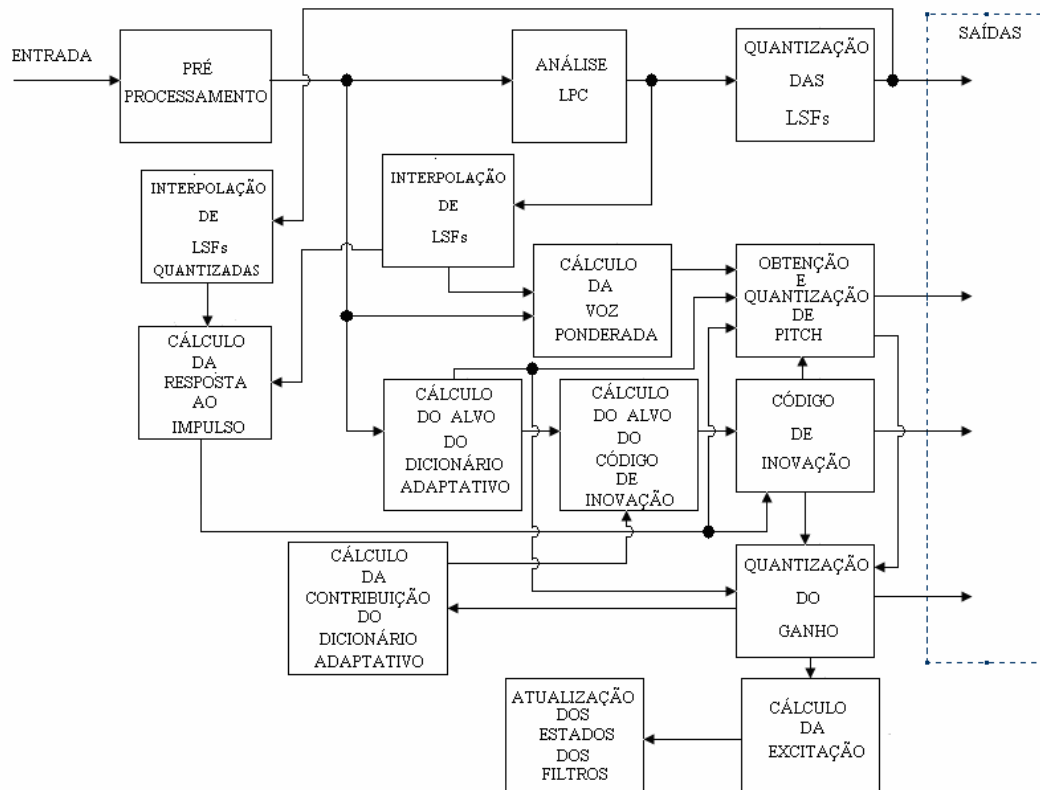


Figura 3.3 – Diagrama de blocos do codificador de voz do AMR-NB

Outra informação bastante importante sobre o codificador são as máscaras de alocação de bits utilizadas nas diversas taxas [54], apresentadas na Tab. 3.3. Essa informação dará subsídio a algumas afirmações sobre o sistema de reconhecimento distribuído a ser analisado quando se utiliza o codificador AMR-NB.

Modo	Parâmetro	1 - subquadro	2 - subquadro	3 - subquadro	4 -subquadro	total por quadro
12.2 kbit/s (GSM EFR)	2 conj. LSF					38
	Retardo de Pitch	9	6	9	6	30
	Ganho de Pitch	4	4	4	4	16
	Código Algébrico	35	35	35	35	140
	Ganho de Código	5	5	5	5	20
	Total					244
10.2 kbit/s	conj. LSF					26
	Retardo de Pitch	8	5	8	5	26
	Código Algébrico	31	31	31	31	124
	Ganhos	7	7	7	7	28
		Total				
7.95 kbit/s	conj. LSF					27
	Retardo de Pitch	8	6	8	6	28
	Ganho de Pitch	4	4	4	4	16
	Código Algébrico	17	17	17	17	68
	Ganho de Código	5	5	5	5	20
	Total					159
7.40 kbit/s (TDMA EFR)	conj. LSF					26
	Retardo de Pitch	8	5	8	5	26
	Código Algébrico	17	17	17	17	68
	Ganhos	7	7	7	7	28
		Total				
6.70 kbit/s (PDC EFR)	conj. LSF					26
	Retardo de Pitch	8	4	8	4	24
	Código Algébrico	14	14	14	14	56
	Ganhos	7	7	7	7	28
		Total				
5.90 kbit/s	conj. LSF					26
	Retardo de Pitch	8	4	8	4	24
	Código Algébrico	11	11	11	11	44
	Ganhos	6	6	6	6	24
		Total				
5.15 kbit/s	conj. LSF					23
	Retardo de Pitch	8	4	4	4	20
	Código Algébrico	9	9	9	9	36
	Ganhos	6	6	6	6	24
		Total				
4.75 kbit/s	conj. LSF					23
	Retardo de Pitch	8	4	4	4	20
	Código Algébrico	9	9	9	9	36
	Ganhos		8		8	16
		Total				

Tabela 3.3 – Tabela de alocação de bits para o codificador AMR-NB

O fluxo do sinal no decodificador é mostrado na Fig. 3.4. Os índices transmitidos são extraídos do fluxo de bits recebido e, então, decodificados para obtenção dos parâmetros do codificador em cada quadro transmitido. Estes parâmetros são os vetores LSF, os períodos de pitch fracionários, os vetores de código de inovação, e os respectivos ganhos de pitch e de inovação. Os vetores LSF são convertidos para os coeficientes LPC do filtro e interpolados para obter filtros LPC em cada sub-quadro. Então, a cada 40 amostras de sub-quadro:

- a excitação é construída adicionando-se os vetores de código adaptativo e de inovação escalados pelos seus respectivos ganhos;
- a fala é reconstruída filtrando-se a excitação através do filtro LPC de síntese. Finalmente, o sinal de fala reconstruído passa por um filtro de formantes que tem como objetivo reduzir o ruído de quantização enfatizando as frequências das formantes e reduzindo os vales presentes no espectro do sinal, sendo sua forma apresentado em [40].

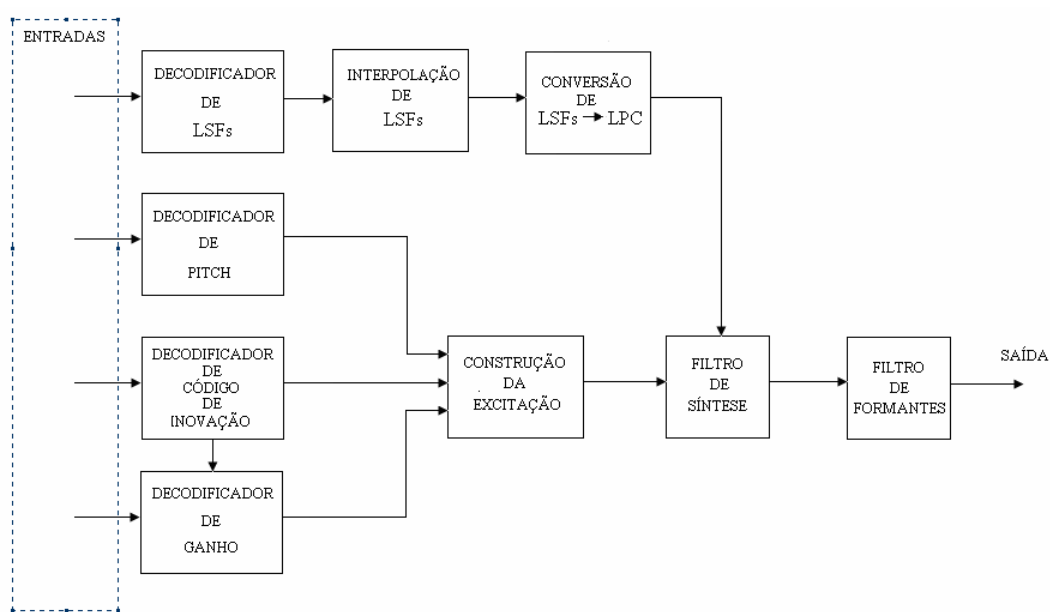


Figura 3.4 – Diagrama de blocos do decodificador de voz do AMR-NB

Nesta tese foi utilizado o código de referência para a aritmética de ponto flutuante definido em [55] para implementar o sistema de reconhecimento de voz distribuído usando o codificador AMR-NB. Porém, é possível também implementar o sistema utilizando o código do mesmo codificador utilizando a aritmética de ponto fixo, conforme definido em [56] quando o processador utilizado, ou o sistema operacional, não for compatível com a aritmética de ponto flutuante.

3.3. Adaptive Multi-Rate Wideband (AMR-WB)

O *codec* AMR-WB [41] é um codificador da família CELP (*Code-Excited Linear Predictive*) que já é utilizado em sistemas celulares (dependerá apenas do

modelo do aparelho em uso pelo usuário – já existem modelos no mercado com esta funcionalidade e da cobertura da operadora – já existem locais onde este codificador já é suportado), porém também é utilizado para redes IP [42], principalmente para a codificação de áudio para vídeo conferências. Como o próprio nome sugere, ele pode operar em diversas taxas, que variam desde 6,6 até 23,85 kbit/s.

O AMR-WB fornece 9 taxas de codificação conforme mostrado na Tab. 3.4. Ele também utiliza um detector de atividade da voz (VAD) [57]. Este detector de atividade decide se cada quadro é composto por voz ou por silêncio com base na energia do sinal amostrado. Os trechos de silêncio no discurso do emissor são codificados a uma taxa denominada SID (*Silence Descriptor*) que reproduz as características do silêncio produzindo o chamado “ruído de conforto” [58]. Além disso, o AMR possui mecanismo de substituição e silenciamento de quadros perdidos que diminui os efeitos da perda de pacotes na rede [59].

Modo	Taxa (kbps)	Classes de bits		Total
		A	B	
AMR-WB 6.60	6.60	54	78	132
AMR-WB 8.85	8.85	64	113	177
AMR-WB 12.65	12.65	72	181	253
AMR-WB 14.25	14.25	72	213	285
AMR-WB 15.85	15.85	72	245	317
AMR-WB 18.25	18.25	72	293	365
AMR-WB 19.85	19.85	72	325	397
AMR-WB 23.05	23.05	72	389	461
AMR-WB 23.85	23.85	72	405	477
SID	1.75	40	0	40

Tabela 3.4 – Taxa de codificação do AMR-WB e alocação de bits nas classes

Após a geração da fala, o codificador AMR-WB faz a amostragem do sinal a uma taxa de 16 kHz para gerar quadros de 20 ms (correspondendo a 320 amostras). Cada quadro de 20 ms de voz produz 132, 177, 253, 285, 317, 365, 397, 461, 477 bits de informação dependendo da taxa de codificação utilizada. Após a codificação da voz, os bits são separados em duas categorias (A, B) conforme a sua importância. Podemos observar, então, que diferentemente do AMR-NB, o AMR-WB não possui bits categorizados como de classe C. Durante a codificação do canal, tais bits são protegidos de acordo com a importância que lhe foi atribuída (codificação de canal mais poderosa ou menos poderosa). No entanto, modernamente a maioria dos equipamentos decodificadores continua a

decodificação dos bits menos significantes (Classe B), mesmo que tenham sido detectados, previamente, erros severos nos bits de Classe A. Parte-se do princípio de que é mais conveniente tentar-se recuperar a inteligibilidade da voz a partir de quadros com erro do que a partir de quadros que tenham sido suprimidos [52].

O fluxo completo de sinal no codificador é mostrado na Fig. 3.5. Depois da sub-amostragem (decimação), filtragem passa-altas e pré-ênfase, a análise LPC é realizada uma vez por quadro. O conjunto de parâmetros LPC é convertido para os ISFs (*Immittance Spectrum Frequencies*) e quantizados vetorialmente usando-se S-MSVQ (*Split-Multistage Vector Quantization*). Detalhes da conversão dos parâmetros LPC para a representação ISF podem ser encontrados em [60]. No capítulo 4 será explicado porque se optou neste codificador pela conversão dos LPC para ISF e não para LSF, apresentando as manipulações matemáticas para esta conversão LPC para ISF. O quadro de fala é dividido em 4 sub-quadros de 5 ms cada (64 amostras com taxa de amostragem de 12,8 kHz). Os parâmetros do dicionário adaptativo e fixo são transmitidos na cadência de sub-quadro. Os parâmetros LPC quantizados e não quantizados ou suas versões interpoladas são usados dependendo do sub-quadro. Um período de pitch em malha aberta é estimado em todos os sub-quadros ou uma vez por quadro baseado no sinal de fala ponderado perceptualmente. Esta estimativa é realizada para simplificar a análise do pitch e para realizar a busca em malha fechada utilizando valores próximos ao que foi estimado em malha aberta [61]. Em adição a estes parâmetros, os índices de ganho da banda alta são computados para o modo de 23.85 kbit/s.

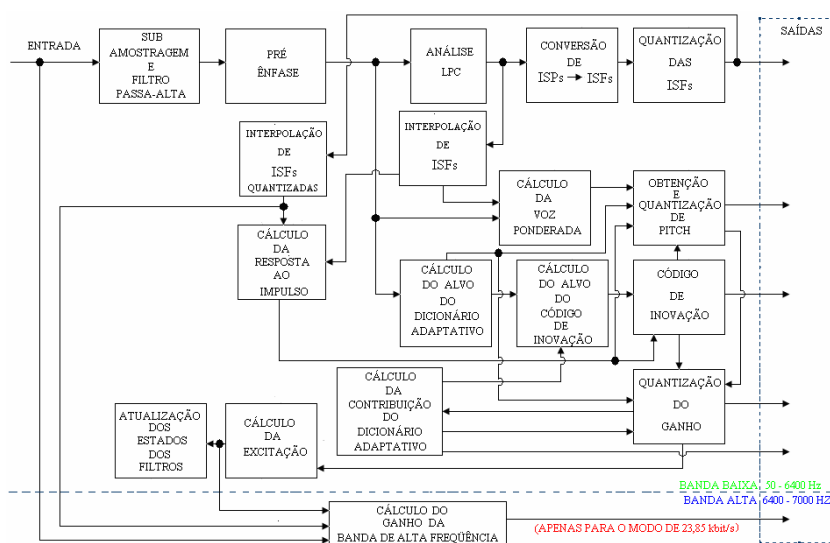


Figura 3.5 – Diagrama de blocos do codificador de voz do AMR-WB

Outra informação bastante importante sobre o codificador são as máscaras de alocação de bits utilizadas nas diversas taxas [61], apresentados na Tab. 3.5. Essa informação dará subsídio a algumas afirmações sobre o sistema de reconhecimento distribuído a ser analisado quando se utiliza o codificador AMR-WB.

Modo	Parâmetro	1 - subquadro	2 - subquadro	3 - subquadro	4 - subquadro	total por quadro
23.85 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	88	88	88	88	352
	Ganho de Código	7	7	7	7	28
	Energia HB	4	4	4	4	16
Total						477
23.05 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	88	88	88	88	352
	Ganhos	7	7	7	7	28
	Total					
19.85 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	72	72	72	72	288
	Ganho de Código	7	7	7	7	28
	Total					
18.25 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	64	64	64	64	256
	Ganhos	7	7	7	7	28
	Total					
15.85 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	52	52	52	52	208
	Ganhos	7	7	7	7	28
	Total					
14.25 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	44	44	44	44	176
	Ganhos	7	7	7	7	28
	Total					
12.65 kbit/s	VAD-flag					1
	ISF					46
	Filtro LTP	1	1	1	1	4
	Retardo de Pitch	9	6	9	6	30
	Código Algébrico	36	36	36	36	144
	Ganhos	7	7	7	7	28
	Total					
8.85 kbit/s	VAD-flag					1
	ISF					46
	Retardo de Pitch	8	5	8	5	26
	Código Algébrico	20	20	20	20	80
	Ganhos	6	6	6	6	24
	Total					
6.60 kbit/s	VAD-flag					1
	ISF					36
	Retardo de Pitch	8	5	5	5	23
	Código Algébrico	12	12	12	12	48
	Ganhos	6	6	6	6	24
	Total					

Tabela 3.5 – Tabela de alocação de bits para o codificador AMR-WB

O fluxo do sinal no decodificador é mostrado na Fig. 3.6. Os índices transmitidos são extraídos do fluxo de bits recebido e, então, decodificados para obtenção dos parâmetros do codificador em cada quadro transmitido. Estes parâmetros são os vetores ISF, os períodos de pitch fracionários, os vetores de código de inovação, os respectivos ganhos de pitch e de inovação e os parâmetros do preditor de período longo (os quais são realizados para a banda baixa de 50 – 6400 Hz). No modo 23.85 kbit/s também são decodificados os índices da banda alta (6400 – 7000 Hz). Os vetores ISF são convertidos para os coeficientes LPC do filtro e interpolados para obter filtros LPC em cada sub-quadro. Então, a cada 64 amostras de sub-quadro:

- a excitação é construída adicionando-se os vetores de código adaptativo e de inovação escalados pelos seus respectivos ganhos;
- a fala é reconstruída filtrando-se a excitação através do filtro LPC de síntese;
- A voz reconstruída passa por um filtro de de-ênfase.

Finalmente, o sinal de fala da banda baixa (50 – 6400 Hz) reconstruído é sobre-amostrado para 16 kHz e o sinal de voz da banda alta é filtrado em (6400 – 7000 Hz) na taxa de 16kHz e adicionado à banda baixa.

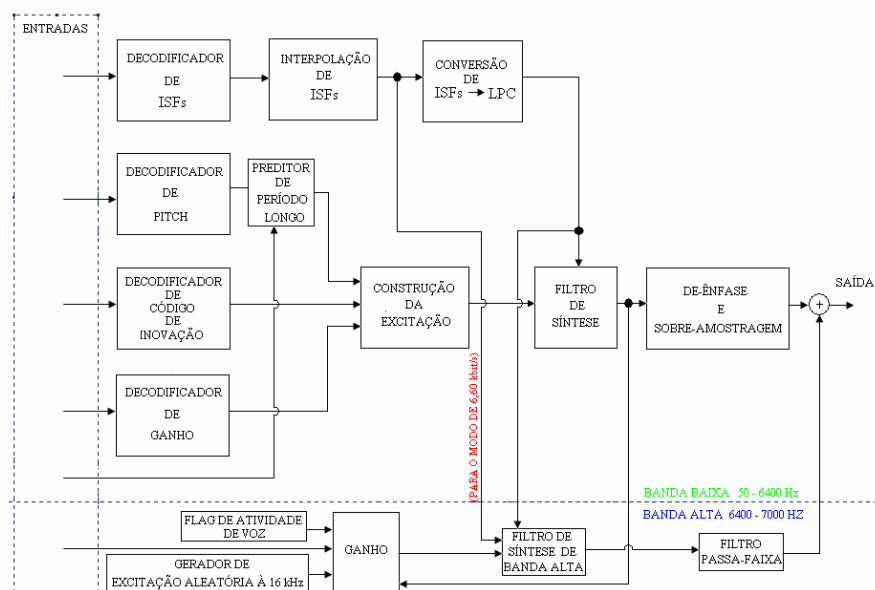


Figura 3.6 – Diagrama de blocos do decodificador de voz do AMR-WB

O código de referência para o codificador AMR-WB na aritmética de ponto flutuante foi definido em [62], porém a norma também implementa o codificador utilizando a aritmética de ponto fixo, conforme definido em [63] para quando o processador utilizado, ou o sistema operacional, não for compatível com a aritmética de ponto flutuante.

3.4. Conclusão

Neste capítulo foi feita a apresentação dos codificadores de voz que serão utilizados na montagem dos sistemas de reconhecimento distribuído no ambiente celular/voz sobre IP.

No capítulo seguinte, será feita a apresentação teórica dos atributos de reconhecimento que serão utilizados para a implementação do sistema de reconhecimento de voz distribuído para o Português Brasileiro com amplo vocabulário.