

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ESCOLA DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO

## **Avaliação de Codificadores de Voz em Ambiente VoIP**

Autor: \_\_\_\_\_  
Rodrigo de Mello Leal Santiago Lamas

Orientador: \_\_\_\_\_  
Prof. Sergio Lima Netto, *Ph.D.*

Examinador: \_\_\_\_\_  
Prof. Luiz Wagner Pereira Biscainho, *D.Sc.*

Examinador: \_\_\_\_\_  
Filipe Castello da Costa Beltrão Diniz, *M.Sc.*

DEL

Dezembro/2005

# **Dedicatória**

Dedico este trabalho aos meus pais (Francisca Maria de Mello Leal Santiago Lamas e Vicinius Santiago Lamas) que nunca deixaram de acreditar em mim, e que desde a minha infância me deram todo o apoio para que chegasse a este momento.

# Agradecimentos

Agradeço:

- À minha mãe, Francisca Maria de Mello Leal Santiago Lamas, por ter sido sempre carinhosa comigo, por ter sempre me apoiado em minhas decisões, pela compreensão de minhas ausências em função dos estudos e por ter me ajudado na formatação deste trabalho tão importante;
- Ao meu pai, Vicinius Santiago Lamas, por toda a dedicação, ajuda em minhas escolhas profissionais e integral apoio em minhas decisões, bem como pela compreensão pelas minhas ausências em função dos estudos;
- Ao meu irmão, Francisco de Mello Leal Santiago Lamas, pela amizade e companheirismo incondicionais;
- À minha tia, Antonina de Mello Leal, por todo carinho despendido ao longo dos anos;
- À minha grande amiga Natasha da Rocha Moura, pelo companheirismo incondicional em todos os momentos da faculdade, desde o primeiro dia de aula, até o momento da conclusão deste projeto;
- Ao meu orientador, professor Sergio Lima Netto, pelas idéias, apoio e compreensão;
- Ao meu orientador acadêmico, professor Luiz Wagner Pereira Biscainho, pelo apoio em minha trajetória acadêmica desde o primeiro período até a conclusão deste curso;

- A todos os meus amigos da turma 2005/2 do curso de Engenharia Eletrônica e de Computação da UFRJ, pelos grandes anos que passamos juntos;
- A todos os meus amigos pelo incondicional apoio e por nunca me deixarem desistir.

## **Resumo**

Neste trabalho será realizada uma apresentação de codificadores de voz e como os mesmos se comportam em ambiente VoIP. Serão mostradas as diferentes famílias de codificadores de voz, bem como representantes das mesmas, as diferentes arquiteturas VoIP existentes no mercado e a importância da qualidade de serviço (QoS) quando da avaliação de codificadores em ambiente VoIP, bem como os parâmetros mais importantes para esta análise.

Além de toda essa abordagem teórica, será implementada uma série de testes seguidos por análises cujo objetivo é chegar a um determinado codificador, ou ainda, a uma determinada família de codificadores que apresente um melhor desempenho, quando expostos às peculiaridades de uma transmissão em ambiente VoIP.

Capítulo 1 - Introdução .....	2
1.1) Introdução .....	2
1.2) Organização do Projeto .....	3
Capítulo 2 – Codificação de Voz .....	5
2.1) Introdução .....	5
2.2) O Problema da Codificação .....	5
2.3) Os Codificadores Forma de Onda .....	7
2.3.1) O PCM .....	8
2.3.2) O ADPCM .....	8
2.4) Os Codificadores Paramétricos .....	10
2.4.1) A Codificação LPC .....	10
2.5) Os Codificadores Híbridos – O CELP .....	13
2.5.1) O Janelamento do Sinal de Voz .....	14
2.5.2) O Modelo do Trato Vocal .....	14
2.5.3) O Filtro de Síntese .....	15
2.5.4) O Filtro de Ponderação (ou Perceptivo) .....	15
2.5.5) Os Dicionários .....	16
2.5.6) Análise por Síntese .....	17
2.5.7) Características do Sistema CELP .....	18
2.6) Os Codificadores Específicos da ITU .....	19
2.6.1) G.711 .....	19
2.6.2) G.726 .....	21
2.6.3) G.728 .....	22
2.6.4) G.729 .....	24
2.6.5) G.723.1 .....	27
2.7) Sumário dos Codificadores .....	28
2.8) Conclusão .....	29
Capítulo 3 - VoIP .....	30
3.1) Introdução .....	30
3.2) O Ambiente VoIP .....	31
3.3) Arquitetura VoIP .....	32
3.4) Protocolo H.323 .....	35
3.5) Protocolo SIP .....	39
3.6) Outras Arquiteturas VoIP .....	42
3.7) Aplicativos VoIP .....	42
3.8) Conclusão .....	46
Capítulo 4 - QoS – Qualidade de Serviço .....	48
4.1) Introdução .....	48
4.2) QoS .....	49
4.3) A Importância Comercial da QoS .....	51
4.4) Parâmetros de Redes com QoS .....	52
4.5) Avaliação de QoS em Sinais de Voz .....	55
4.5.1) P.800 (MOS) .....	55
4.5.2) P.861 (PSQM) .....	57
4.5.3) PSQM+ .....	58

4.5.4) PAMS .....	58
4.5.5) PESQ (P.862).....	59
4.5.6) Mapeamento PESQ x MOS.....	61
4.6) Conclusão .....	62
Capítulo 5 - Testes, Resultados e Análises .....	63
5.1) Introdução .....	63
5.2) Métodos Utilizados.....	64
5.3) Testes .....	65
5.3.1) Atraso.....	66
5.3.2) Jitter .....	68
5.3.3) Perda de Pacotes .....	70
5.4) Testes Subjetivos .....	72
5.5) Conclusão .....	77
Capítulo 6 - Conclusão.....	79
6.1) Resumo do Projeto e Conclusão .....	79
6.2) Propostas de Trabalhos Futuros.....	81
Referências Bibliográficas.....	82

# Capítulo 1 - Introdução

## 1.1) Introdução

A evolução dos sistemas de comunicação ao longo do século XX, principalmente em suas duas últimas décadas, e também no início do século XXI, despertou o interesse pela procura de sistemas que sejam cada vez mais eficientes, ou melhor, capazes de prover serviços de qualidade alta por um preço competitivo. Em outras palavras, cada vez mais se buscam ambientes onde possa ocorrer transmissão de dados de forma mais eficiente (menores custos e consumo de energia) mantendo sempre um padrão de qualidade alto. Podemos destacar como grande exemplo desta busca o rápido crescimento da integração entre dados e voz. Dessa maneira, criou-se a técnica de transmissão VoIP (do inglês, *Voice over IP*), cujo objetivo é utilizar a mesma estrutura (protocolo IP) tanto para a comunicação de dados quanto para a transmissão de voz, através de uma mesma rede convergente.

Este projeto possui objetivos tanto teóricos quanto práticos, sendo o seu tema fundamental a análise do comportamento de diversos codificadores de voz em ambiente VoIP, com ênfase na forma com que os parâmetros do ambiente citado são capazes de alterar a qualidade do serviço (do inglês QoS – *Quality of Service*).

Como colocamos, este projeto pode ser dividido em duas partes, uma teórica e uma prática. O projeto inicia-se com a parte teórica, que tem como objetivo estudar os princípios da codificação, mencionando quais são as estruturas de codificação existentes, quais as vantagens e desvantagens de cada uma delas e quais os principais representantes de cada uma das famílias de codificadores.

Em seguida, destacaremos as principais características de um ambiente VoIP, bem como os principais protocolos de comunicação VoIP atualmente utilizados (H.323 e SIP), mencionando vantagens e desvantagens de cada um deles.



O último parâmetro para o estudo teórico é sabermos como medir a qualidade de serviço fornecida por um determinado sistema de comunicação por voz. Este talvez seja o parâmetro mais importante de nosso estudo. Sabemos que sistemas como a rede pública de telefonia e os PBXs (do inglês, *Private Branch eXchange*) são considerados serviços de alta disponibilidade. Essa é uma necessidade fundamental para o sucesso da migração de tecnologias. Dessa maneira, entendemos que a voz deve ser tratada de forma correta para que a qualidade de serviço para os usuários seja alta, o que ressalta a importância deste parâmetro para a popularização de sistemas VoIP.

No que diz respeito à parte prática foram efetuados alguns testes com diferentes codificadores de voz, variando uma série de parâmetros intrínsecos ao ambiente VoIP como atraso, variação do atraso (*jitter*) e taxa de perda de pacotes. Nossos objetivos práticos são verificar quais codificadores de voz apresentam um comportamento mais amigável no ambiente VoIP e com isso apontar uma tendência para o desenvolvimento de novas normas especificamente construídas para o tal ambiente.

## **1.2) Organização do Projeto**

No Capítulo 2 deste projeto abordaremos primeiramente o problema da codificação. Serão vistas as principais famílias de codificadores existentes atualmente. Procuraremos expor quais são os principais membros de cada uma das famílias, enfatizando suas vantagens e desvantagens relativas. Finalmente, iremos enumerar os codificadores que serão utilizados para análise posterior.

No Capítulo 3, procuraremos abordar o que é o ambiente VoIP, ressaltando a requisição de seus possíveis clientes. Inicialmente vamos abordar os diferentes protocolos IP (TCP/IP, UDP e RTP/RTCP) e as estruturas de *hardware* utilizadas (*gateway*, *gatekeeper* e *IP-phones*). Em seguida, iremos abordar especificamente os protocolos VoIP, destacando as características, vantagens e desvantagens das duas arquiteturas mais utilizadas (H.323 e SIP), bem como mencionando outros protocolos menos utilizados.

Finalmente, iremos mencionar alguns aplicativos VoIP existentes, como o MSN, o Skype e o Google-Talk.

No Capítulo 4, iremos abordar o que é qualidade de serviço, qual a sua importância comercial, quais os parâmetros para sua aferição (atraso, *jitter* e perda de pacotes) e através de quais sistemas e normas a mesma pode ser medida. Ao seu final, iremos descrever um procedimento detalhado para os testes e avaliações que serão descritos no capítulo seguinte.

No Capítulo 5 iremos descrever os testes realizados, apresentaremos os resultados dos mesmos e faremos uma discussão com relação a estes. Na parte final deste capítulo procuraremos estabelecer uma determinada classe de codificadores que apresenta um melhor desempenho no ambiente VoIP.

Finalmente, no Capítulo 6, efetuaremos a conclusão do projeto. Ressaltaremos os principais pontos expostos e desenvolvidos no trabalho, procurando explicitar suas contribuições, sem deixarmos de apresentar sugestões de trabalhos futuros, que possam dar continuidade a este trabalho.

# Capítulo 2 – Codificação de Voz

## 2.1) Introdução

Este capítulo tem como principal objetivo explicitar o problema da codificação e estimular o seu estudo.

Primeiramente, na seção 2.2, iremos analisar os principais pontos do “problema da codificação”.

Em seguida, nas seções 2.3, 2.4 e 2.5, será feita uma explicitação dos codificadores mais utilizados atualmente nos sistemas de comunicação por voz, destacando as principais características, vantagens e desvantagens de cada um deles.

Depois, na seção 2.6, enumeraremos os principais codificadores da ITU (*International Telecommunication Union*) a serem avaliados no decorrer desta dissertação, destacando mais uma vez suas características, vantagens e desvantagens dos mesmos.

A seguir será feito na seção 2.7 um pequeno sumário de todos os codificadores apresentados neste capítulo.

Finalmente, na seção 2.8, faremos uma pequena conclusão para o capítulo, ressaltando os principais tópicos vistos ao longo do mesmo.

## 2.2) O Problema da Codificação

Sabe-se que a transmissão em sistemas telefônicos (sejam eles através de redes tradicionais, celulares ou mesmo através da internet) é bastante custosa. Dessa maneira,

qualquer forma de compressão dos dados a serem transmitidos, desde que não haja perda significativa na qualidade do sinal recebido, é sempre bem-vinda. Para tanto, ao longo do tempo, vêm sendo implementadas diversas soluções (codificadores de voz) que buscam aperfeiçoar a relação entre taxa de transmissão e qualidade.

Dentre as soluções de grande sucesso (e utilizadas atualmente), podemos destacar três grandes classes de codificadores: os codificadores em formato de onda, onde destacamos o codificador PCM (utilizado, por exemplo, no padrão G.711) e suas variantes, utilizados em praticamente todos os sistemas de telefonia fixa no mundo; os codificadores paramétricos, dentre os quais podemos destacar o codificador LPC (utilizado, por exemplo, no padrão FS1015), que, apesar de apresentar uma qualidade precária, apresenta uma taxa de transmissão extremamente baixa; e os codificadores híbridos, dentre os quais se destaca o CELP (utilizado, por exemplo, nos padrões G.729, com as suas variantes, e GSM).

A diferença básica entre a codificação em formato de onda e a codificação paramétrica, de maneira geral, está na filosofia básica utilizada para efetuarla. Codificadores em formato de onda transmitem o sinal de voz propriamente dito (ou ainda alguma variante do mesmo), enquanto que os codificadores paramétricos realizam um “tratamento matemático” e transmite uma série de parâmetros resultantes das operações anteriores.

Em linhas gerais, o resultado da comparação das soluções adotadas é um tanto intuitivo: os codificadores em formato de onda, por transmitirem o sinal de voz propriamente dito, acabam por apresentar uma qualidade muito boa. Porém, devido à ausência de compressão do sinal (manipulação matemática), as taxas de transmissão apresentadas por tais codificadores são bastante altas. Em contrapartida, os codificadores paramétricos, por realizarem uma manipulação matemática do sinal de voz a ser transmitido e transmitirem apenas os parâmetros resultantes de tal manipulação, apresentam taxas de transmissão muito baixas, mas geram no receptor sinais de qualidade precária.

Já os codificadores híbridos procuram, até certo ponto, unir as boas características de cada um dos sistemas anteriores, objetivando apresentar uma boa qualidade de codificação/decodificação a taxas relativamente baixas.

De um modo geral os codificadores podem ser descritos pelas informações contidas na tabela 2.1. Um detalhamento das três famílias citadas será feito nas seções a seguir.

**Tabela 2.1:** Comportamento dos codificadores com relação à taxa de codificação e qualidade.

<b>Tipo de Codificador</b>	<b>Taxa de Codificação</b>	<b>Qualidade</b>
Forma de Onda	32 a 64 kbits/s	Boa – Excelente
Paramétrico	1 a 4 kbits/s	Ruim - Regular
Híbrido	3 a 12 kbits/s	Boa - Excelente

### **2.3) Os Codificadores Forma de Onda**

Tais codificadores utilizam para a codificação do sinal de entrada propriedades de caráter temporal e/ou espectral do mesmo. Porém, isso não pressupõe analisar o que realmente o sinal está representando, o que significa que tais codificadores apenas se concentram na necessidade de reconstruir a forma de onda a ser codificada. Dessa maneira, qualquer tipo de sinal pode vir a ser codificado utilizando este método.

Entre tais codificadores podemos citar o PCM, o ADPCM, ambos já citados na seção anterior. Tais codificadores apresentam uma qualidade muito boa, mas acompanhada de taxas extremamente elevadas. Para sinais de voz, tais codificadores só devem ser escolhidos para taxas superiores a 16 kbits/s, uma vez que para taxas inferiores a essa, existem soluções mais atrativas, como os codificadores paramétricos.

### **2.3.1) O PCM**

PCM é uma sigla inglesa que significa *Pulse Code Modulation*. Podemos classificar esta forma como a mais básica de se codificar qualquer espécie de sinal da forma digital. Tal codificação baseia-se em atribuir a cada amostra um nível discreto de amplitude. O número de níveis discretos é dado como a  $n$ -ésima potência de dois, sendo  $n$  o número de bits disponíveis para a quantização. Isso significa que caso tenhamos 10 níveis distintos de quantização, poderemos ter 1024 níveis discretos, sendo cada uma das amostras representada pela forma binária de sua amplitude.

Como já foi visto, a codificação PCM apresenta a melhor qualidade dentre os codificadores utilizados, mas acompanhada de uma alta taxa de codificação. A fim de exemplificarmos isto, podemos mencionar a telefonia digital fixa.

Na telefonia digital fixa utiliza-se uma taxa de amostragem de 8 kHz (devido a propriedades dos sistemas de fala e auditivo do ser humano) e uma resolução de 8 bits. Multiplicando os dois números obtemos a taxa de codificação de 64 kbits/s, o que é uma taxa relativamente alta para ser usada em ambientes de transmissão digital como a internet e a rede de telefonia celular.

Porém, o sistema PCM se revela extremamente eficiente em termos de qualidade quando codificando sinal de voz. Isso nos leva a tratá-lo como uma referência no que se refere à taxa de codificação. Dessa maneira, se falarmos de um sistema que codifique dados de voz a uma taxa de 16 kbits/s, dizemos que o mesmo possui uma taxa de compressão de 4:1.

### **2.3.2) O ADPCM**

Como já vimos o sistema PCM é extremamente eficiente para codificação de sinais de voz. Contudo, a sua utilização é restrita, devido à alta taxa de codificação. Isso

sugere que se procurem algumas formas de se codificar sinais de voz em baixas taxas. Para realizarmos tal operação com os codificadores por forma de onda, exploramos as características estatísticas dos sinais de voz e também as características do sistema auditivo humano. Mais especificamente, devemos ter dois objetivos em mente:

- A máxima remoção de redundâncias do sinal de voz;
- A alocação dos bits disponíveis para codificar a parte não-redundante do sinal de voz.

Porém, ao passo que tentamos baixar as taxas de codificação, os sistemas utilizados para remoção de redundância e eficiente atribuição dos bits passam a ter sua complexidade computacional muito elevada. Para uma mesma qualidade de sinal de voz, cada vez que dividimos por dois o número de bits a serem utilizados, a complexidade computacional cresce aproximadamente de uma ordem de grandeza [5].

Dessa maneira, devemos buscar soluções intermediárias. Isso levou ao desenvolvimento da tecnologia ADPCM (do inglês, *Adaptive Differential Pulse Code Modulation*), que permite a codificação a taxas inferiores a 64 kbits/s, sendo que o número de bits por amostra é que é reduzido, ao passo que se mantém a frequência de amostragem. O número de bits por amostra pode ser reduzido de 8 para 7, 6, 5, 4, 3 ou 2, sendo as taxas reduzidas respectivamente para 56, 48, 40, 32, 24 ou 16 kbits/s, ao custo de se aumentar a complexidade computacional. Cabe ao projetista decidir qual a melhor estratégia a ser adotada.

O sistema ADPCM baseia-se no uso combinado da quantização adaptativa e da predição adaptativa, o que é suficiente (e necessário) para que se atinja o melhor desempenho possível, independentemente da qualidade do sinal de voz original e da qualidade do material eletrônico utilizado (microfones, alto-falantes, caixas de som etc).

Tal sistema obteve um sucesso bastante grande, tanto que é utilizado até hoje na telefonia digital fixa nos EUA. Maiores detalhes do funcionamento do ADPCM podem ser encontrados em [17].

## **2.4) Os Codificadores Paramétricos**

Esta segunda classe de codificadores trata daqueles que utilizam, no processo de codificação, determinadas características da fonte de sinal representada. Então, quando codificando sinais de voz, deve-se fazer um mapeamento (estudo) detalhado do trato vocal do ser humano, bem como das propriedades da natureza da voz humana, levando em consideração características como idade, sexo, timbre etc.

Dessa maneira, ao contrário dos codificadores forma de onda, os codificadores paramétricos fazem uso de aspectos específicos, intrínsecos à voz humana, o que os torna específicos para cada tipo de sinal a ser codificado.

Entre tais codificadores podemos destacar o LPC (do inglês *Linear Predictive Coding*), utilizado normalmente no padrão FS1015. Esse tipo de codificação será especificado na subseção a seguir.

### **2.4.1) A Codificação LPC**

O LPC é o mais importante dos codificadores da família dos codificadores paramétricos. Como é característica dos membros desta família, o LPC se baseia em uma série de parâmetros inerentes à voz humana.

De modo geral, processos estocásticos estacionários podem ser interpretados como a saída de um filtro digital onde a entrada é um modelo para o ar que sai dos pulmões e atravessa as cordas vocais e o filtro propriamente dito é dado pelo trato vocal.



O trato vocal é a parte do corpo humano que vai desde a glote até os lábios, passando pela epiglote, laringe, faringe, garganta, língua e dentes.

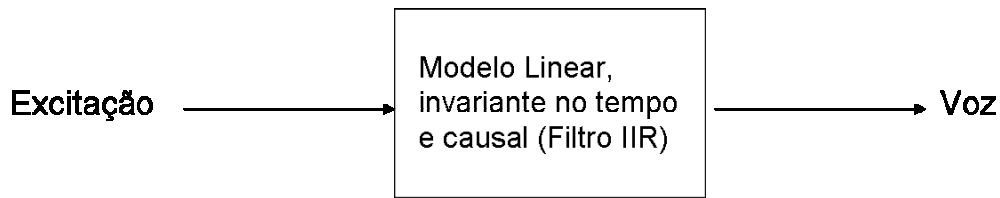
Uma vez tendo definidos os elementos que compõem o nosso sistema (entrada, sistema propriamente dito e saída) podemos classificar os sinais de voz (sinais de saída) em dois tipos: falas sonoras (sinais vozeados) e falas surdas (sinais não-vozeados).

Os trechos sonoros são aqueles que, quando emitidos, geram vibração das cordas vocais, sendo os mesmos modelados por um trem de pulsos praticamente periódico. O período relacionado com a sua frequência fundamental é denominado período de *pitch*, um parâmetro determinante no projeto de codificadores atuais. Tais sons podem ser exemplificados através daqueles relativos às vogais.

Os sons surdos ou não-vozeados, ao contrário dos sonoros ou vozeados, são aqueles que não geram vibração das cordas vocais. Dessa maneira, são modelados como ruído branco, apresentando como característica marcante a taxa de cruzamento por zero significativa. Sons surdos podem ser exemplificados como som da letra “s” na palavra “sapo” ou como o som do encontro consonantal “ch” na palavra “chocalho”.

Ainda podemos considerar um terceiro tipo de som que seria uma classificação intermediária entre as duas anteriores. Chamamos esses sons de plosivos, pois os mesmos têm como excitação um único pulso de ar. Como exemplo, podemos citar as consoantes “p” (som plosivo vozeado) e “b” (som plosivo não-vozeado).

A voz é um sinal fundamentalmente não estacionário e não periódico o que nos levaria a uma análise muito complicada. Porém, se considerarmos um intervalo de tempo (janela) relativamente pequeno (de 10 a 30 ms), o sinal de voz pode ser considerado estacionário em cada uma dessas janelas. Dessa maneira, para cada um destes pequenos intervalos, podemos interpretar a geração da voz através do modelo colocado na figura 2.1 [1].



**Figura 2.1:** O processo de geração de voz através do trato vocal

O trato vocal é modelado por um filtro digital do tipo IIR (*Infinite Impulse Response*) que possui os zeros na origem e cujos pólos são determinados pela técnica da regressão linear. A cada uma das janelas os pólos (e conseqüentemente os coeficientes) do filtro são atualizados. A essa análise damos o nome de “Análise LPC”.

Para a geração de sons sonoros no codificador, utiliza-se um trem de pulsos com período igual ao período de *pitch* da voz. Para sons surdos utilizamos como excitação uma seqüência de ruído branco, sem nunca deixarmos de estimar um ganho para o modelo.

Dessa maneira, deixamos de transmitir as amostras do sinal de voz quantizadas, passando a ser transmitidos os coeficientes determinados pela análise LPC, um sinal de *flag* que indica se o som é sonoro ou surdo, o ganho estimado e, caso o som seja sonoro, o *pitch*. Com isso, passamos a transmitir uma quantidade extremamente inferior de parâmetros com relação aos codificadores forma de onda o que diretamente resulta em uma taxa de codificação bem inferior. Se considerarmos o PCM padrão, cuja freqüência de amostragem é de 8 kHz, e janelas de 20ms, teríamos que transmitir 160 valores por janela, enquanto que no LPC, supondo um filtro de 10ª ordem, transmitiríamos 13 valores (10 correspondentes aos coeficientes, 1 ao ganho, 1 ao *flag* e 1 ao *pitch*).

No entanto, apesar de esse tipo de codificador apresentar uma redução de taxa de codificação, ele possui uma grande restrição. Como existe uma decisão do tipo binária com relação ao sinal ser ou não vozeado, não há espaço para transmissão de sons intermediários. O resultado disso é o sinal codificado de voz ser um tanto “robótico”. Esta é uma característica marcante não só do codificador LPC, mas de todos os

codificadores de fonte, o que é uma grande desvantagem. Porém, sua baixa taxa de codificação é uma grande vantagem. Como os codificadores por forma de onda apresentam um comportamento antagônico em relação aos vistos nessa seção, devemos procurar sistemas que reúnam e maximizem os pontos fortes de cada uma das famílias. Tais sistemas se encontram nos codificadores híbridos, objeto de estudo da próxima seção.

## **2.5) Os Codificadores Híbridos – O CELP**

Esta família de codificadores se caracteriza pela união de características das duas famílias anteriores. Em linhas bem gerais, os codificadores híbridos mantêm a parametrização dos codificadores paramétricos, enquanto geram a excitação pelo formato de onda. Com o auxílio de dicionários é possível determinar a melhor excitação, o que permite que obtenhamos taxas de codificação bem reduzidas mantendo uma qualidade bem superior a dos codificadores paramétricos e comparável com a dos codificadores forma de onda.

A técnica mais usada neste tipo de codificador é a CELP (do inglês *Code Excited Linear Prediction*). Esse sistema reúne os mesmos princípios da análise LPC para redução de parâmetros a serem transmitidos. Porém, como característica dos codificadores híbridos, o CELP lança mão de um artifício para anular a desvantagem do codificador LPC. De fato, para manipular as excitações da entrada do sistema, o CELP faz o uso de dicionários, tanto fixos quanto adaptativos (o que será explicitado posteriormente). Sendo assim, um número bem maior de excitações (com relação ao LPC) pode ser utilizado, o que significa que uma gama maior de sinais de voz pode ser reconstituída. Isso torna o sistema bastante mais completo e capaz de gerar um sinal de voz (após a transmissão) de qualidade bem superior àquele gerado pelo LPC.

Existe uma série de codificadores baseados na técnica CELP disponíveis no mercado e na literatura, tal grande é a eficiência desse tipo de codificação. Como

exemplos, podem ser citados alguns codificadores do ITU que serão descritos posteriormente neste capítulo. Sendo assim descrever com alguma riqueza de detalhes cada um dos blocos desse sistema é de grande importância. Para tanto, iremos descrever, nas próximas subseções, os módulos do sistema CELP desenvolvido e descrito em [1].

### 2.5.1) O Janelamento do Sinal de Voz

Como já foi visto, a voz humana possui uma natureza extremamente não estacionária, o que nos leva a necessidade de segmentá-la, ou melhor, que sejam considerados intervalos relativamente pequenos do sinal de voz, individualmente.

No caso do sistema considerado, as janelas possuem 20 ms, o que, a uma taxa de amostragem de 8 kHz, representa 160 amostras, sendo utilizada a janela de *Hamming* [2]. Por motivos a serem explicados mais adiante, cada um dos blocos de 20 ms será dividido em quatro sub-blocos de 5 ms cada.

### 2.5.2) O Modelo do Trato Vocal

Com a análise LPC, são calculados (para cada uma das janelas de 20 ms) os coeficientes do filtro digital que melhor representa o trato vocal do usuário no instante em questão. Isso é feito da seguinte forma [3].

- Calcula-se a matriz de autocorrelação da janela do sinal de voz, que denominaremos  $R_s$ . Nesta matriz  $r_s(i)$  representa a correlação do sinal com atraso de  $i$  amostras e  $M$  representa o número total de amostras do sinal.

$$R_s = \begin{bmatrix} r_s(0) & r_s(1) & r_s(2) & \dots & r_s(M-1) \\ r_s(1) & r_s(0) & r_s(1) & \dots & r_s(M-2) \\ r_s(2) & r_s(1) & r_s(0) & \dots & r_s(M-3) \\ \dots & \dots & \dots & \dots & \dots \\ r_s(M-1) & r_s(M-2) & r_s(M-3) & \dots & r_s(0) \end{bmatrix}$$

- Monta-se a seguinte equação, onde  $\hat{a}$  é o vetor dos coeficientes de predição linear que desejamos calcular.

$$R_s \times \begin{bmatrix} \hat{a}(1) \\ \hat{a}(2) \\ \hat{a}(3) \\ \dots \\ \hat{a}(M) \end{bmatrix} = \begin{bmatrix} r_s(1) \\ r_s(2) \\ r_s(3) \\ \dots \\ r_s(M) \end{bmatrix} \rightarrow R_s \times \hat{a} = r_s$$

- Resolve-se a equação anterior pelo algoritmo de Levinson-Durbin, descrito também em [3].

### 2.5.3) O Filtro de Síntese

Na predição linear, o filtro de síntese  $H(z)$  é da seguinte forma:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}}$$

A constante  $p$  tem como função representar a ordem do modelo LPC usado e identifica a precisão com a qual o trato vocal será modelado através de  $H(z)$ . Normalmente, os codificadores CELP utilizam  $p=10$ , o que estabelece um bom compromisso entre a qualidade do sinal de saída e a taxa de codificação[4]. Os valores  $a_i$  representam os coeficientes de predição linear, obtidos através da análise LPC.

### 2.5.4) O Filtro de Ponderação (ou Perceptivo)

Um estudo realizado de acordo com [4] verificou que o ouvido humano é mais sensível a erros e ruídos em componentes de mais baixa amplitude do que em componentes de alta amplitude, no domínio da frequência. Isso nos induz a dar mais importância às componentes de menor amplitude quando calculamos o erro contido no processo de análise por síntese.

Sendo assim, temos que a função do filtro de ponderação é enfatizar as componentes de mais baixa amplitude, ao passo que realiza o oposto com as componentes de maior amplitude. Este filtro é denotado  $W(z)$ , e sua função de transferência é dada pela seguinte forma:

$$W(z) = \frac{A(z)}{A\left(\frac{z}{\gamma}\right)},$$

onde  $\gamma$  representa o fator de percepção, que possui valor típico de 0,8.

No entanto, nos sistemas CELP, é utilizada uma cascata entre o filtro perceptivo e o filtro de síntese, com a intenção de acelerar o processamento.

### 2.5.5) Os Dicionários

Dicionários (ou, em inglês, “*codebooks*”) são conjuntos de excitações que podem ser descritos da seguinte forma:

$$C = \{\{X_0(n)\}, \{X_1(n)\}, \{X_2(n)\}, \dots, \{X_{n-1}(n)\}\},$$

onde cada um dos termos é uma seqüência de um processo estocástico de média zero. Através desta representação também podemos verificar que o dicionário armazena  $K$  seqüências  $X_I(n)$ , onde  $I$  é o índice de cada uma.

No caso do CELP, geralmente são utilizados dois dicionários, um com as excitações fixas e o outro com as excitações adaptativas.

Em geral, o dicionário fixo tem as suas seqüências geradas aleatoriamente, e para reduzirmos os cálculos, as amostras das mesmas sofrem um processo de *clipping* para simplificação dos cálculos, o que significa que as amostras abaixo de um determinado valor são zeradas.

Já o dicionário adaptativo tem as suas seqüências atualizadas a cada bloco de 5 ms, o que é realizado através de uma malha de realimentação. Essas atualizações se dão baseadas na soma das melhores excitações dos dicionários fixo e adaptativo. A cada sub-bloco, a melhor excitação encontrada é incluída no dicionário adaptativo que tem seu conteúdo mais antigo descartado, estando, este dicionário, inicialmente zerado por convenção.

Finalmente, cada um dos dicionários deve ter o seu ganho calculado, sendo que ambos podem ser calculados da mesma maneira, segundo a fórmula a seguir:

$$G = \frac{Corr_{Sinal-Alvo, Re\ posta-Dic}}{Corr_{Re\ posta-Dic, Re\ posta-Dic}}$$

onde o numerador representa a correlação entre o sinal-alvo (sinal de voz com o qual deve ser feita a análise por síntese) e a resposta contida no dicionário em questão, e o denominador representa a autocorrelação da resposta citada.

### **2.5.6) Análise por Síntese**

A excitação que deve ser utilizada para reconstituir a voz do usuário na saída é determinada por um processo chamado “Análise por Síntese”. Tal processo é realizado a cada sub-bloco de 5 ms, uma vez que as características da excitação variam mais

rapidamente que as características do trato vocal, o que impõe esse requisito ao sistema [4].

O procedimento é o seguinte:

- Para cada sub-bloco, cada uma das excitações contida no dicionário passa pelo filtro calculado pela análise LPC, gerando uma determinada resposta;
- Subtrai-se essa resposta do sinal de voz na referência;
- Armazena-se a norma quadrática da diferença obtida no passo anterior;
- A excitação que gerar a menor norma quadrática da diferença será escolhida para reconstituir o sinal na saída.

### 2.5.7) Características do Sistema CELP

O funcionamento do sistema CELP descrito nas subseções anteriores é plenamente descrito em [1]. Além disso, em [1] também são propostas uma série de melhorias para o sistema CELP, e o seu funcionamento com tais melhorias é detalhado.

No sistema CELP, são transmitidos apenas o índice da melhor excitação, o ganho do filtro e seus respectivos coeficientes, ao invés de todas suas amostras quantizadas, como por exemplo, no sistema PCM. A alocação dos bits (como utilizada em [1]) está demonstrada pela Tabela 2.2.

**Tabela 2.2:** Características do sistema CELP.

<b>Parâmetro</b>	<b>Faixa</b>	<b>Número de Bits</b>
<b>Ganho do Dicionário Fixo (Gf)</b>	-0,05 a 0,05	5 bits X 4



<b>Índice do Dicionário Fixo (I)</b>	0 a 511	9 bits X 4
<b>Ganho do Dicionário Adaptativo (Ga)</b>	0 a 2	4 bits X 4
<b>Índice do Dicionário Adaptativo (L)</b>	0 a 511	9 bits X 4
<b>Coefficientes do Filtro de Síntese</b>		32 bits
<b>Total</b>		<b>140 bits</b>

Dessa maneira, podemos verificar um ganho promovido pelo uso do sistema CELP, uma vez que a taxa de codificação é diminuída dos 64 kbits/s do PCM para 7 kbits/s. Sendo assim, este codificador será considerado para os testes que serão desenvolvidos ao longo deste trabalho. Já os demais codificadores a serem considerados serão descritos na próxima seção.

## 2.6) Os Codificadores Específicos da ITU

Ao longo das seções anteriores viemos colocando os tipos de codificação mais utilizados. Para cada um dos tipos, a ITU (*International Telecommunication Union*), desenvolveu uma série de recomendações.

A seguir descreveremos as seguintes recomendações: G.711, G.726, G.728, G.729 e G.723.1. Procuraremos abordar os aspectos históricos, os aspectos funcionais e as atualizações (anexos) que vieram surgindo ao longo do tempo (para cada uma das recomendações) com o intuito de melhorar os sistemas originais.

### 2.6.1) G.711

O G.711 [16] foi aprovado no início da década de 70, mais precisamente em 1972. Ele consiste na codificação PCM e apresenta uma taxa de codificação de 64 kbits/s. A norma não definiu um escopo preciso, mas a mesma foi concebida de modo a ser uma forma de digitalizar o sinal de voz para o mesmo ser tratado da forma mais eficiente possível pelos sistemas de comunicação digital.

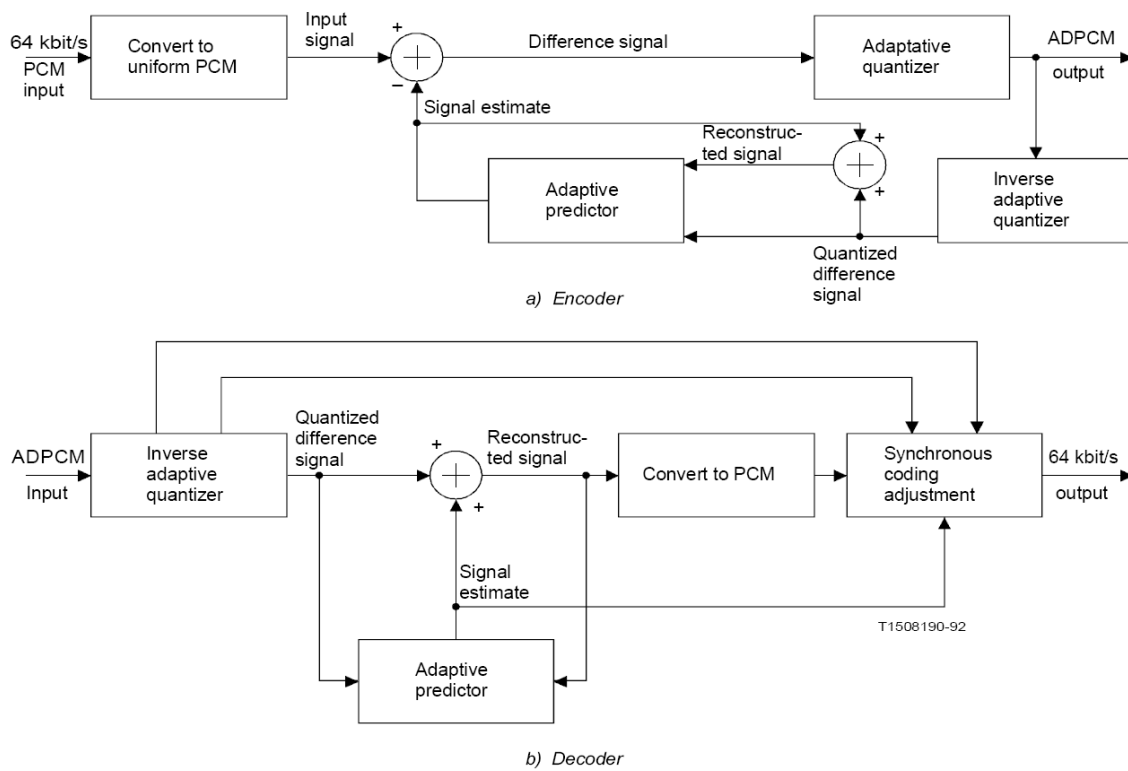
Como no G.711 são realizadas 8.000 amostras por segundo, o atraso gerado pelo algoritmo, para cada informação de voz digital (um octeto), é de 125  $\mu$ s. Na prática, podemos agrupar os bits em pacotes, aumentando ligeiramente o atraso.

Como sabemos, o princípio de Nyquist deve ser levado em consideração em qualquer sistema que sofra amostragem de seus dados contínuos. Dessa maneira, levando-se em conta parâmetros dos sistemas auditivo e de fala do ser humano, estabelece-se que nos sistemas telefônicos uma frequência de cerca de 3,3 kHz é suficiente para boa compreensão. Com isso, a fim de se respeitar o critério de Nyquist e de não “sufocar” as especificações de outros módulos do sistema (como o filtro *anti-aliasing*, por exemplo), escolhe-se uma taxa de amostragem de 8 kHz. Isso significa que a cada segundo teremos 8.000 amostras, que em seguida são quantizadas em 256 níveis, com o uso de 8 bits para tanto. Com isso, o canal imediatamente demanda 64 kbits/s de banda para transmissão.

O processo de quantização refere apenas à atribuição de valores discretos aos níveis de amplitude contínuos. Conseqüentemente verifica-se a existência de um erro (ou ruído) de quantização, dado pela diferença entre os valores do sinal na saída do quantizador e na entrada do mesmo. A fim de que tais erros sejam percentualmente próximos, independentemente da ordem de grandeza da amplitude, os níveis de quantização têm espaçamento exponencial. Para tanto, como o G.711 foi e é largamente utilizado, o mesmo prevê o uso de duas curvas para o tratamento dos erros acima mencionados. Estas curvas são as leis A (inicialmente usada no Brasil e na Europa) e  $\mu$  (desenvolvida posteriormente nos EUA e lá adotada como padrão), conforme mostrado em [5].

## 2.6.2) G.726

O G.726 é uma recomendação que data de dezembro de 1990. O seu princípio de funcionamento se baseia na utilização da codificação ADPCM (descrita na seção anterior), para a redução do número de bits a serem codificados e transmitidos no canal.



**Figura 2.2:** Codificador e decodificador da recomendação G.726 [16].

O funcionamento do G.726 pode ser descrito através da figura 2.2, que mostra os blocos de codificação e de decodificação do sistema.

Primeiramente, verificamos que o sinal PCM de entrada é convertido, através da lei A ou da lei  $\mu$ , para PCM uniforme. Em seguida, o sinal convertido é inserido em uma

malha de realimentação, sofre os efeitos combinados da predição e da quantização adaptativa e em seguida é transmitido.

No decodificador existe um processo exatamente inverso que, ao seu final, reconverte o sinal para PCM. Deve-se notar aqui que no processo de decodificação há a inserção de um bloco de sincronismo. Esse bloco contribui para que sejam evitadas distorções oriundas de conversões entre PCM e ADPCM, conexões diversas etc. Esse sincronismo é alcançado através do ajuste dos códigos PCM de saída de uma forma que almeja eliminar as distorções de quantização no próximo estágio de codificação ADPCM.

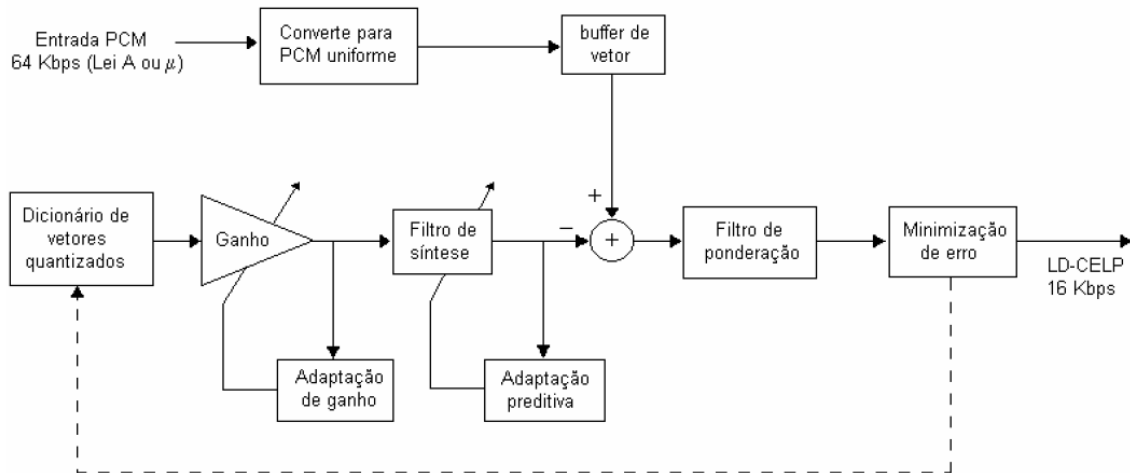
Finalmente, devemos notar que o G.726 foi concebido para realizar a codificação ADPCM com diversas taxas de codificação, dependendo da aplicação a ser utilizada. Podem ser utilizados de 2 a 5 bits para a quantização, sendo que sempre 1 bit é reservado para o sinal da amplitude. Dessa maneira, taxas de 16 a 40 kbits/s podem ser utilizadas.

Para o nosso trabalho, iremos utilizar o G.726 com taxa de compressão 2:1, ou de 32 kbits/s, lançando mão de 4 bits para quantização. Escolhemos este valor por ser o mais comum na literatura e o utilizado na maior parte dos sistemas que utilizam ADPCM.

### **2.6.3) G.728**

O G.728 é uma recomendação relativamente moderna, datando de 1º de setembro de 1992. Ela utiliza como tipo de codificação a LD-CELP (do inglês *Low-Delay Code Excited Linear Prediction*), que é uma variação do sistema CELP original. Ela apresenta um consumo de banda de 16 kbits/s.

Entrando no aspecto específico da codificação, verificamos que o codificador trabalha com blocos de cinco amostras PCM. Como já sabemos, cada uma dessas amostras possui um atraso de 125  $\mu$ s, o que nos leva a concluir que o atraso do algoritmo dessa recomendação é de apenas 625  $\mu$ s.



**Figura 2.3:** Diagrama de blocos simplificado da codificação referente à recomendação G.728 [18].

Com auxílio da figura 2.3 podemos entender melhor como tal recomendação funciona. Inicialmente, temos uma conversão do sinal de entrada de PCM lei A (ou lei  $\mu$ ) para PCM uniforme. Em seguida, o mesmo é agrupado em blocos de cinco amostras consecutivas. Cada um desses blocos é submetido a uma comparação com cada um dos 1.024 vetores armazenados no dicionário de vetores quantizados, tendo estes vetores atravessado as unidades de ganho e o filtro de síntese. A comparação que apresentar o menor erro indicará qual índice do dicionário a ser transmitido. Dessa maneira, sabendo que temos 8.000 amostras por segundo, que as mesmas são agrupadas em grupos de 5 amostras e que o índice do dicionário é formado por 10 bits, chegamos a uma taxa de codificação de 16 kbits/s.

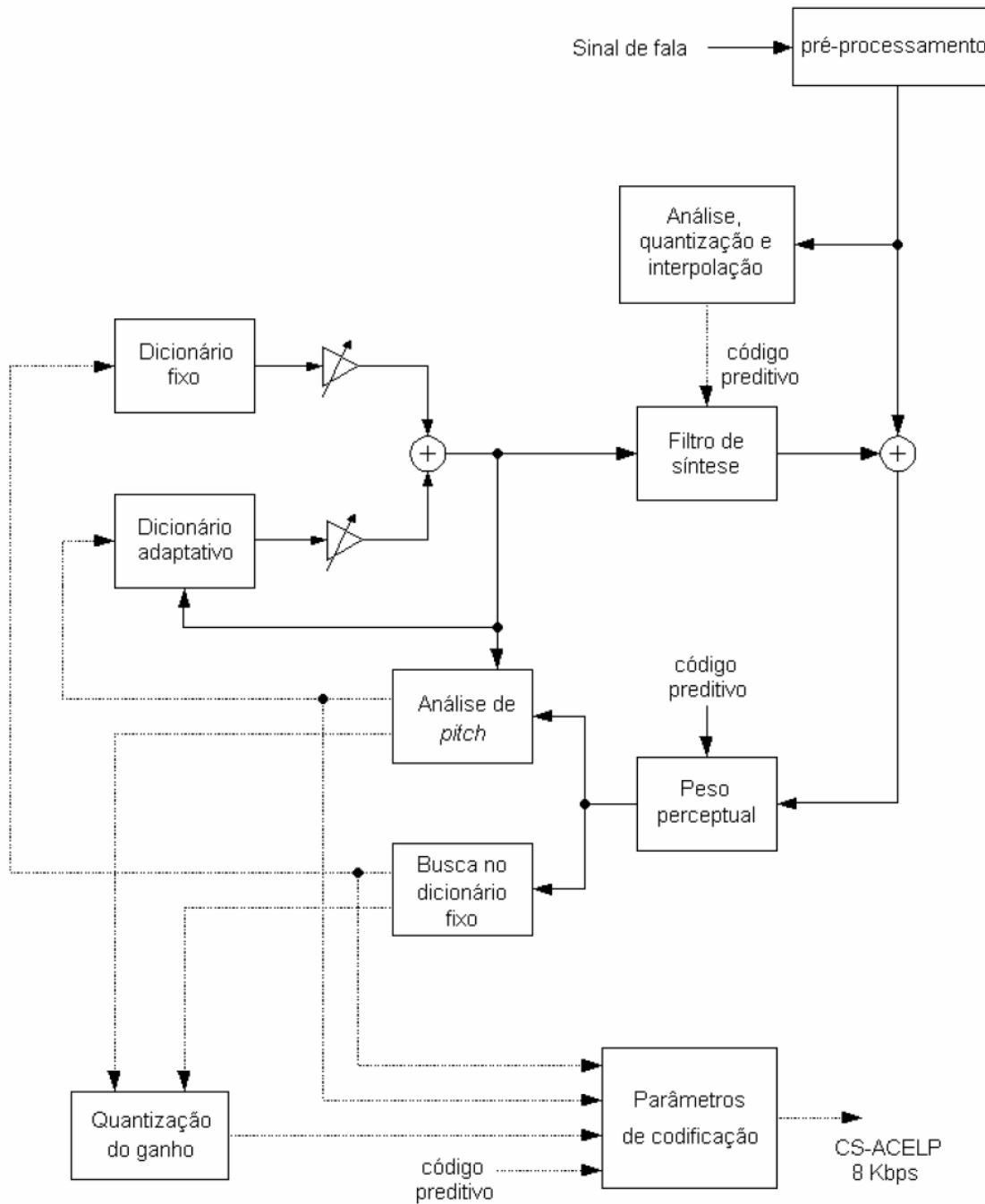
É importante notar também que três parâmetros são atualizados constantemente (periodicamente). Estes são o ganho e os coeficientes dos filtros de ponderação e de síntese, sendo tais parâmetros derivados do vetor imediatamente anterior ao corrente. O ganho é atualizado a cada vetor e os coeficientes a cada 4 vetores (ou 20 amostras, ou ainda 2,5ms).

#### **2.6.4) G.729**

O G.729 é uma recomendação posterior ao G.728, e também utiliza uma variante do sistema CELP tradicional. A mesma foi aprovada em 19 de março de 1996, lançando mão da codificação CS-ACELP (*Conjugate-Structure Algebraic-Code Excited Linear Prediction*) e apresenta uma taxa de codificação duas vezes menor que o G.728, ou seja, de 8 kbits/s.

Tal recomendação foi concebida para transmitir sinais de voz com qualidade em ambientes onde baixas taxas de codificação são de extrema importância, como por exemplo, aplicações de comunicação sem fio e circuitos transoceânicos.

Com relação ao atraso o G.729 codifica os sinais de áudio em janelas de 10 ms, que é sempre sucedido de um tempo de *look-ahead* de 5 ms, o que resulta em um atraso total de 15 ms para essa recomendação.



**Figura 2.4:** Diagrama de blocos simplificado da codificação referente à recomendação G.729 [14].

O codificador em questão foi projetado para operar com sinais de entrada obtidos através da filtragem de sinais analógicos, conforme exposto na recomendação G.712 do

ITU. Depois de uma discretização à taxa de 8.000 amostras em um segundo, o sinal é convertido em PCM linear de 16 bits, servindo assim de entrada para o codificador (na saída realiza-se a operação inversa).

Este codificador, como já foi dito, baseia-se no modelo CELP de codificação, assim como o G.728. Opera com quadros de 10 ms, o que corresponde a 80 amostras do PCM. Em cada um desses quadros, o sinal de voz é analisado para a obtenção dos parâmetros inerentes ao modelo CELP (ganho, índices dos dicionários fixo e adaptativo e coeficientes dos filtros de síntese e de ponderação). Em seguida há a codificação de tais dados, seguida da transmissão. Esse procedimento resulta em 80 bits por janela amostrada, resultando na taxa de codificação de 8 kbits/s.

Apesar do G.729 de ser uma recomendação extremamente eficiente em termos de taxa de codificação, seus requisitos computacionais são extremamente elevados, inclusive superiores inclusive aos do G.723.1, que será visto a seguir. Dessa maneira, em maio de 1996 foi lançada o Anexo A da recomendação, mantendo a operabilidade com o G.729 original e reduzindo a sua complexidade computacional. A operação do G.729A é bastante similar à do G.729, sendo as principais alterações referentes à forma de busca nos dicionários e à forma de operação de cada um dos filtros. Coloquialmente, o G.729A ficou conhecido como G.729 de baixa complexidade.

Em outubro de 1996, foi homologado o Anexo B do G.729. Tal anexo descreve o gerador de ruído de conforto e o detector de voz, utilizados na implementação da compressão de silêncio, tanto no G.729 quanto no G.729A.

Uma série de outros anexos para o G.729 já foram publicados após a respectiva homologação. Porém, para os fins deste estudo, os anexos citados acima são os de maior importância para a realização do mesmo, o que significa que iremos nos restringir aos dois primeiros anexos desta recomendação.



### 2.6.5) G.723.1

Aprovada em março de 1996, esta recomendação, diferentemente das anteriores, especifica uma determinada codificação a ser utilizada para compressão de sinal de voz ou sinal de áudio de um serviço multimídia qualquer para meios de baixíssima velocidade de transmissão.

Os dois tipos de codificação utilizados também são o ACELP (do inglês, *Algebraic-Code Excited Linear Prediction*) com taxa de 5,3 kbits/s e o MP-MLQ (do inglês, *Multi-Pulse Maximum Likelihood Quantization*) com taxa de 6,3 kbits/s.

Com relação ao atraso, independentemente de qual das duas velocidades está se usando, são necessários 30 ms para a formação de cada uma das janelas, além de 7,5 ms de *look-ahead*, gerando um atraso total de 37,5 ms.

O principal requisito dessa recomendação é garantir que as duas velocidades (taxas) estejam disponíveis em qualquer momento. Para tanto podem ser usadas duas operações. A primeira consiste em simplesmente trocar as velocidades entre um quadro e outro, enquanto a segunda baseia-se em utilizar os períodos de descontinuidade de transmissão (nos intervalos de silêncio) para efetuar a troca.

Com relação ao processo de codificação do sinal de entrada, o procedimento adotado é idêntico àquele adotado no G.729 e explicitado na subseção anterior.

Internamente, o G.723.1 se baseia no princípio da análise do sinal para síntese do resultado, visando a minimizar o peso percentual do erro. A operação se dá em blocos de 240 amostras, que são obtidas através do enquadramento, a cada 30 ms, das 8.000 amostras por segundo. Cada um desses quadros é submetido a um filtro passa-altas (para remover eventuais componentes DC) e em seguida é dividido em sub-quadros de 60 amostras cada.

Já o processo de codificação propriamente dito é bastante similar àquele descrito na recomendação G.729. No caso do G.723.1, uma janela deslizante de observação com tamanho de 180 amostras é centrada em cada um dos sub-quadros. Dessa maneira, quatro grupos de parâmetros serão gerados; tais parâmetros (referente aos valores de ganho, índices de dicionários e coeficientes de filtros utilizados) serão agrupados, codificados e transmitidos.

Quando chegamos ao último sub-quadro de um determinado quadro, a janela considera também o primeiro sub-quadro do quadro subsequente. Como cada sub-quadro contém 60 amostras e cada uma destas leva 125  $\mu$ s, chegamos a um atraso de *look-ahead* de 7,5 ms.

Na codificação MP-MLQ, que possui a maior banda, são transmitidos 189 bits por quadro, enquanto na codificação ACELP são 158 bits por quadro. Dessa maneira, as taxas de codificação são respectivamente 6,3 kbits/s ( $189 \times 8000 / 240$ ) e 5,3 kbits/s ( $158 \times 8000 / 240$ ).

## 2.7) Sumário dos Codificadores

Nesta seção iremos buscar fazer um sumário dos codificadores apresentados nas seções e subseções anteriores. Conforme vimos, cada um deles possui suas características intrínsecas e seu comportamento em ambiente VoIP será estudado ao longo deste trabalho. Como referência, vamos construir aqui a tabela 2.3 que resume as principais características de cada um dos codificadores a ser estudado. Deve ser observado que a coluna MOS denota uma pontuação subjetiva atribuída aos codificadores de voz, que pode variar de 1,0 a 5,0. No Capítulo 4 entraremos em detalhes de como esta é atribuída aos codificadores de voz. Uma pontuação superior a 3,0 indica que o codificador é de boa qualidade.

**Tabela 2.3:** Taxa de codificação e pontuação MOS dos codificadores apresentados.

<b>Codificador</b>	<b>Taxa de codificação (kbits/s)</b>	<b>MOS</b>
CELP – UFRJ	8	2,8
G.711	64	4,3
G.726	32	4,2
G.728	16	4,0
G.729	8	3,9
G.723.1	5,3/6,3	3,8

## **2.8) Conclusão**

Neste capítulo analisamos primeiramente o problema da codificação de forma bastante genérica, destacando sua importância para a concepção dos sistemas de comunicação por voz.

Em seguida vimos que existem basicamente três técnicas de codificação, cada uma destas com as suas vantagens e desvantagens. Além disso, buscamos mostrar exemplos de codificação usuais para cada uma das técnicas mencionadas, com destaque para o sistema CELP desenvolvido na UFRJ.

Finalmente analisamos os codificadores do ITU, que serão alvo de estudo no decorrer deste trabalho, bem como sumarizamos todos os codificadores vistos no capítulo na seção seguinte.

No próximo capítulo nosso trabalho irá se concentrar no que realmente é um ambiente VoIP, no qual serão realizados os testes e análises deste trabalho.

# Capítulo 3 - VoIP

## 3.1) Introdução

Neste capítulo temos como objetivo expor as principais características de um ambiente VoIP. Procuraremos ressaltar quais os principais aspectos do mesmo, observar como se formam suas respectivas estruturas de *hardware*, verificar quais as arquiteturas atualmente utilizadas (efetuando uma comparação entre as mais usuais) e buscar quais aplicativos mais comuns atualmente lançam mão dessa tecnologia.

Na seção 3.2 faremos uma breve apresentação do que é um ambiente VoIP, destacando quais são as características e aspectos que devem ser levados em consideração quando utilizamos uma transmissão de dados através de um ambiente como este.

Na seção 3.3, iremos discutir os diferentes elementos da arquitetura VoIP, procurando mostrar a participação de cada um deles e como o funcionamento de cada um deles pode influenciar na transmissão de dados.

Nas seções 3.4, 3.5 e 3.6 iremos detalhar os protocolos VoIP mais usuais atualmente. Na seção 3.4 iremos detalhar o protocolo H.323, enquanto na seção 3.5 iremos especificar o protocolo SIP, tentando enfatizar nessas seções as distinções entre os protocolos. Já na seção 3.6 apresentaremos protocolos alternativos.

Na seção 3.7 mostraremos os aplicativos VoIP mais utilizados hoje em dia, como por exemplo Skype, MSN e Google-Talk, explicitando como são concebidos e quais são algumas de suas diferenças.

Finalmente, na seção 3.8 faremos uma conclusão, resumizando todos os aspectos discutidos ao longo do capítulo.

## 3.2) O Ambiente VoIP

Ao longo das últimas duas décadas um grande problema se estabeleceu. “Como utilizar os sistemas telefônicos para transmissão de dados?” Algum sucesso foi obtido, com a utilização de modems, porém sua banda de transmissão foi sempre um tanto restrita, o que motivou o desenvolvimento de outros tipos de canal de transmissão, o que acabou por gerar os sistemas atuais de internet de banda larga.

Com o passar do tempo (mais especificamente com o desenvolvimento dos codificadores de voz) passou-se a interpretar a voz como um pacote de dados, assim como qualquer outro transmitido na internet. Isso levou a uma mudança de modelo, ou melhor, levou a busca pela utilização do computador para comunicação por voz, utilizando a rede mundial, procurando dispensar todo o sistema telefônico disponível. A esses novos sistemas foi dado o nome de VoIP (do inglês, *Voice over IP*), ou simplesmente, voz sobre IP. Dessa maneira, abriu-se um largo campo de estudos e desenvolvimento.

Isso se deve ao fato de que uma vez que temos uma tecnologia estabelecida, robusta e eficiente, temos certa aversão a dispensá-la para passarmos a utilizar uma novidade. Isto nos obriga a nos certificar que os sistemas VoIP apresentem ganhos em relação aos sistemas de telefonia. Por isso, devemos avaliar quais os aspectos contribuem para a eficiência dos sistemas de telefonia, ter a certeza de que os mesmos são preservados nas arquiteturas VoIP a serem utilizadas e que os mesmos apresentem vantagens marcantes em relação aos sistemas tradicionais.

As possibilidades de uso de VoIP são diversas. Primeiramente, como este é um serviço baseado no protocolo IP, o VoIP pressupõe dois computadores conectados à internet. Já existem uma série de *softwares* gratuitos e de interface amigável (como MSN, Skype e Google-Talk) que são capazes de prover serviços VoIP, mas também outros serviços, como vídeo sobre IP. Além disso, temos a possibilidade de através de um

computador ligar para um terminal telefônico comum e vice-versa, com tarifas telefônicas bastante inferiores àquelas praticadas pelas operadoras de telefonia tradicional.

Um outro ponto marcante que deve ser citado é o fato de que diferentes tipos de clientes de serviços VoIP apresentam requisitos distintos. É perfeitamente concebível que um usuário residencial tenha necessidades de uso bastante distintas de uma empresa que deseja lançar mão de VoIP para uma eventual otimização de suas operações. Além disso, as operadoras que passam a fornecer serviços VoIP passam a ter que realizar uma análise com relação a seus clientes, análise esta que pode ser colocada com uma pergunta: “Até que ponto o usuário troca uma baixa qualidade de serviço por uma tarifa gratuita ou com altos descontos quando comparado com a telefonia tradicional?” Uma resposta para esta pergunta depende de uma série de fatores, sendo o principal o tipo de cliente, residencial ou comercial. Um cliente residencial tem uma tendência a aceitar serviços de qualidade inferior, buscando sempre a menor tarifa possível. Já um cliente comercial possui uma preocupação grande com relação a seus fornecedores, clientes, funcionários etc. Isso significa que o mesmo será um tanto mais cauteloso que o residencial quando da migração para uma tecnologia mais barata, mas com qualidade inferior. Estas questões, atualmente sem uma resposta precisa, colocam para os usuários, e principalmente para as operadoras, um desafio de adaptar-se ao uso do VoIP.

Agora, que já vimos os aspectos gerais do ambiente VoIP, bem como as suas possibilidades de uso e os interesses dos usuários e fornecedores desse tipo de serviço, apresentaremos como são estruturadas as arquiteturas VoIP.

### **3.3) Arquitetura VoIP**

Um sistema VoIP pode ser basicamente definido pela simples conversão de amostras de voz (ou uma manipulação matemática das mesmas) em uma série de pacotes de dados para transmissão por uma rede IP. Seus três elementos básicos são os seguintes: o *gateway* de voz, o agente de chamada e o equipamento do usuário.

O *gateway* de voz é um dispositivo responsável pela digitalização, compressão, demodulação e determinadas funções de empacotamento IP do sinal de voz recebido. É mesmo é normalmente tratado como o elemento de uma rede comum através do qual saem os pacotes.

Já o agente de chamada é um módulo que recebe uma sinalização do *gateway* indicando quando uma chamada está sendo iniciada. Em seguida, o agente de chamada analisa como a chamada será gerenciada.

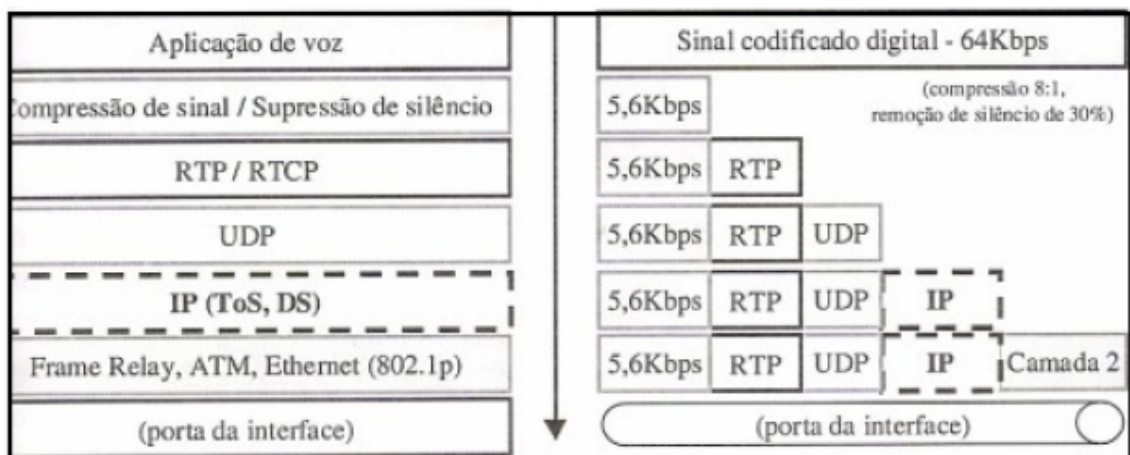
O equipamento do usuário é o dispositivo final utilizado pelo usuário, para solicitar um serviço de voz.

A figura 3.1, de [6], mostra a estrutura VoIP com os elementos acima descritos. Os telefones que estão ligados aos PBXs são os equipamentos do usuário (pontos finais). Já os *gateways* digitalizam a voz vinda do PBX, empacotam-na e a enviam para o seu destino, que é um outro *gateway*.



**Figura 3.1:** Pontos finais ligados a PBXs interligados através de rede IP [15].

As soluções VoIP possuem um modelo de camadas hierárquico. Esse modelo pode ser comparado ao modelo OSI, sendo que no processo de comunicação VoIP o efeito de cada camada é a adição de um cabeçalho de controle e transmissão da informação, o que pode ser observado na figura 3.2.



**Figura 3.2:** Adição dos cabeçalhos RTP, UDP e IP à amostra de voz codificada [15].

Primeiramente, após ser digitalizada, a voz é comprimida, através de um algoritmo de codificação. Em seguida a mesma deve ser empacotada utilizando o protocolo RTP (do inglês, *Real Time Protocol*). Este protocolo transporta os pacotes de voz codificada, incluindo outros pacotes multimídia interativos, porém sem nenhuma garantia com relação à perda de pacotes e à qualidade de serviço. Numa rede IP não existe nenhuma garantia de qual caminho um determinado pacote irá tomar. Isso significa que no destino os pacotes podem chegar totalmente desordenados. Nesse ponto, o protocolo RTP tem outra função fundamental: ordenar os pacotes de voz antes de a voz ser decodificada.

Juntamente com o RTP, utiliza-se o protocolo RTCP (do inglês, *Real Time Control Protocol*), que realiza o controle durante a chamada. Além disso, o RTCP tem como objetivo medir e informar aos pontos finais as condições de rede. Porém, ao passo que o RTCP detecta quaisquer condições de rede (sejam elas boas ou ruins), não toma qualquer providência.

Como existe uma grande necessidade de entrega veloz e há uma falta de tempo caso ocorra uma retransmissão de pacote, utiliza-se um canal UDP (do inglês, *User Datagram Protocol*) para transmissão de voz através de RTP. Para que o pacote possa ser transmitido através da rede é fundamental a transmissão do protocolo IP, o qual é



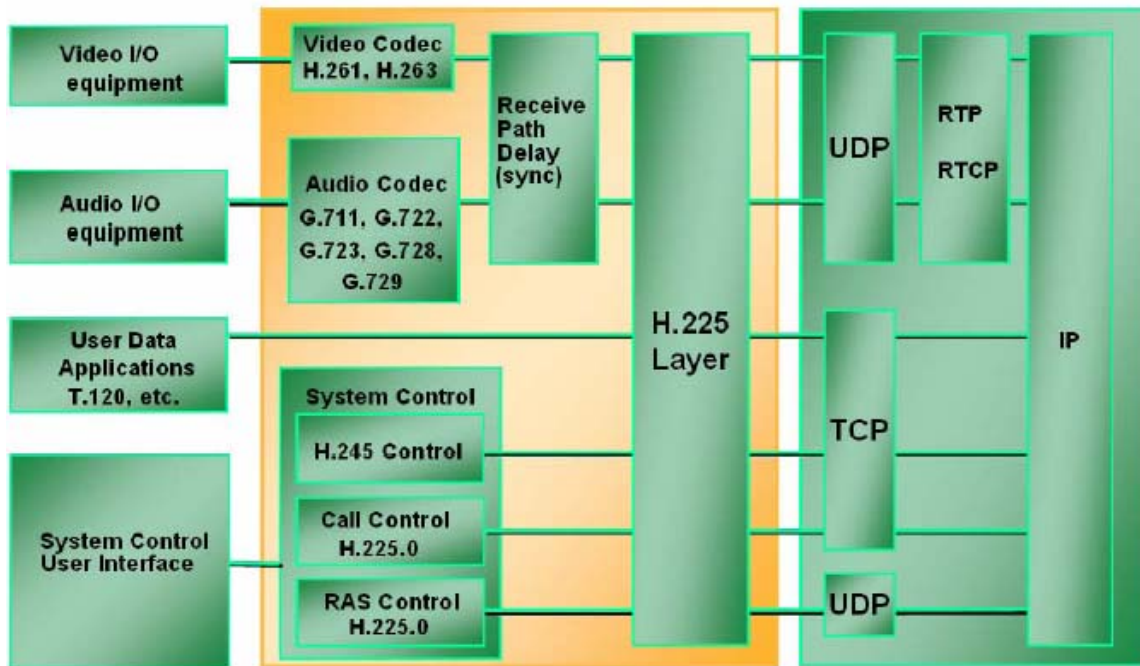
necessário para que os roteadores achem o endereço de destino. No final desse processo de adição de cabeçalhos (explicitado pela figura 3.2), o pacote de voz tem um *overhead* de 40 bytes, sendo 20 bytes IP, 8 UDP e 12 RTP [6].

Como sabemos cada um dos pontos finais pode ter especificações bastante diferentes dos outros, como por exemplo, requisitos de banda, áudio, dados etc. Logo, antes de iniciarmos qualquer conexão é necessário garantir que os dois pontos possuam as mesmas capacidades, bem como implementar uma arquitetura de controle de chamadas ou de sinalização que habilite o controle dos recursos da rede. Tal natureza distinta dos pontos finais torna a telefonia IP, no que se refere à tecnologia, muito mais versátil e complexa que a telefonia convencional.

Verificamos em linhas gerais, como se realiza a arquitetura de um sistema VoIP. Nas próximas seções iremos conhecer os principais protocolos existentes que representam realizações de arquiteturas largamente utilizadas nos aplicativos atuais.

### **3.4) Protocolo H.323**

O padrão H.323 é uma fonte de fundamentos para transmissão de áudio, vídeo e dados através de redes baseadas em pacotes, dentre as quais se inclui a internet. Uma grande vantagem deste padrão é o fato de os usuários poderem fazer uso do mesmo sem se preocuparem com compatibilidade, ou melhor, programas e aplicativos de diversos fornecedores podem trabalhar concomitantemente. O H.323 requer que todos os dispositivos que o utilizam dêem suporte ao tráfego de voz, além de opcionalmente poderem dar suporte a vídeo e dados. A interface é feita de seguinte forma: os dispositivos devem ser integrados a um PC através de alguma solução em *software* ou ainda devem ser implementados através de algum tipo de *hardware* do tipo *stand-alone*.



**Figura 3.3:** Protocolos formadores do H.323 [15].

Como pode ser visto através da figura 3.3, o H.323 pode ser caracterizado como um protocolo formado por uma série de outros protocolos. Na mesma figura pode ser observada a pilha de protocolos implementada pelos pontos finais (terminais e *gateways*) do H.323 numa rede VoIP.

Da mesma forma que os codificadores apresentados no capítulo anterior, o padrão H.323 foi formalizado pelo ITU (no ano de 1996). Seus elementos formadores também foram regulamentados pelo mesmo órgão. A seguir descrevemos cada um dos padrões que formam o H.323.

- O H.225 tem como função especificar mensagens de controle de chamada, dentre as quais se incluem a sinalização, os registros, as admissões, o empacotamento e a sincronização de *streams*.

- O H.245 escolhe o codificador a ser utilizado na chamada, especifica as mensagens para abertura e fechamento dos canais de transmissão de *streams* e outros comandos.
- Já as recomendações H.261 e H.263 se referem à escolha de codificadores de vídeo, sendo o primeiro um codificador audiovisual a 64 kbits/s e o segundo um novo tipo de codificador para vídeos PTOS (*Plain Old Telephony Service*).

Além destas recomendações, na composição da arquitetura padrão do H.323 existe uma série de componentes que serão especificados a seguir.

- Terminal – trata-se de um ponto final numa rede local que promove comunicação em tempo real com qualquer outro terminal H.323, *gateway* ou MCU (esses elementos serão descritos a seguir). Tal comunicação é composta por sinais de controle, indicações, áudio, vídeo e dados entres dois terminais. Um terminal obrigatoriamente provê voz, mas adicionalmente também pode prover dados e/ou vídeo.
- *Gateway* (GW) – é um ponto final numa rede local que provê comunicação em tempo real entre quaisquer terminais H.323 numa LAN (*Local Area Network*), outros terminais ITU numa WAN (*Wide Area Network*), ou ainda outro *gateway* H.323. Entre os outros terminais ITU podemos destacar o H.310, o H.320 (ISDN), o H.321(ATM), o H.322(GQOS-LAN), o H.324 (GSTN), o H.324M (móvel), dentre outros.
- *Gatekeeper* – é o componente que provê o controle de serviços de chamadas para os pontos H.323. Pode ser que exista mais de um *gatekeeper*, sendo que caso isso aconteça, eles podem se comunicar entre si. Além disso, cada um destes elementos é separado logicamente dos

pontos finais, porém sua implementação pode coexistir fisicamente com outro dispositivo H.323, ou mesmo não-H.323.

- Unidade de Controle Multiponto (*Multipoint Control Unit* - MCU) – trata-se de um ponto final numa LAN que permite a participação numa conferência multiponto de três ou mais terminais ou *gateways*. Uma outra função é permitir a conexão entre dois terminais para uma conferência ponto a ponto, possibilitando que posteriormente possam ser adicionados novos usuários, transformando a conferência em multiponto.
- Controlador Multiponto (*Multipoint Controller* - MC) – este é um elemento H.323 em uma rede local que providencia o controle para que três ou mais terminais possam participar de uma conferência multiponto. Como a MCU, possibilita que após tenha sido estabelecida uma conferência ponto a ponto, que a mesma se torne multiponto.
- Processador Multiponto (*Multipoint Processor* - MP) – é um outro elemento H.323 que provê processamento de *streams* de áudio, vídeo e dados em conferências multiponto, em redes locais. Os processamentos realizados pelo MP estão sempre sob o controle de um MC.
- Conferência ponto a ponto (*Point to point Conference*) – uma conferência deste tipo é realizada entre dois terminais ou entre um terminal H.323 e um SCN através de um *gateway*.
- *Switched-circuit Network* (SCN) – é uma rede de telecomunicações chaveada pública ou privada, como GSTN, N-ISDN ou B-ISDN.

O protocolo H.323 possui pontos finais razoavelmente inteligentes, pois são capazes de manter suas próprias chamadas. Ele é um sistema *peer-to-peer* de sinalização. Isso significa que cada ponto pode chamar um outro lançando mão dos procedimentos-

padrão, desde que um saiba o endereço IP do outro. As mensagens de sinalização inicial são as mesmas do modelo tradicional ISDN Q.931 [7], utilizando o formato ASN.1[8] encapsulado em pacotes TCP (do inglês, *Transmission Control Protocol*). Após a configuração inicial, os dois pontos são capazes de negociar qual codificador de voz deve ser utilizado e finalmente quais portas serão usadas para transmissão de pacotes RTP, ou melhor, a transmissão propriamente dita. Deve-se notar aqui que há uma dificuldade de operarmos com a presença de *firewalls*, uma vez que as portas RTP são escolhidas dinamicamente numa faixa muito grande de possibilidades.

Dessa maneira, apresentamos o H.323 e os seus componentes, descrevendo como atuam cada um dos seus protocolos e elementos. Na próxima seção iremos falar sobre o grande “concorrente” do H.323, o SIP.

### **3.5) Protocolo SIP**

O SIP (do inglês, *Session Initiation Protocol*) é um protocolo de sinalização mais recente que o H.323 e que utiliza uma arquitetura *peer-to-peer* bastante semelhante à do protocolo anterior. Porém, ao contrário do H.323, o SIP é um protocolo projetado diretamente para ser usado na internet.

O SIP depende de pontos finais com certa inteligência, o que traz o benefício de requerer quase nenhuma interface com os servidores, uma vez que cada ponto final gerencia sua sinalização, havendo ainda um controle de sessão. Tais características trazem grandes vantagens ao SIP. Outra característica marcante deste protocolo é que o mesmo possui uma estrutura relativamente simples de mensagens, permitindo uma configuração de chamada menos duradoura que a do H.323, o que, com a utilização do mesmo *hardware*, acarreta melhor desempenho.

Por ter um modelo de chamada bem mais distribuído que o H.323, o SIP é bem mais escalável que o H.323. Porém, sua grande vantagem está no fato de que ele é um protocolo concebido para internet. O SIP usa mensagens no formato ASCII [9] baseado

em HTTP/1.1 [10], o que resulta em mensagens mais facilmente decodificadas e depuradas. Uma consequência direta disto é que outras aplicações para a *web* podem suportar serviços SIP com poucas alterações. Mais especificamente, como este protocolo suporta nomes de URL, por exemplo, o seu modelo permite que o e-mail de um usuário seja o seu número de telefone, o que tem como consequência uma grande facilidade no que se refere à comunicação de pontos finais distintos.

Embora o SIP seja baseado na filosofia *peer-to-peer*, é regido por uma lógica cliente-servidor. Um exemplo de um típico cliente SIP seria um Telefone IP, um PC ou ainda um PDA (*Personal Digital Assistant*). Para a comunicação se fazem necessários dois elementos, um UAC (*User Agent Client*) para originar a requisição e um UAS (*User Agent Server*) para recebê-la. Opcionalmente podem ser suportados alguns outros servidores que, normalmente, são úteis e valiosos em aplicações SIP recentes. Estes são o *SIP Proxy Server*, o *SIP Redirect Server*, o *Registrar* e o *Location Server*.

- O *Proxy Server* é uma entidade que atua tanto como cliente quanto como servidor, além de inicializar requisições SIP de acordo com o interesse do UAC.
- Já o *Redirect Server* atua recebendo requisições SIP e mapeando o destino para um ou mais endereços e respondendo para estes endereços.
- O *Registrar* atua aceitando requisições de registro da localização atual dos UAC, sendo esta entidade tipicamente colocada junto a um *Redirect Server*.
- Finalmente, temos o *Location Server*, que provê as informações sobre as possíveis localidades das chamadas, sendo normalmente contatado por um *Redirect Server*, podendo também coexistir com ele.

O protocolo SIP, como já foi dito, é bastante simples no que se refere ao seu vocabulário de requisição/resposta. Seis palavras-chave controlam a comunicação:

- REGISTER (registra localização atual de um servidor);
- INVITE (mensagem enviada para inicialização de chamada);
- ACK (enviada por quem inicia a chamada para confirmar sua aceitação);
- BYE (término de chamada);
- CANCEL (término de chamada não conectada);
- OPTIONS (requisição de capacidades).

O endereçamento SIP, como colocado acima, é modelado como o de URLs. A estrutura do endereço indica os parâmetros, o tipo de transporte e os endereços. Um exemplo pode ser visto a seguir:

**sip: "lamas" [rodlamas@email.com](mailto:rodlamas@email.com); transport=udp**

Dessa maneira, verificamos as principais características de cada um dos comandos e elementos da arquitetura SIP. Procuramos enfatizar nesta seção em quais pontos ela apresenta distinção em relação ao H.323. Verificamos como principal característica que o SIP apresenta uma escalabilidade superior ao H.323. Porém, devido a uma segurança maior com os pacotes de dados transmitidos, o H.323 vem sendo mais utilizado.

Contudo, ainda existe uma grande discussão na comunidade VoIP sobre qual dos dois protocolos será mais utilizado. Mas temos a certeza de que até o momento esses dois

apresentados são os líderes de mercado. Na próxima sessão faremos menção a outros protocolos menos utilizados atualmente.

### **3.6) Outras Arquiteturas VoIP**

Dentre as arquiteturas VoIP menos utilizadas do que as duas anteriormente citadas, podemos citar outras quatro que até certo ponto têm suas participações no mercado. São elas:

- O MGCP (*Media Gateway Controller Protocol*) da IETF[11];
- O MEGACO (*Media Gateway Controller*), que é o sucessor do MGCP, também da IETF[11];
- O H.248 que é um “sinônimo” do MEGACO, mas homologado pela ITU, além de usar o padrão ASN.1[8] de codificação, enquanto o anterior usa o padrão ASCII;
- O Skinny que é um protocolo comercial desenvolvido pela Cisco.

### **3.7) Aplicativos VoIP**

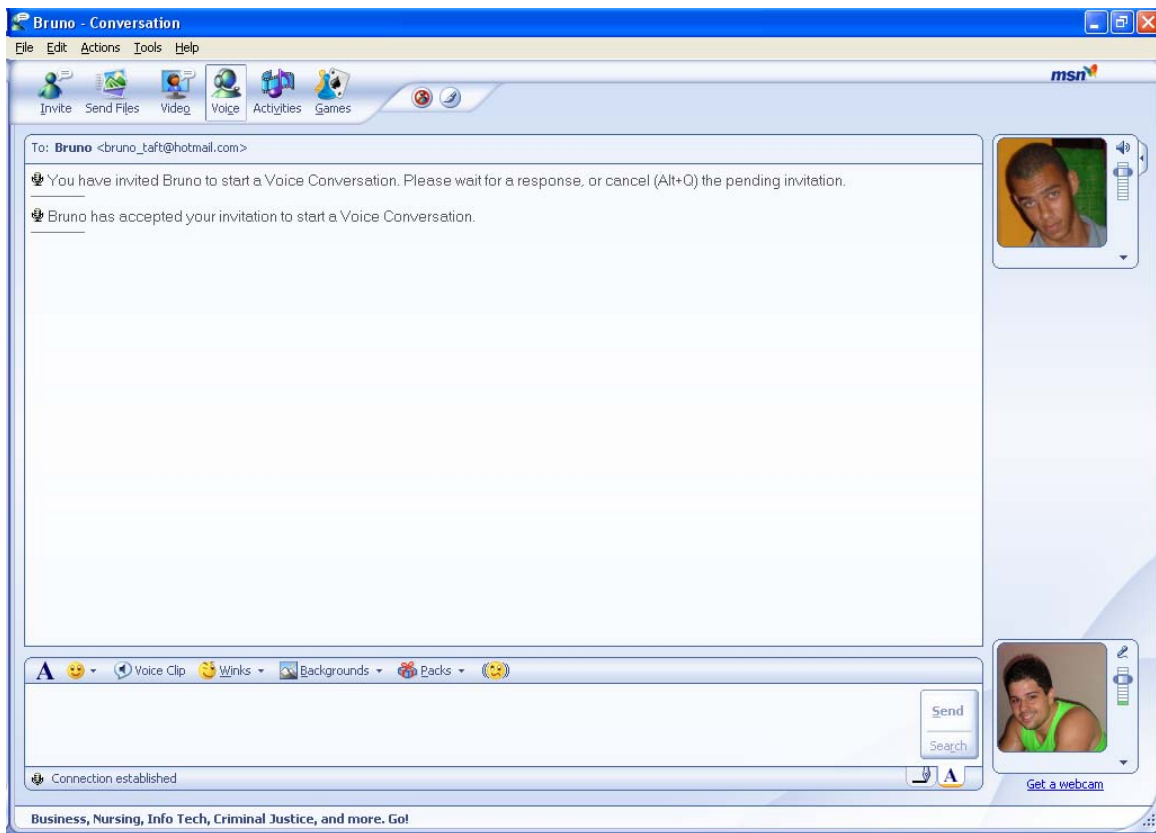
Hoje em dia, a maior parte das pessoas está em contato com aplicativos VoIP sem mesmo saber como os mesmos operam. Eles vêm se tornando cada vez mais comuns em nosso dia-a-dia. Como grandes exemplos, podemos citar 3 ferramentas, as quais iremos mostrar a seguir.

Um primeiro exemplo é o “MSN Messenger” que foi desenvolvido (e vem sendo atualizado regularmente) pela Microsoft desde o final do século passado. Inicialmente ele surgiu como um aplicativo de troca de mensagens em tempo real, porém, com o passar do



tempo e o surgimento de aplicativos VoIP de outras empresas, o mesmo passou a incorporar tal função. Sua aceitação na comunidade da Internet é extremamente alta, uma vez que o mesmo é totalmente amigável com o usuário e vem pré-instalado no Windows, sistema operacional mais difundido no mundo, atualmente. O MSN utiliza como protocolo de arquitetura o SIP. Vale aqui lembrar que o MSN é o único aplicativo dentre os mais comuns que suporta transmissão de vídeo sobre IP.

A figura 3.4 mostra um exemplo de uma conexão VoIP estabelecida entre dois usuários do MSN Messenger. Na mesma pode ser verificado que existe uma série de outros serviços que podem ser estabelecidos entre seus usuários, dentre os quais se destaca o serviço de vídeo sobre IP, desde que cada um dos usuários tenha uma *webcam*.

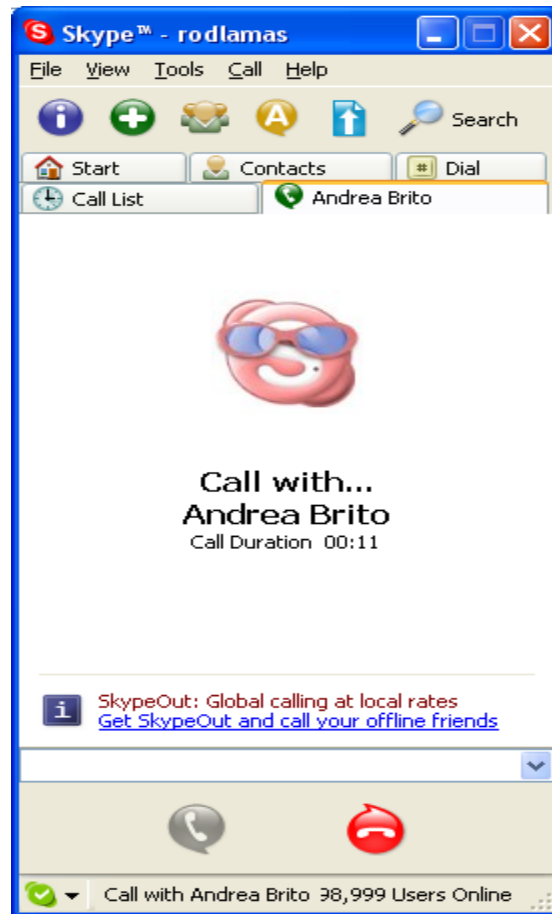


**Figura 3.4:** Conexão VoIP estabelecida via MSN Messenger

Um segundo aplicativo que pode ser citado é o “Skype”, que nos últimos meses vem crescendo de maneira até um certo ponto espantosa. Mais de 170 milhões de

*downloads* já foram feitos através de sua página principal o que revela uma grande aceitação na comunidade. Seu grande objetivo é fazer com que cobranças sobre tarifas telefônicas sejam extintas a médio-longo prazo, sendo o usuário tarifado apenas pelo serviço de internet banda larga que usa, sendo que o preço deste é normalmente fixo e não apresenta limites com relação ao tráfego de dados (pacotes). Com relação às suas funcionalidades, o mesmo apresenta certa vantagem relação à concorrência, uma vez que permite que o usuário, por exemplo, utilize serviços como *voicemail* e até mesmo faça ligações para terminais telefônicos, sendo este serviço tarifado, embora com tarifas cerca de 80 a 90% inferiores às praticadas pelas empresas de telecomunicações. Este serviço é denominado Skype-out. Com relação à arquitetura VoIP, utiliza um protocolo próprio.

A figura 3.5, mostrada a seguir, mostra uma conexão VoIP estabelecida entre dois usuários do Skype. Pode ser visto na figura também uma “aba” onde há a opção de discagem (em inglês, *dial*), através da qual pode ser feita a conexão com um terminal telefônico fixo por tarifas bem atraentes.

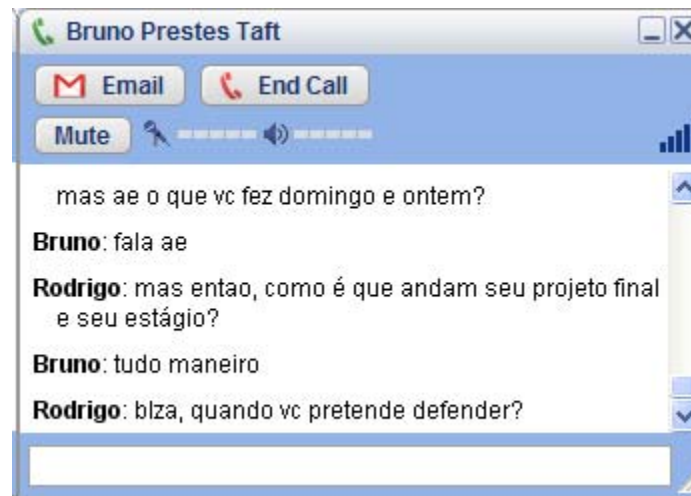


**Figura 3.5:** Conexão VoIP estabelecida via Skype

Um terceiro (e mais recente aplicativo) é o “Google-Talk”, cujo lançamento foi realizado no final de agosto de 2005. Segue os mesmos moldes do Skype, porém ainda possui funcionalidades limitadas, restringindo-se à transmissão VoIP e ao serviço de mensagens em tempo real. Utiliza protocolo e arquitetura próprios, mas em versões futuras estará baseado no SIP. Espera-se bastante desta ferramenta, tanto em tecnologia e funcionalidades quanto em relação ao sucesso. Isto é devido à enorme aceitação que existe do serviço de buscas do “Google” e o serviço de e-mails do “Gmail” que apresenta uma forma única de apresentação dos e-mails, muito mais amigável ao usuário do que a maior parte dos outros.

A seguir, na figura 3.6, podemos verificar uma conexão VoIP estabelecida entre dois usuários do Google-Talk. Na mesma observamos uma integração entre os serviços

“Google”, uma vez que os usuários têm acesso aos seus respectivos e-mails através de um botão na janela de conversação estabelecida.



**Figura 3.6:** Conexão VoIP estabelecida via Google-Talk

Podemos ainda destacar um quarto e último aplicativo, que é uma solução nacional, cujo nicho alvo de mercado são pessoas que desejam fazer ligações de longa distância, cujas tarifas praticadas pelas operadoras de telefonia são bastante altas. Trata-se do “UOL Fone” que possibilita, assim como o Skype, a comunicação VoIP gratuita e ligações de longa distância com tarifas pelo menos 80% mais baratas que as tradicionais.

### 3.8) Conclusão

Neste capítulo procuramos enfatizar qual a importância do estudo de VoIP atualmente. Claramente pudemos verificar que cada vez mais existem novas arquiteturas surgindo e novos aplicativos sendo desenvolvidos, o que no final só acaba por trazer benefícios para os usuários comuns.

Primeiramente vimos o que é um ambiente VoIP e quais são os tipos de serviço que existem atualmente, bem como quais são as perspectivas que deverão ser consideradas pelas operadoras e pelos usuários comerciais e residenciais.

Em seguida, mostramos as principais características e elementos que compõem uma arquitetura VoIP, além de que dedicamos duas seções para a explicação detalhada das duas arquiteturas mais utilizadas atualmente no mercado. Posteriormente, mencionamos outros exemplos de arquiteturas menos comuns.

Finalmente, demos exemplos de aplicativos comuns e extremamente difundidos na comunidade da Internet e observamos que seu crescimento vem aumentando a cada momento e uma expansão deste mercado é iminente.

No próximo capítulo abordaremos a questão da qualidade de serviço (QoS) e tentaremos examinar qual a importância de termos QoS em arquiteturas/redes VoIP.

# Capítulo 4 - QoS – Qualidade de Serviço

## 4.1) Introdução

Quando nos referimos à qualidade de um serviço, estamos basicamente interessados em saber como o mesmo se comporta em relação a parâmetros determinados. No caso, dos serviços VoIP, tais parâmetros (como esperado) são bastante específicos e o comportamento de um sistema perante os mesmos permite dizer se ele é eficiente ou não.

Nossos objetivos nesse capítulo são expor o que é de fato QoS (do inglês, *Quality of Service*), indicar quais são seus parâmetros relevantes no ambiente VoIP, indicar como aferi-la (medi-la) e colocar como iremos proceder com os testes de QoS a serem apresentados no capítulo seguinte.

Na seção 4.2, iremos expor o que é qualidade de serviço em um ambiente VoIP, procurando ressaltar a sua importância quando da escolha de um sistema.

Na seção 4.3, iremos explicitar os aspectos comerciais da QoS e procuraremos colocar como as operadoras de telecomunicações terão de utilizá-la de modo a fazer os potenciais usuários utilizarem os serviços VoIP.

Na seção 4.4, serão expostos quais os parâmetros que devem ser considerados para a análise da qualidade de um sistema (codificador) num ambiente VoIP. Mostraremos quais parâmetros são mais e quais são menos importantes, procurando especificar quais as razões para a maior ou menor relevância de cada um deles.

Na seção 4.5, procuraremos explicitar quais formas existem para avaliar e medir a qualidade dos sistemas num ambiente VoIP. Apresentaremos uma série de avaliadores,

com suas características, destacando suas respectivas importâncias. Ao final escolheremos um deles para que seja a referência para nossas avaliações (testes).

Finalmente, na seção 4.6, iremos concluir este capítulo procurando sumarizar os itens abordados ao longo do mesmo.

## **4.2) QoS**

Como estamos verificando ao longo deste projeto, a convergência de dados, voz e vídeo em um único meio (muitas vezes comandada por uma única aplicação) é uma realidade cada vez mais presente nos sistemas VoIP. Dessa maneira, as redes, bem como a própria internet, precisam se adequar para acomodar as demandas dessas novas aplicações, especialmente aquelas como voz e vídeo, que possuem requerimentos maiores com relação à QoS.

Com relação à voz, quando esta é dividida, empacotada e submetida ao mesmo meio físico que um volume imprevisível de dados, a QoS é chamada para proteger o tráfego de voz. A especificação correta de QoS que implica a requisição de um serviço é uma exigência da operação da rede, de seus componentes e dos equipamentos utilizados para viabilizar a operação com máxima qualidade possível para uma determinada aplicação baseada em parâmetros definidos, os quais serão detalhados na próxima seção.

Porém, o protocolo IP, pelo qual é efetuado o transporte dos dados de voz, não possui nenhum compromisso com qualidade de serviço. Este é considerado um protocolo que provê um serviço de melhor esforço, no qual os recursos disponibilizados pela rede são distribuídos de forma igualitária, não havendo distinção alguma entre os pacotes. Isso significa que há uma probabilidade de o serviço ser de péssima qualidade, o que não seria interessante. No entanto, quando a QoS se faz presente, passamos a considerar uma mudança significativa, uma vez que passamos a tratar diferentes tipos de dados de formas distintas. Contudo, esta alteração modifica o princípio fundamental da simplicidade que trouxe o sucesso à internet. Porém, com o grande crescimento da quantidade dos mais

variados serviços sobre IP (*everything over IP*), novas oportunidades vêm sendo geradas e apoiando o suporte à QoS, o que é imprescindível em aplicações multimídia em tempo real, como aquelas que envolvem áudio, voz e vídeo.

Atualmente a definição de QoS é dada pela capacidade da rede de prover serviço de encaminhamento de dados de forma consistente, coerente e previsível. Também existe a possibilidade de definirmos QoS como a habilidade de um elemento da rede de ter algum nível de garantia que seu tráfego e exigências de serviço possam ser satisfeitos. A provisão de QoS em redes IP é normalmente feita através de dois mecanismos (arquiteturas) básicos. Vale a pena mencionar aqui, antes mesmo de detalharmos cada um deles, que não são mutuamente exclusivos, mas sim complementares.

A Reserva de Recursos é o primeiro destes mecanismos. Baseia-se no princípio de que os recursos da rede são divididos de acordo com os requisitos de QoS de cada uma das aplicações em uso. O RSVP (*Resource ReSerVation Protocol*), por exemplo, provê os mecanismos para implementação de serviços integrados (*IntServ*) baseado na reserva de recursos. Uma vez que um determinado recurso esteja alocado para uma determinada aplicação, ninguém pode utilizá-lo, mesmo se a aplicação em questão não estiver fazendo uso do mesmo.

Uma segunda arquitetura de provisão de QoS é a Priorização. Esta se baseia em o tráfego de rede ser classificado e os seus recursos divididos entre as diversas classes de tráfego de acordo com critérios de administração da rede. Para que a QoS seja habilitada, os mecanismos de classificação dão preferência às aplicações identificadas com requisitos maiores.

Sendo assim, colocamos o que de fato é QoS. Procuramos além de fazer uma breve definição, colocar a sua importância e os mecanismos através dos quais ela é provida.



### 4.3) A Importância Comercial da QoS

Sabemos que a qualquer tipo de serviço ou produto oferecido à sociedade deve apresentar uma relação custo - benefício atraente para seu público alvo. Caso contrário, o serviço estará fadado ao insucesso. Dessa maneira, assim como qualquer outro serviço, os sistemas VoIP, devem ser capazes de fornecer aos seus eventuais clientes um sistema de tráfego de voz que apresente uma boa relação entre custo e qualidade do mesmo. Nesse ponto, a QoS passa a ser extremamente importante, uma vez que (como já diz o seu próprio nome) ela passa a controlar a qualidade do serviço que está sendo fornecido.

Como vimos no capítulo anterior, cabe aos clientes comerciais e residenciais estabelecer um limite de qualidade do tráfego de dados de voz e o valor que irá ser cobrado pelo mesmo. As empresas fornecedoras de transmissão VoIP irão, com isso, saber de que forma poderão aplicar a QoS para que o(s) serviço(s) que elas ofereçam sejam adequados às requisições dos seus mais diversos usuários.

Atualmente, a utilização comercial de serviços VoIP é bastante restrita, não sendo ainda explorada comercialmente pelas grandes operadoras. Como estas irão enfrentar o problema, ainda é uma incógnita para todos. Isto acontece porque nos sistemas VoIP, como já foi visto, os pacotes de voz são tratados como pacotes de dados quaisquer, o que implica numa grande redução de custos tanto para operadora como para o consumidor final. O resultado direto que podemos esperar disto é que, caso os sistemas VoIP se revelem um sucesso entre os seus potenciais consumidores, as empresas terão de readequar suas políticas de *marketing*, vendas, tecnologia, estratégia e outras mais. O que irá acontecer a partir daí, ainda é incerto, mas podemos ter uma certeza: o que quer que aconteça, será em prol da sociedade e de seus membros.

Tendo visto como a QoS irá influenciar comercialmente a sociedade, iremos na próxima seção dissertar sobre os parâmetros essenciais para sistemas de tráfego de voz com QoS.

## 4.4) Parâmetros de Redes com QoS

Os parâmetros considerados essenciais para avaliar a QoS em sistemas VoIP são: o atraso fim-a-fim, a variação do atraso (do inglês, *jitter*), taxa de perda de pacotes e a largura de banda de transmissão. A seguir descreveremos tais parâmetros, procurando ressaltar suas peculiaridades.

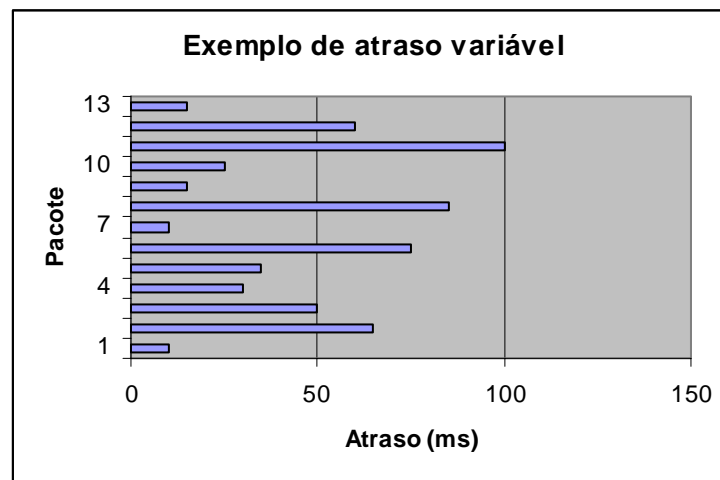
- Atraso fim-a-fim – Este é definido como o tempo entre o envio de uma mensagem por um ponto-final e a recepção da mesma pelo ponto-final de destino. Este atraso ocorre no caminho de transmissão ou ainda em dispositivos que estejam nesse caminho. Para transmissão de voz, a ITU recomenda (através da norma G.114, levando em consideração parâmetros do sistema auditivo humano) que tais atrasos não sejam superiores a 150 ms em uma direção. Atrasos entre 150 e 400 ms ainda são aceitáveis, porém acima desse último valor são inaceitáveis. A tabela 4.1 mostra a classificação do atraso fim-a-fim. Quando o atraso se torna demasiadamente alto, os usuários passam a ter a sensação de que a conexão entre os dois foi perdida e isso acaba por deixá-los bastante irritados. Dessa maneira, o atraso se torna um parâmetro bastante importante, uma vez que caso o usuário não esteja satisfeito com um serviço, ele dificilmente irá continuar usando o mesmo.

**Tabela 4.1:** Atraso fim-a-fim segundo a recomendação G.114 do ITU.

Atraso fim-a-fim	Classificação
Inferior a 150 ms	Desejável
Entre 150 e 400 ms	Aceitável
Superior a 400 ms	Inaceitável

- Variação do Atraso (*Jitter*) – Este parâmetro apresenta a seguinte definição [6]: “O *jitter* é o valor médio de todas as diferenças de atrasos incrementais, do início da medição até o pacote atual. Os pacotes mais recentes têm um peso maior do que aqueles menos recentes. A curva de *jitter* precisa de pelo menos 100 pacotes para ser estabilizada.” Uma situação onde temos o *jitter*, por exemplo, é quando

fluxos de voz e/ou vídeo são transmitidos em uma rede e os pacotes correspondentes chegam ao seu destino de maneira distorcida, ou melhor, fora de uma ordem sucessiva, fora de uma cadência, ou ainda com tempos de atraso variantes. Tal distorção é bastante prejudicial ao tráfego de pacotes multimídia, prejudicando o sinal de áudio/vídeo/voz na recepção. Para transmissão de voz a recomendação G.114 do ITU sugere que tenhamos um *jitter* médio de no máximo 30 ms.



**Figura 4.1:** Exemplo de pacotes com variação de atraso

Na figura 4.1 encontramos um exemplo de uma sucessão de pacotes transmitidos cujos atrasos variam de forma excessiva. Esta é uma situação altamente indesejada, podendo resultar no embaralhamento de informações, por conseguinte dificultando a compreensão do usuário de certa informação. Dessa maneira, como no caso do atraso fim-a-fim, o usuário ficará incomodado com esta situação e uma vez que ela aconteça constantemente, não usará um serviço que apresente tal característica. Este fato demonstra que o *jitter* possui uma grande importância na avaliação de sistemas VoIP.

Porém, em sistemas com níveis elevados de *jitter* pode ser inserido um *buffer*, a fim de que as informações sejam sincronizadas. Contudo, este *buffer* possui uma capacidade de entrada de dados e de armazenamento muito superior à de saída de dados (como por exemplo, uma caixa d'água), o que implica em imediato aumento do atraso fim-

a-fim. Neste momento, quando nos deparamos com uma situação como esta, deve ser adotada a solução mais conveniente, entre a presença de um *buffer* maior (que implicará um maior atraso) ou um de tamanho menor (que implicará um *jitter* maior).

- Perda de Pacotes – Já este parâmetro representa o numero de pacotes que foram transmitidos na rede, porém não alcançaram seu destino em um tempo pré-determinado. Isto ocorre porque o protocolo IP é um protocolo não totalmente confiável, que tem como prioridade a velocidade de transmissão e não a confiabilidade e a robustez. Dessa maneira, este é um parâmetro que deve ser considerado, uma vez que caso muitos pacotes sejam perdidos, a compreensão da informação de voz por parte do usuário ficará prejudicada e este fato acarretará as mesmas conseqüências apresentadas no caso do atraso fim-a-fim e do *jitter*. Nas transmissões de voz, a ITU recomenda que não haja perda superior a 1% dos pacotes.
- Largura de Banda – Trata-se de uma medida de capacidade transmissão de dados, expressa ou em *Quilobits* por segundo (kbits/s) ou em *megabits* por segundo (Mbits/s). Quando do projeto de uma rede, é necessário que seja definido o consumo de banda por cada uma das aplicações da rede, de modo que cada uma destas seja capaz de realizar suas funções, sem que prejudique o funcionamento das outras.

Destes quatro parâmetros, quando da realização dos nossos testes, iremos considerar os três primeiros, não levando em consideração a largura de banda. A justificativa para isso se encontra no fato de que para o nosso propósito, a análise de codificadores num ambiente VoIP, a taxa de codificação máxima é de 64 kbits/s e os serviços de internet de banda larga, sejam eles a cabo ou ainda via ADSL, possuem taxas de *download* e *upload* bastante superiores a esta taxa máxima, o que torna este parâmetro, para os testes considerados neste projeto, de menor importância em relação aos demais.

Existem ainda outros parâmetros que poderiam ser considerados, como eco, ruído de ambiente, outros tipos de ruído, distorções etc. Porém, como estes aspectos são dependentes da aplicação em questão e/ou da rede que será utilizada, os mesmos não serão abordados neste trabalho e ficarão como sugestões para desenvolvimentos futuros.

Na próxima seção iremos expor como avaliamos a QoS de sinais de voz e quais são as ferramentas disponíveis atualmente para que façamos tais medidas.

## **4.5) Avaliação de QoS em Sinais de Voz**

Quando nos referimos aos sinais de voz, seja quando estamos falando de telefonia convencional ou ainda de voz digitalizada através de uma rede IP, a qualidade do sinal é o parâmetro que é levado em consideração. Quanto mais próximo for o sinal de saída do sinal de entrada, melhor a qualidade do serviço em questão. Uma outra necessidade para o uso destes métodos é a possibilidade de comparação da eficiência das respostas de diversos codificadores.

Existem diversos métodos para este propósito, sendo alguns deles recomendações da ITU. O mais disseminado destes é o MOS (*Mean Opinion Score*), empregado em muitos trabalhos acadêmicos, artigos e publicações sobre avaliação de codificação de voz. Porém, este é um procedimento subjetivo, que exige um grande número de ouvintes, e conseqüentemente é extremamente custoso.

Nas subseções a seguir apresentaremos os diferentes avaliadores existentes e ao final selecionaremos aquele(s) que considerarmos mais eficiente(s) para procedermos como as nossas avaliações.

### **4.5.1) P.800 (MOS)**

Esta recomendação descreve métodos e procedimentos que possibilitam uma avaliação subjetiva da qualidade de transmissão de um sinal de voz. Esta é conhecida como MOS, sendo apenas uma das formas de pontuação citadas nesta recomendação [12], que foi aprovada em 30 de agosto de 1996.

A P.800 possui sugestões para a condução de testes subjetivos de qualidade de transmissão em laboratório. Provê métodos considerados eficientes para a determinação do quão satisfatório é o desempenho de uma dada conexão telefônica. Como estes são testes subjetivos, deve ser notado que apenas procedimentos de conversação ou de simples escuta devem ser utilizados como métodos para teste, com o fim de medir a qualidade de transmissão dos sistemas em questão.

De maneira bem simplificada, podemos dizer que o procedimento para os testes ocorre da seguinte forma. Vários usuários são submetidos a diversos trechos de escuta/conversação e em seguida devem atribuir uma nota que varia entre 1 e 5, conforme a tabela 4.2 a seguir.

**Tabela 4.2:** Pontuação MOS

Qualidade da Fala	Pontos
Excelente	5
Bom	4
Regular	3
Pobre	2
Ruim	1

Observamos que este método é de bastante simples execução, uma vez que qualquer usuário leigo pode ser submetido a ele. A sua grande desvantagem é o fato de requerer, para um alto grau de precisão, que seja utilizado um grande número de usuários.

Dessa maneira, podemos concluir que este método é bastante útil para a identificação da qualidade da fala, tanto em escuta como em conversação, uma vez que,

por realizar testes subjetivos, é capaz de identificar a interferência causada pelo atraso, pelo *jitter* e pela perda de pacotes. Porém, como colocamos acima, é um método muito custoso, por demandar muitos usuários para os testes, além de uma série de exigências quanto ao ambiente de realização dos mesmos [12]. Além disso, este é um método subjetivo, o que até certo ponto impede a sua utilização direta em trabalhos mais rigorosos. Os dados com relação às notas atribuídas aos diferentes codificadores existentes foram apresentados no Capítulo 2.

#### **4.5.2) P.861 (PSQM)**

Esta recomendação descreve, um método para estimar de forma objetiva a qualidade de codificadores de voz, operando na banda telefônica. O cálculo da qualidade é determinado através de uma medida denominada PSQM (do inglês, *Perceptual Speech Quality Measure*), que foi escolhida pela ITU entre uma série de outras por apresentar o melhor desempenho entre estas com relação à qualidade subjetiva de codificadores de voz[13]. Esta recomendação é posterior à P.800, tendo sido publicada em fevereiro de 1998.

A P.861 é uma recomendação que mede a qualidade de codificadores de voz apenas baseada em testes de escuta, ao contrário da P.800, que leva em consideração testes de escuta e de conversação. Ela apresenta uma acurácia aceitável quando avalia a resposta de codificadores forma de onda ou híbridos com taxas superiores a 4 kbits/s; o nível de fala na entrada dos codificadores; e a dependência das características do sinal original. Porém, não existe informação disponível suficiente para que esta recomendação seja acurada quando são avaliados os seguintes aspectos: erros no canal de transmissão; ruído ambiente na geração do sinal; música como sinal de entrada; e ainda corte (do inglês, *clipping*) temporal ou de amplitude na fala. Além disso, sabe-se que este método não possui acurácia em experimentos com variação no atraso entre os sinais de entrada e saída do sistema testado. Ainda, este método assume, para todos os efeitos, que a fala é tomada como fonte limpa (conforme já citado acima) e que não existe degradação no canal devido à transmissão equivocada de bits e/ou perda de pacotes. Estes fatos

comprometem largamente a utilização do mesmo em nossos testes, uma vez que estes problemas são extremamente comuns nas redes IP que servem como infra-estrutura para o transporte de voz. Dessa maneira, apesar de esta recomendação possuir uma grande relevância por representar o primeiro método objetivo de medição de qualidade de voz (e com larga aceitação/utilização), o mesmo não será adotado neste trabalho.

### **4.5.3) PSQM+**

Este é um método sobre o qual não há grandes referências na literatura. O mesmo é tido como uma “evolução” do PSQM (descrito na subseção anterior) que tenta corrigir certos erros com relação à taxa de erro de bit e perdas de pacotes. Mas em linhas gerais, o PSQM+ mantém a mesma filosofia do seu antecessor.

As melhorias do PSQM+ em relação ao PSQM nos levariam a crer que o mesmo é capaz de suprir as nossas necessidades, uma vez que o mesmo considerada aspectos importantes do ambiente VoIP, ignoradas pelo PSQM. Porém, como este é um método cujos padrões não são formalmente estabelecidos e como sua documentação é bastante restrita, o mesmo não será usado no nosso trabalho.

Maiores detalhes sobre o funcionamento do PSQM+ podem ser obtidas em [14].

### **4.5.4) PAMS**

O PAMS (do inglês, *Perceptual Analysis Measurement System*) é um método de medida objetiva de qualidade de voz para sistemas telefônicos, que tem como objetivo prever resultados obtidos em testes subjetivos. O mesmo foi desenvolvido em 1998 e inicialmente teve uma grande aceitação na Europa, porém não foi adotado por nenhum organismo internacional de normatização como padrão.



Assim como no PSQM, os testes de medida de qualidade objetivo com o PAMS são realizados apenas através de escuta. Os mesmos apresentam uma boa acurácia em seus resultados quando o sinal de fala é submetido aos seguintes processos/parâmetros: codificadores forma de onda, paramétricos e híbridos; transcódificações; nível do sinal de entrada no codificador; variação de oradores; súbita variação do atraso; *clipping* temporal ou de amplitude; e inserção de ruído.

Porém, este avaliador não se propõe a medir os efeitos dos seguintes impactos sobre o sinal original de fala: atraso; lentas variações do atraso; ganho/atenuação total do sistema; filtragem analógica; ruído de fundo presente na entrada; e música como sinal de entrada.

Como verificamos, este é um avaliador que apresenta características bastante positivas no que se refere à análise de sinais de fala. Porém, como o mesmo não é um padrão formalmente estabelecido e apresenta documentação restrita, o mesmo não será adotado neste trabalho.

Informações mais detalhadas sobre o funcionamento deste método podem ser obtidas em [14].

#### **4.5.5 PESQ (P.862)**

O PESQ (do inglês, *Perceptual Evaluation of Speech Quality*), assim como seus antecessores, é direcionado para medidas de sinais de voz na faixa telefônica (300 a 3.400 Hz), sendo aplicável a sistemas cujas características englobem fala codificada, atraso variável, erros de canal e perdas de pacotes. Esta recomendação combina a precisão do modelo perceptual do PSQM+, com a robustez das técnicas de alinhamento no tempo do PAMS.

Assim como no PSQM, a medida de qualidade objetiva do PESQ é obtida apenas em testes de escuta. O mesmo apresenta uma precisão bastante aceitável de seus

resultados, quando o sinal de voz original é exposto a: codificadores forma de onda, paramétricos ou híbridos; transcódificações; erros no canal de transmissão; efeitos da variação do atraso nos testes de escuta; perdas de pacotes com codificadores paramétricos ou híbridos; ruído ambiente na transmissão; e deformações temporais do sinal de áudio.

Por um outro lado, não há precisão nas medidas afetadas pelos seguintes parâmetros (ou mesmo não é pretendido medi-los): atraso; eco percebido pelo orador; som da própria voz ouvida pelo orador por retorno; níveis de escuta (ganho/atenuação total do sistema).

Ainda não há na literatura informações precisas sobre como esta recomendação se comporta quando sob a influência dos seguintes parâmetros: perda de pacotes com codificadores por forma de onda; *clipping* temporal e de nível; dependência do orador; múltiplos oradores; música como sinal de entrada; eco percebido pelo ouvinte; diferença entre as taxas de codificação e decodificação.

O processo de atribuição da pontuação final é obtido calculando-se a distância entre o sinal original e o sinal degradado, sendo o procedimento completo de obtenção do mesmo descrito em [14]. A faixa de pontuação varia entre -0,5 e 4,5, sendo que na maior parte dos casos esse valor obtido é bastante semelhante àqueles observados nos teste de escuta do tipo MOS.

Na recomendação em questão não há nenhuma referência relacionando os números obtidos por um codificador, quando submetido ao PESQ, com a qualidade real do sinal de voz distorcido. Porém há um consenso (até certo ponto) entre os profissionais de que pontuações iguais ou superiores a 3,0 representam um sinal de boa qualidade, enquanto valores inferiores a este limite representam um sinal mais pobre no que se refere à qualidade. Codificadores que apresentem valor abaixo deste limite não são adequadas ao uso comercial/corporativo, a não ser quando não há soluções superiores ou esta representa uma solução de baixo custo.

Como pudemos ver, este é um método de avaliação de codificadores de voz em testes de escuta simples. Porém, este é um padrão formalmente estabelecido e cuja documentação foi/é amplamente divulgada pelo ITU. Esta recomendação possibilita que seja notada a interferência causada pela perda de pacotes e variação do atraso, que são os pontos de observação utilizados para identificação das modificações de clareza notadas. Dessa maneira, este método será o escolhido para as medidas a serem efetuadas no decorrer deste trabalho.

#### **4.5.6) Mapeamento PESQ x MOS**

Como dissemos na subseção anterior, o PESQ (recomendação P.862 da ITU) será o nosso avaliador escolhido para as medidas que serão efetuadas posteriormente. Porém, os resultados finais que serão analisados não serão os números fornecidos pela saída do PESQ e sim através de um mapeamento PESQ x MOS, conforme o anexo P.862-1 da recomendação P.862 da ITU.

Vimos anteriormente que o MOS apresenta uma grande dificuldade de ser medido, por requerer que uma série de ouvintes estejam disponíveis para os testes. Porém, apesar disso, o MOS é largamente difundido na literatura, sendo referência principal quando um codificador de voz é citado. Dessa maneira, agregaram-se as vantagens das características analisadas pelo PESQ com a difusão do MOS e criou-se um mapeamento dos resultados obtidos pelo PESQ para que os mesmos sejam convertidos para o valor MOS. Este mapeamento é dado pela equação 4.1:

$$MOS = 0.999 + \frac{4}{1 + \left( e^{-1.495 \times PESQ + 4.6607} \right)} \quad (4.1)$$

Sendo assim, iremos utilizar como fator de análise final para os testes realizados os valores MOS obtidos através do mapeamento aqui exposto.

## 4.6) Conclusão

Neste capítulo pudemos verificar, primeiramente, qual o papel da QoS nos serviços VoIP. Ressaltamos o que de fato é QoS, levantando quão importante ela é para os serviços VoIP, uma vez que se trata uma novidade e sabemos que uma coisa nova só entra no mercado caso ela represente um ganho em relação à tecnologia anterior, já que as pessoas, em geral, têm uma certa restrição com relação a mudanças.

Além disso, verificamos qual o papel comercial da QoS e como as empresas/operadoras deverão atuar quando forem lançar mão da QoS em seus serviços, bem como observamos que o sucesso dos serviços VoIP com seu público alvo depende bastante da QoS.

Em seguida, passamos à verificação de quais parâmetros devem ser levados em consideração (variados) quando da análise da QoS. Chegamos à conclusão de que para os propósitos deste trabalho, três parâmetros (atraso, *jitter* e perda de pacotes) são de extrema importância, e serão considerados no desenvolvimento da parte prática deste trabalho.

Finalmente, consideramos uma série de avaliadores de QoS, buscando caracterizar cada um deles, levantando vantagens e desvantagens e os motivos pelos quais iríamos ou não utilizá-los ou não ao longo deste trabalho. Ao final desta seção (e do capítulo, por consequência) optamos por somente utilizar o PESQ, da recomendação P.862 da ITU.

Com isso, a parte teórica deste trabalho está concluída. No próximo capítulo serão mostrados os testes realizados, bem como uma análise dos resultados obtidos.

# Capítulo 5 - Testes, Resultados e Análises

## 5.1) Introdução

Neste capítulo faremos uma descrição detalhada dos testes realizados ao longo deste trabalho. Como dissemos nos capítulos anteriores, o foco deste trabalho está na avaliação do desempenho dos codificadores em ambiente VoIP. Consideraremos, para tais testes, os codificadores apresentados anteriormente. Para suas respectivas análises, iremos variar os parâmetros mencionados no Capítulo 4 e utilizaremos o analisador de QoS o mencionado ao final daquele capítulo.

Na seção 5.2 descreveremos os testes realizados, bem como quais dos codificadores apresentados anteriormente foram escolhidos para os testes. Nesse momento deixaremos claro quais motivos nos levaram a usar cada um destes. Além disso, será apresentado qual banco de dados foi adotado para execução dos testes.

Já na seção 5.3 serão apresentados os resultados dos testes e será feita uma extensa análise dos resultados obtidos. Nesta seção encontra-se o principal objetivo principal deste trabalho: estaremos nela comparando o desempenho de cada um dos codificadores analisados relativamente, ou seja, buscaremos qual deles apresenta, de modo geral, melhor desempenho no ambiente VoIP.

Na seção 5.4 serão mostrados os resultados de testes subjetivos que foram inseridos neste trabalho de modo a verificar os resultados dos testes objetivos. Iremos buscar, através destes, validar os resultados obtidos nos testes da seção 5.3.

Finalmente, na seção 5.5, iremos buscar uma conclusão para os resultados obtidos ao longo deste trabalho, procurando estabelecer uma tendência de comportamento das

famílias de codificadores no ambiente VoIP, de modo a criarmos uma referência para futuros estudos nessa área.

## 5.2) Métodos Utilizados

No capítulo 2 apresentamos uma série de codificadores (de diversas famílias) aprovados pelo ITU bem como um codificador CELP desenvolvido em [1], o qual passará a ser notado como *CELP-UFRJ*. Além disso, colocamos que estes seriam as ferramentas utilizadas como parâmetros em nossos testes. Dentre os codificadores apresentados, selecionamos para testes:

- O G.711, codificador PCM, normalmente utilizado como referência na maioria dos trabalhos, pela sua difusão, praticidade, utilização e relativa simplicidade, além de ser utilizado na telefonia digital fixa no Brasil e em outros países;
- O G.726, codificador ADPCM, variante do PCM original, também bastante difundido devido às suas qualidades, conforme apresentado no capítulo 2, utilizado neste trabalho com a taxa de 32 kbits/s, conforme utilizado na telefonia digital fixa nos EUA;
- O G.729, codificador da família CELP de bastante sucesso. Além deste codificador propriamente dito, submetemos o G.729 aos anexos A e B (bem como os dois juntos), uma vez que estes tratam de aspectos bastante relevantes no que se refere a sinais de voz (redução de complexidade computacional e supressão de silêncio, respectivamente);
- E o CELP-UFRJ, outro codificador híbrido com características semelhantes ao G.729. Este codificador, como já colocado, não foi desenvolvido pelo ITU e, sendo assim, possui uma relevância bastante

importante para este trabalho, porque apesar de ser da mesma família do G.729, foi desenvolvido de forma distinta em relação aos outros 3 codificadores.

Estes codificadores foram obtidos através de *downloads* gratuitos no site do ITU, a exceção do CELP-UFRJ que foi disponibilizado para utilização pelo seu autor.

No capítulo 2 ainda foram apresentados os codificadores G.723.1 e G.728, também representantes da família dos híbridos. Porém, devido a dificuldades com relação à operação adequada destes, bem como obtenção dos códigos-fonte correspondentes, os mesmos não serão testados neste trabalho, ficando como sugestão para trabalhos futuros.

Para a realização dos testes, precisamos de um banco de dados de sinais de voz. Para tanto, lançamos mão de um conjunto de 20 frases, todas em português (do Brasil), foneticamente balanceadas que nos serviram como entrada e referência para todos os procedimentos adotados. Os resultados que serão apresentados na próxima seção, são frutos da média dos resultados individuais de cada uma dessas frases. O objetivo de trabalharmos com um banco de dados relativamente extenso está na minimização dos possíveis erros, ou melhor, na busca por uma melhor acurácia.

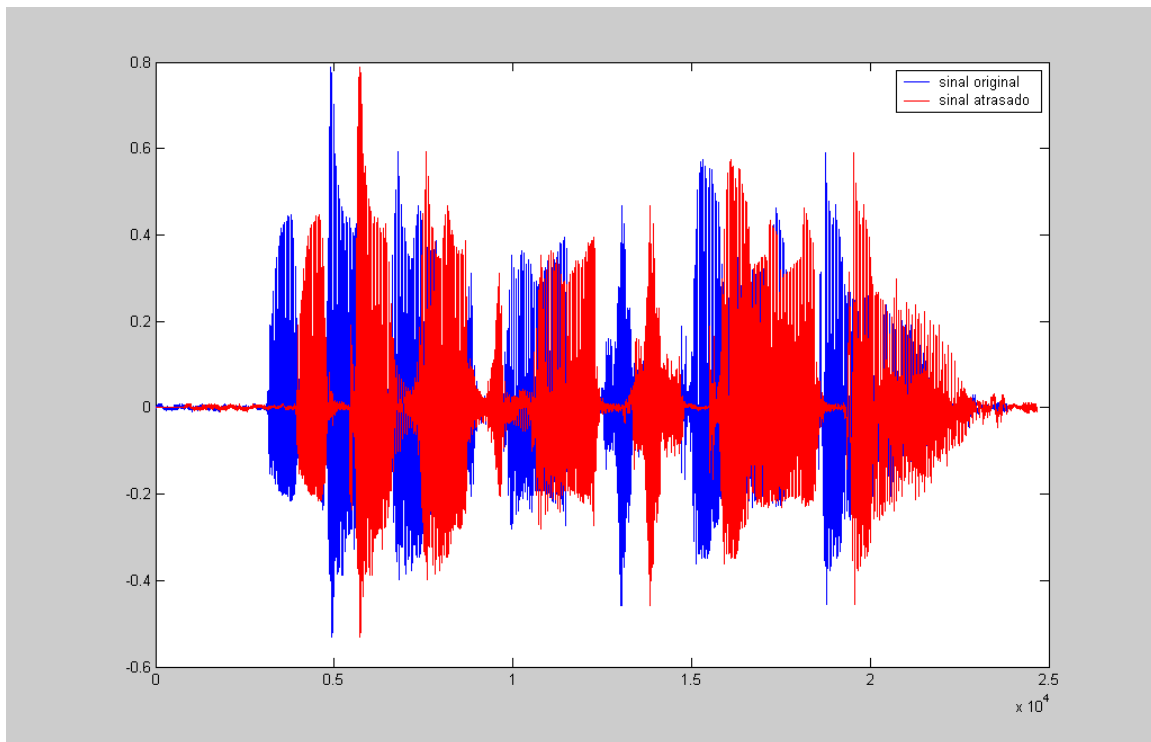
### **5.3) Testes**

Nesta seção apresentaremos os resultados obtidos para os testes realizados ao longo deste trabalho, bem como uma análise absoluta e relativa de cada um dos codificadores. Conforme foi descrito no capítulo 4, três parâmetros seriam colocados em prova, de modo a verificarmos a qualidade de cada um dos codificadores em ambiente VoIP, sendo eles o atraso, o *jitter* (variação do atraso) e a perda de pacotes. Nas subseções a seguir cada um destes será o parâmetro em questão e, como já foi colocado, procuraremos identificar quais codificadores que apresentam um melhor desempenho quando submetidos a elementos constantemente presentes em comunicações VoIP.

### 5.3.1) Atraso

Como exposto anteriormente o atraso fim-a-fim é um parâmetro de extrema importância em comunicações. Quando dois usuários estabelecem uma conexão de voz, não é de interesse dos mesmos que haja uma grande demora para recepção dos sinais. Isso causa irritação e, quando acontece constantemente, faz com que o usuário não utilize mais aquele serviço.

Para simularmos o atraso nos sinais de voz, adicionamos zeros em cada um dos sinais. A figura 5.1 mostra um sinal de voz e o mesmo sinal de voz atrasado de 100 ms.

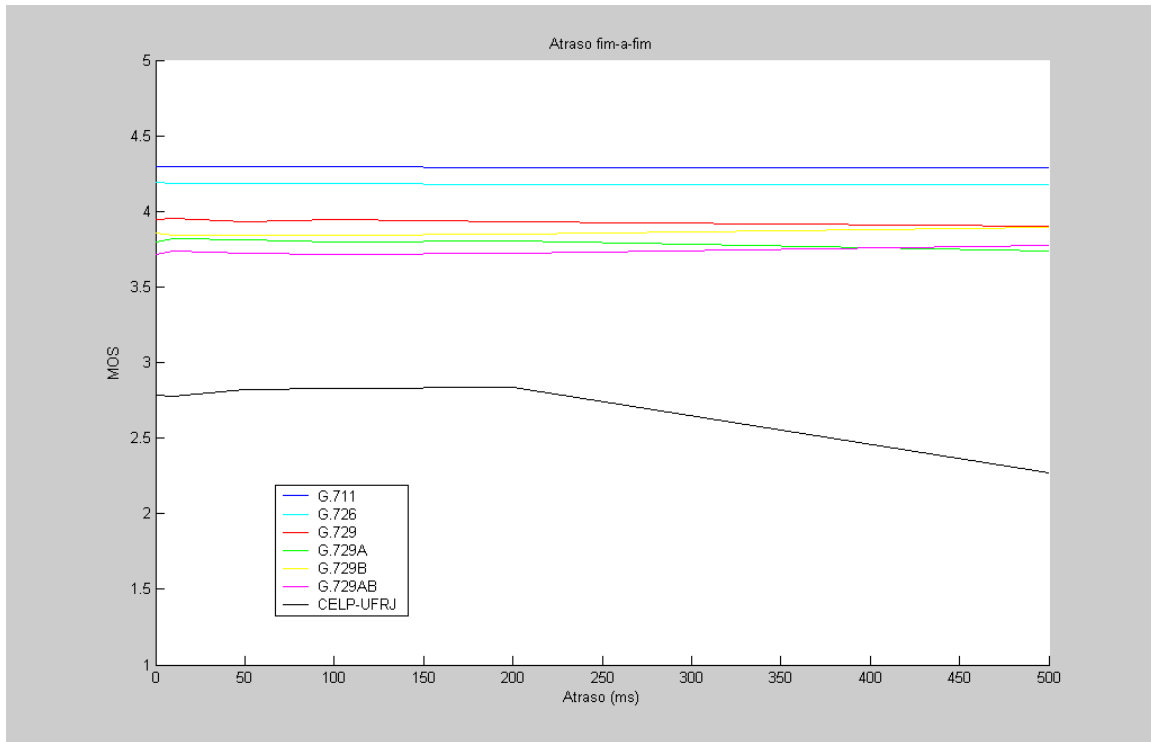


**Figura 5.1:** Sinal de voz e sua reprodução atrasada de 100 ms.

Consideramos no nosso trabalho atrasos de 10, 50, 100, 200 e 500 ms. Submetemos cada uma das vinte frases do nosso banco de dados a inserção de zeros proporcional ao atraso desejado e em seguida passamos cada uma delas por cada um dos codificadores. Em seguida, submeteremos cada uma das frases geradas ao PESQ e



calculamos a média. Estas médias são os valores representados no gráfico mostrado na figura 5.2, onde cada uma das curvas tem como parâmetro cada um dos codificadores previamente mencionados.

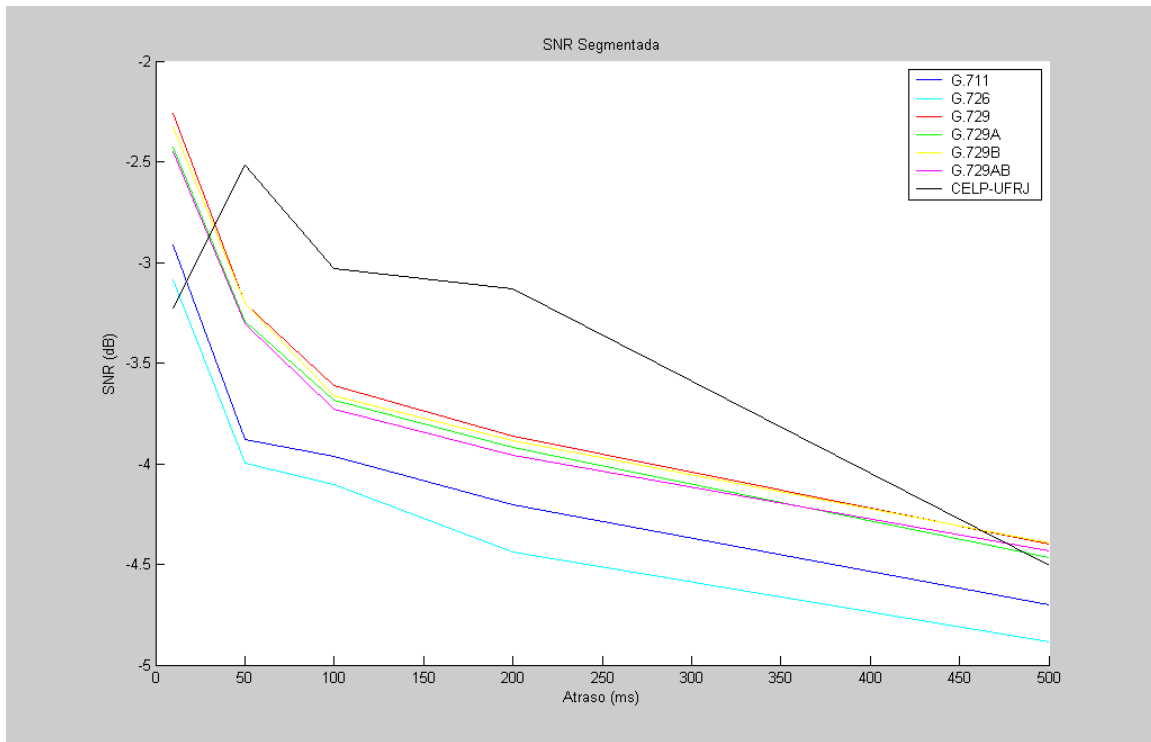


**Figura 5.2:** Desempenho dos codificadores com relação à variação do atraso fim-a-fim

Dessa maneira, as curvas apresentadas na figura 5.2 nos levam a uma conclusão bastante direta. Tendo em vista a uniformidade de desempenho apresentada, podemos inferir que o avaliador em questão, apesar de sua larga difusão na literatura referente, é bastante insensível com relação aos atrasos como foi previsto na seção 4.5.5.

Sendo assim, fomos levados a buscar uma alternativa para o PESQ para medirmos o efeito do atraso sobre os codificadores de voz. Para isso, lançamos mão da razão sinal-ruído perceptual segmentada (do inglês, *Perceptual Segmented-SNR*), uma medida alternativa. Os resultados podem ser observados a seguir na figura 5.3.

De modo geral, a figura 5.3 indica que quanto maior o atraso, maior é a distância (de acordo com a métrica escolhida) entre os dois sinais, como era de se esperar. Este teste, porém, não gera nenhum resultado concreto do aspecto perceptual. Ou seja, não conseguimos avaliar como o atraso afeta na prática, para um usuário humano, o desempenho de um codificador. Procuraremos fazer este tipo de avaliação subjetiva mais adiante neste capítulo.



**Figura 5.3:** Desempenho dos codificadores de voz com relação ao atraso fim-a-fim

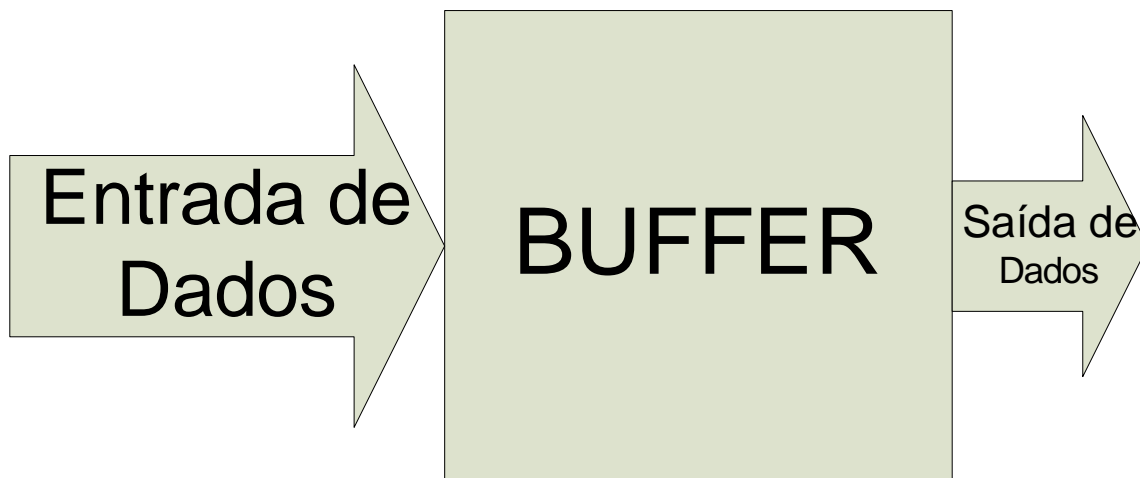
Dessa maneira, tendo estabelecido um primeiro parâmetro com relação ao comportamento de codificadores de voz, vamos agora partir o próximo parâmetro considerado, a variação do atraso, ou simplesmente, *jitter*.

### 5.3.2) Jitter

Este parâmetro não é comumente observado entre os usuários comuns, pelo menos de maneira direta. Isto ocorre porque o *jitter* é a variação do atraso fim-a-fim

médio, conforme descrito no capítulo 4. Esta variação pode levar os pacotes de voz a chegarem numa ordem não-coerente, fazendo com que o usuário tenha a sensação de que existe a superposição ou mesmo a perda de uma parte da fala. Por exemplo, se temos um atraso fim-a-fim de 100 ms e um jitter médio de 30ms, um determinado pacote pode chegar com 130 ms de atraso e o pacote subsequente com 70ms de atraso. Este fato, ocorrendo repetidamente pode levar o usuário a se irritar e com isso não mais utilizar o serviço de comunicação em questão. Por isso uma boa administração desse parâmetro é importante nos serviços VoIP.

Conforme explicado no capítulo 4, níveis excessivos de *jitter* podem ser contornados com a presença de um *buffer*, que, como todos os outros componentes da rede, deve ser dimensionado. Lembrando o que foi citado no capítulo anterior, um *buffer* teria o papel semelhante ao de uma caixa d'água. Sua vazão de dados na entrada e sua capacidade de armazenamento de dados são bem superiores àquela de saída de dados. A figura 5.4 mostra a atuação de um *buffer*.



**Figura 5.4:** *Buffer* de dados, cujo papel é a minimização do *jitter*.

Este é um ponto crucial de nossa análise: a análise da variação do *jitter* em função de um determinado atraso passa a ficar em segundo plano, uma vez que bem dimensionemos um *buffer* que seja capaz de reduzir o *jitter* para níveis bastante aceitáveis, recaindo no problema do atraso, analisado na subseção anterior.

Dessa maneira, fizemos a opção por não fazer uma extensa análise do *jitter* como fizemos para o atraso, uma vez que julgamos que um *buffer* adequado é capaz de solucionar este problema, desde que os níveis de atraso não sejam exacerbados. Este fato, como já colocado, levaria o usuário a dispensar o uso deste serviço.

Sendo assim, passamos para análise do terceiro e último parâmetro de nossa análise de codificadores em ambiente VoIP, a perda de pacotes.

### **5.3.3) Perda de Pacotes**

Este é um parâmetro de alta importância (bem como os dois anteriores) quando nos referimos a comunicações VoIP. Como já foi mencionado neste trabalho, o protocolo IP possui uma preocupação bastante grande com a velocidade, deixando de lado, até certo ponto, a confiabilidade com relação à entrega dos pacotes. Este fato realça mais ainda a importância de estudarmos como os codificadores se comportam com relação à perda de pacotes.

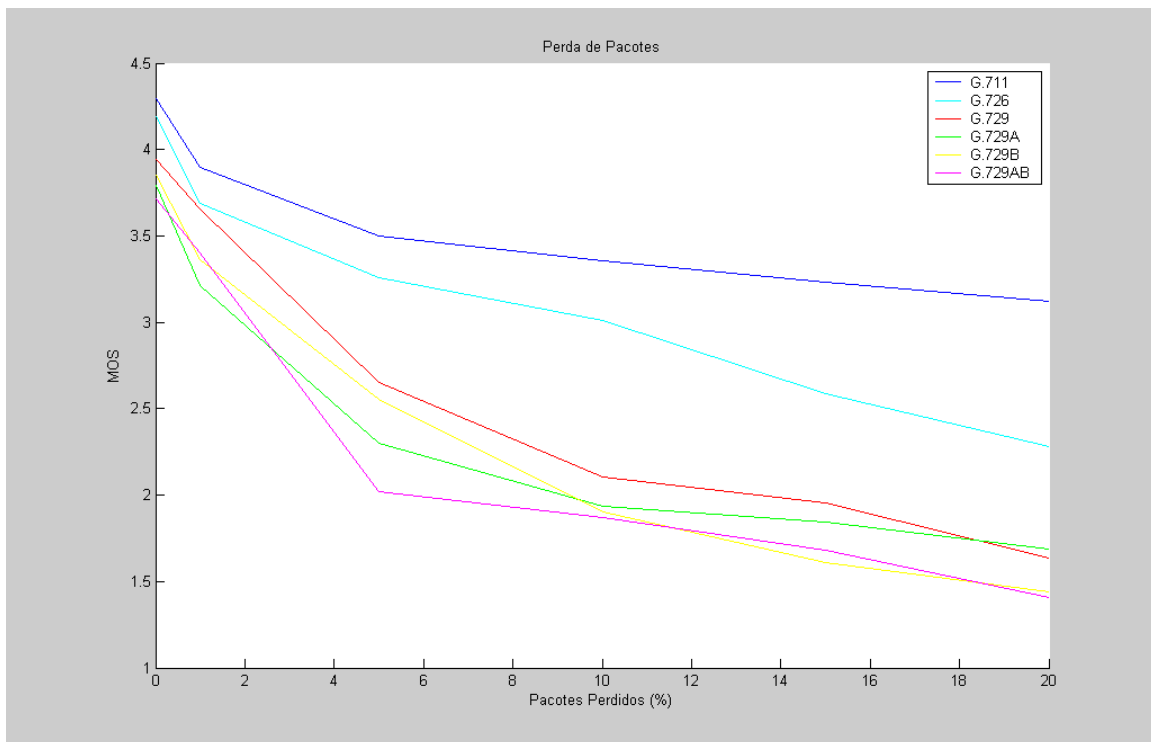
Vimos nas últimas duas subseções que a partir de uma determinada condição (alocação de *buffer* adequado) os problemas de atraso e de *jitter* podem ser interpretados como um único problema, o do atraso. Quando paramos para analisar o atraso, verificamos que este se trata de um problema estacionário. Isto ocorre porque o atraso significa uma simples demora para o recebimento do sinal, e se repete uniformemente ao longo da comunicação estabelecida.

Para a perda de pacotes passamos a ter que considerar em qual momento do sinal este é perdido, pois, com exceção do G.711, todos os outros codificadores possuem um módulo adaptativo. Este fato pode levar um sinal cujas perdas foram concentradas no início da transmissão a ter um desempenho bastante pior do que um sinal cujas perdas ocorreram quando do término da transmissão.

Porém, os sinais de voz avaliados neste trabalho são de duração relativamente pequena, não superiores 2 segundos, o que, através de testes preliminares, nos levou a descartar esta hipótese da não estacionariedade da perda de pacotes. No entanto, este fenômeno pode estar presente quando da transmissão de sinais de voz mais longos.

Para a realização dos testes variamos a perda de pacotes entre 1 e 20%, aleatoriamente, o decorrer de cada um dos 20 sinais de entrada. Utilizamos todos os codificadores apresentados nas subseções anteriores deste Capítulo, à exceção do CELP-UFRJ cuja implementação impede este tipo de teste.

A figura 5.5 mostra a seguir o desempenho dos codificadores quando submetidos à perda de pacotes entre a codificação e a decodificação.



**Figura 5.5:** Desempenho dos codificadores de voz com relação à perda de pacotes.

Com relação a este parâmetro, podemos verificar que os codificadores se comportam de forma bastante diferente do que com relação ao atraso fim-a-fim. Os codificadores em forma de onda apresentam um desempenho superior do que os codificadores híbridos, sendo o G.711 o que melhor lida com a perda de pacotes.

Esta conclusão já era esperada, uma vez que o G.711, como já foi colocado, não possui um módulo adaptativo em sua arquitetura, o que implica na não-propagação de eventuais perdas ao longo da transmissão. Por outro lado, esta propagação ocorre nos outros codificadores avaliados, o que justifica o seu desempenho inferior.

Dessa maneira, podemos concluir, diferentemente do que ocorreu com relação ao atraso, que os codificadores por forma de onda apresentam um desempenho superior aos da família dos codificadores híbridos. Este é um fato bastante importante, uma vez que verificamos que a escolha de um codificador, ou ainda, de uma família de codificadores de voz para a transmissão sob os parâmetros avaliados neste trabalho não é trivial.

## **5.4) Testes Subjetivos**

De modo a complementarmos os testes objetivos realizados na seção anterior, vamos neste momento realizar alguns testes subjetivos. O objetivo neste momento é verificar se as conclusões obtidas para os testes objetivos se repetem com os subjetivos ou se encontramos resultados divergentes.

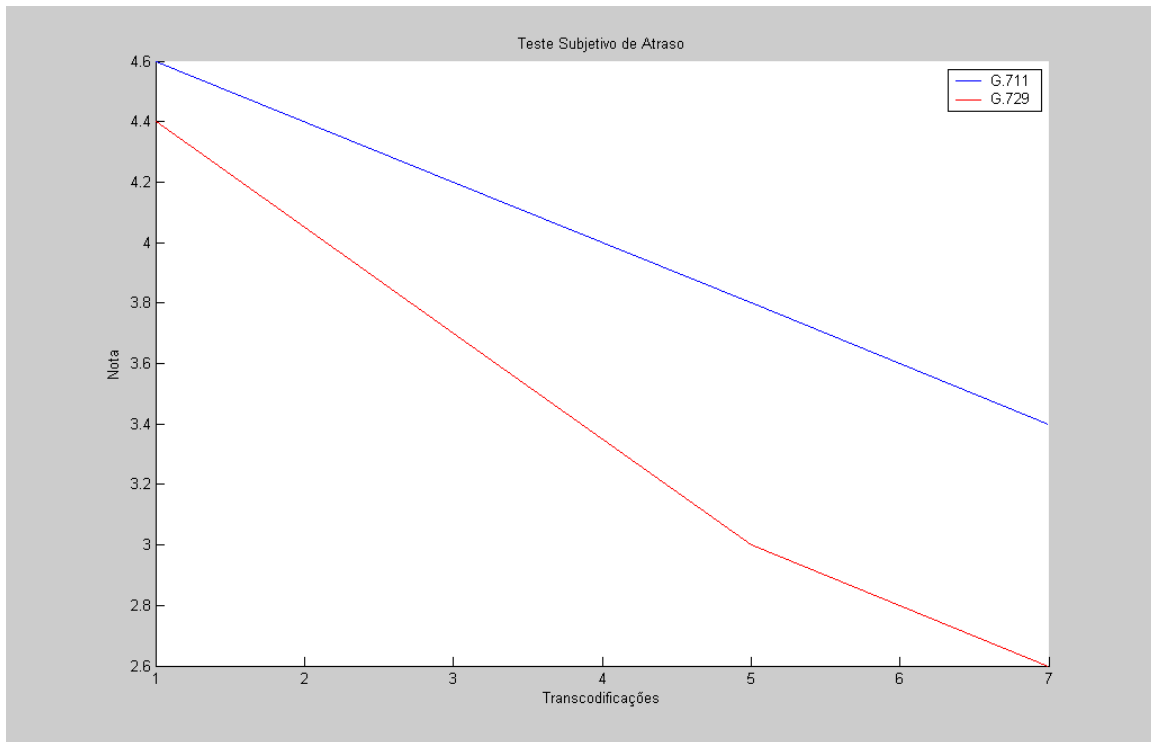
Para tais testes procuramos adotar um procedimento padronizado, de modo ao mesmo apresentar resultados fidedignos.

Para os testes referentes ao atraso, submetemos simulações de conversa de duração média (de 5 a 10 segundos) a várias transcódificações. Nosso objetivo, que foi apresentado ao ouvinte antes do teste era procurar saber quanto à demora no recebimento de uma resposta incomodaria o mesmo. As sucessivas transcódificações representam o efeito do atraso crescente. Para estes testes, três simulações de conversa de média

duração a 1, 3, 5 e 7 transcodificações sucessivas através dos codificadores G.711 e G.729, a fim de verificarmos quais diferenças de comportamento de duas famílias de codificadores distintas. Para a execução dos mesmos, foram escolhidos 10 usuários que realizaram os testes em um mesmo computador, sob um treinamento prévio que os instruíu a atribuir uma nota de 1 a 5, sendo a nota 1 correspondente a um atraso intolerável e a nota 5 correspondente a um atraso que não causava incômodo algum. Os usuários não tiveram quaisquer restrições quanto ao número de vezes que poderiam escutar cada uma das frases nem com relação ao treinamento, que poderia ser repetido ao longo do teste, caso fosse do interesse do usuário.

É importante notar aqui que chegamos ao número de 10 usuários durante a realização deste teste. Com este número de usuários, chegamos a uma convergência das notas atribuídas, sem a presença de discrepâncias. Além disso, é importante colocarmos que todos os usuários eram leigos com relação ao assunto, com idade entre 20 e 30 anos, sendo 50% dos usuários do sexo masculino e 50% do sexo feminino.

Os resultados destes testes podem ser observados através da figura 5.6, a seguir.



**Figura 5.6:** Desempenho dos codificadores de voz com relação ao atraso.

Através da figura acima podemos observar que o codificador G.711 apresenta um comportamento superior em relação ao G.729. Este se deve ao fato de que cada transcódificação do G.711 leva um tempo inferior ao G.729 e sua consequência é a acentuada divergência entre as curvas ao passo que aumentamos o número de transcódificações.

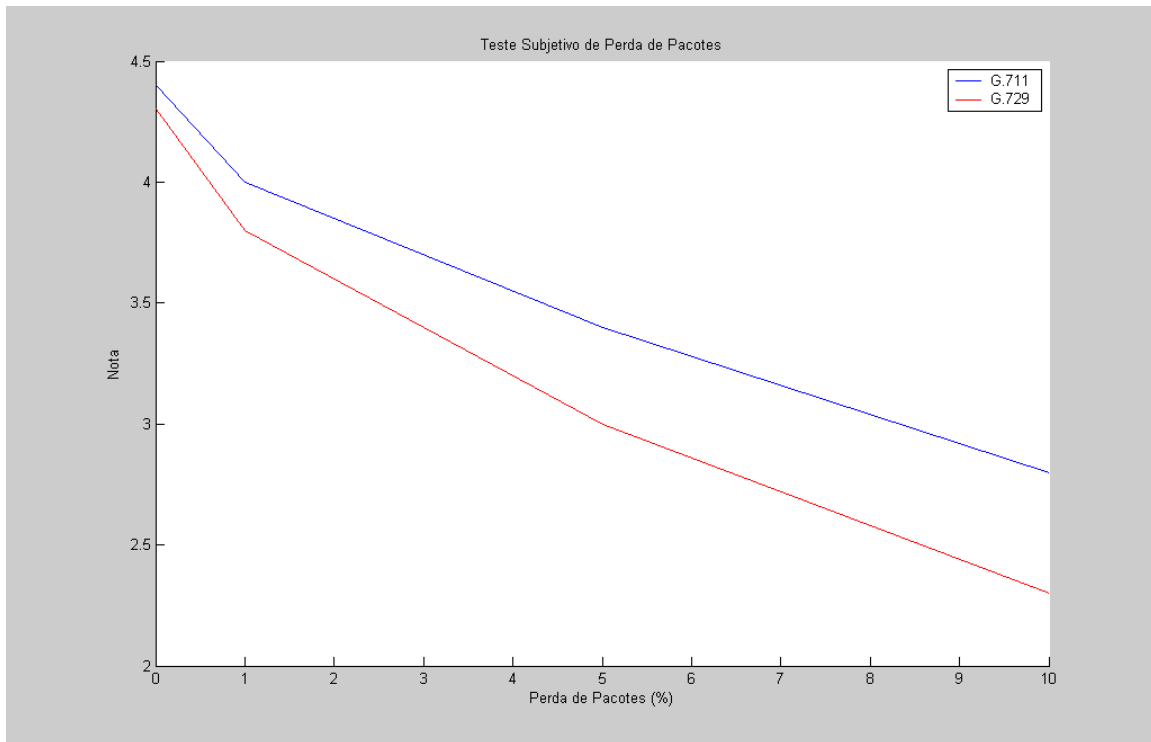
Este gráfico nos proporciona um dado bastante importante. Quando analisamos o atraso objetivamente não pudemos verificar como um determinado usuário se incomodaria com excessos de atraso. Já os testes subjetivos puderam nos levar a uma conclusão mais precisa. Devido a um atraso interno maior, os codificadores CELP possuem uma tendência maior de provocar atrasos superiores em relação aos codificadores PCM, o que revela, pelo menos segundo esta análise subjetiva para o atraso, que os codificadores PCM são superiores em relação aos CELP.



Já para os testes subjetivos com relação à perda de pacotes, um procedimento similar foi adotado. Submetemos cinco das frases utilizadas nos testes objetivos a perdas de 0, 1, 5 e 10% de seus pacotes entre a codificação e a decodificação, também dos codificadores G.711 e G.729, mais uma vez a fim de verificarmos o comportamento de duas famílias distintas de codificadores. Conforme nos testes com relação ao atraso, foram utilizados 10 usuários, previamente treinados, que deveriam atribuir notas de 1 a 5, sendo a nota 1 correspondente a uma qualidade de voz sofrível e a nota 5 correspondente a uma qualidade de voz inaceitável. Mais uma vez não houve restrições quanto ao número de vezes que um determinado usuário poderia escutar uma frase nem com relação ao treinamento, que poderia ser repetido a qualquer momento do teste.

Mais uma vez, chegamos ao número de 10 usuários com o decorrer do teste, uma vez que esta quantidade apresentava uma convergência dos resultados, bem como uma ausência de discrepância. As características com relação a sexo, faixa etária e conhecimento do assunto são iguais àquelas do teste subjetivo de atraso.

Os resultados dos testes subjetivos com relação à perda de pacotes podem ser observados através da figura 5.7, apresentada a seguir.



**Figura 5.7:** Desempenho dos codificadores de voz com relação à perda de pacotes.

Observando a figura 5.7 verificamos que um comportamento semelhante aos testes objetivos, onde verificamos que os codificadores PCM possuíam um desempenho superior ao dos codificadores CELP. Estes resultados, primeiramente servem para confirmar a eficiência do avaliador objetivo que utilizamos (PESQ). Além disso, mostram que realmente há uma tendência de melhor desempenho dos codificadores em forma de onda do que os codificadores híbridos, quando submetidos a este parâmetro das redes onde há tráfego de voz sobre IP.

Contudo, os testes subjetivos nos levaram a uma importante conclusão. Ao contrário dos testes objetivos, onde obtivemos conclusões distintas para os dois parâmetros avaliados, aqui obtivemos conclusões semelhantes, de que os codificadores em forma de onda representam uma melhor solução para transmissão de voz sobre IP (com relação aos codificadores híbridos) pelo menos com relação aos parâmetros avaliados neste trabalho.

## 5.5) Conclusão

Neste capítulo procuramos validar de maneira prática, tudo aquilo que foi mencionado na introdução e nos três capítulos anteriores.

Primeiramente, apresentamos os codificadores que utilizamos para os testes, procurando enfatizar as características, vantagens e desvantagens de cada um deles, complementando o que já fora mencionado no capítulo 2. Além disso, expusemos com qual banco de dados de sinais de voz iríamos realizar os testes, bem como enfatizamos a utilização do avaliador MOS de qualidade de voz, devido à sua difusão na literatura e aos outros fatos apresentados no capítulo 4.

Em seguida, descrevemos os procedimentos adotados para cada um dos parâmetros testados. Conforme exposto, obtivemos sucesso na análise do atraso, do *jitter*, cuja solução recai no problema do atraso, e da perda de pacotes.

Com relação ao atraso verificamos que o PESQ é insensível a este fenômeno. Desta maneira, procuramos utilizar uma forma alternativa de análise, que foi a relação sinal-ruído segmentada. A partir desta, concluímos que, pelo menos quando submetemos codificadores de voz a atrasos, codificadores híbridos são superiores a codificadores forma de onda.

Em contra partida, quando avaliamos a perda de pacotes na transmissão, os codificadores por forma de onda apresentaram um desempenho superior em relação aos híbridos. Este fato nos levou a uma importante conclusão: os testes objetivos não nos proporcionaram uma certeza com relação a que família de codificadores é mais adequada para transmissão VoIP.

Isto nos levou a procurar por respostas mais precisas o que foi realizado através de testes subjetivos. Ao contrário dos objetivos, estes testes proporcionaram conclusões

bem mais precisas: codificadores em forma de onda se comportam realmente de forma melhor que os codificadores híbridos, ainda que estes últimos apresentem um desempenho bastante satisfatório. Esta conclusão deste capítulo nos leva a uma tendência de utilização dos codificadores em forma de onda quando da transmissão VoIP. Porém, nada impede que codificadores híbridos sejam utilizados, ainda mais em ambientes restritos com relação a largura de banda, tendo em vista sua baixa taxa de codificação e desempenho satisfatório, apesar de inferior em relação aos em forma de onda.

Dessa maneira, concluímos o nosso capítulo de testes, resultados e avaliações, com sucesso. Sendo assim, agora iremos para o último capítulo deste trabalho, onde faremos um sumário de todos os tópicos (teóricos e práticos) abordados ao longo deste projeto, verificaremos que contribuições este trabalho pode proporcionar e apresentaremos uma lista de propostas de trabalhos futuros, de modo a convergirmos para uma resposta com relação a que codificador e/ou família de codificadores usar quando trabalhando em ambiente VoIP.

# Capítulo 6 - Conclusão

## 6.1) Resumo do Projeto e Conclusão

Neste capítulo faremos um pequeno resumo de tudo que foi abordado, tentando ressaltar os aspectos, resultados e conclusões mais importantes. Iremos também apontar quais as formas com que este projeto pode contribuir para os estudos nesta área de comunicações VoIP e tentaremos traçar uma diretriz para trabalhos futuros.

Como propusemos inicialmente, este projeto teve objetivos teóricos e práticos. Isto se traduziu em um estudo teórico de codificadores de voz, do ambiente VoIP e da QoS. Já os objetivos práticos foram traduzidos em uma série de testes, que objetivaram observar o comportamento dos codificadores quando da variação de parâmetros pré-determinados.

Dedicamos o Capítulo 2 à compreensão do que são e como funcionam os codificadores de voz existentes hoje. Neste capítulo foram descritas as três famílias de codificadores mais comuns (por forma de onda, paramétricos e híbridos), foram apresentados exemplos de codificadores pertencentes a cada uma destas famílias, bem como pré-selecionamos alguns destes para os testes que se seguiram.

Já o Capítulo 3 foi totalmente dedicado à descrição do ambiente VoIP. Procuramos inicialmente deixar bem claro o que é um ambiente desse porte, ressaltando que os possíveis clientes dos serviços VoIP podem ter requisitos distintos. Em seguida descrevemos brevemente como é a arquitetura básica VoIP, seguido de uma descrição detalhada dos dois protocolos mais comuns atualmente, bem como citamos outras alternativas menos comuns. Finalizamos este capítulo dando exemplos de aplicativos VoIP existentes no mercado, enumerando e destacando suas respectivas características, vantagens e desvantagens.

No Capítulo 4 procuramos deixar bem claro o que é a QoS. Inicialmente destacamos a importância da QoS em um ambiente VoIP. Isto foi seguido por uma análise dos aspectos comerciais da QoS, procurando deixar claro como as operadoras de telecomunicações, que foram explorar serviços VoIP (ou ainda “tudo sobre IP”), terão de explorar a QoS, tendo em vista os seus usuários. Em seguida, foram analisados uma série de parâmetros a serem considerados na análise da QoS, onde procuramos mostrar a importância/relevância de cada um deles. Foram selecionados três destes (atraso fim-a-fim, *jitter* e perda de pacotes) para os testes que foram realizados a seguir. Finalmente, apresentamos uma série de avaliadores de QoS e concluímos que um destes deveria ser utilizado em nossos testes.

De maneira a complementar aos objetivos teóricos de nosso trabalho, apresentamos os objetivos práticos no Capítulo 5. Estes foram mostrados através de testes com um grupo de codificadores selecionados, previamente apresentados e descritos no capítulo 2. Consideramos os parâmetros escolhidos no capítulo 4, bem como o avaliador de QoS selecionado neste mesmo capítulo. Com relação ao atraso fim-a-fim e ao *jitter*, pudemos obter conclusões bastante satisfatórias, uma vez que o comportamento de cada um dos codificadores pôde ser observado ao passo que variamos o atraso, onde verificamos uma insensibilidade do principal avaliador selecionado para detecção do mesmo. Isto nos levou a analisarmos este fenômeno com uma ferramenta auxiliar, que nos revelou um melhor desempenho dos codificadores híbridos com relação aos codificadores por forma de onda. Já com relação ao *jitter*, que nada mais é do que a variação do atraso, pode ser comprovado que com o bom dimensionamento de um *buffer* este problema é minimizado e nos faz recair no problema do atraso fim-a-fim. Já com relação à perda de pacotes obtivemos uma constatação oposta àquela que obtivemos em relação ao atraso. Este fato nos levou a lançar mão de testes subjetivos para que obtivéssemos uma conclusão um tanto mais concreta. E através destes pudemos verificar que os codificadores em forma de onda apresentam um desempenho superior em relação aos híbridos. Porém, os híbridos ainda sim apresentam um comportamento satisfatório e seu uso pode ser bastante importante em ambientes cuja taxa de codificação é um parâmetro importante.

## 6.2) Propostas de Trabalhos Futuros

A seguir, mencionamos algumas propostas para trabalhos futuros que tenham como objetivo a continuação e aprimoramento deste projeto.

- O estudo de novos codificadores, principalmente da família dos codificadores híbridos;
- A análise de novos protocolos VoIP, bem como do papel futuro dos protocolos mais presentes hoje no mercado;
- A análise dos aplicativos comerciais VoIP que já existem hoje, dos que estão para surgir, verificando qual influência os mesmos terão nos serviços de comunicação de todos os tipos num futuro próximo;
- A verificação da importância de outros parâmetros na QoS de serviços VoIP, como por exemplo, eco e supressão de silêncio.

## Referências Bibliográficas

- [1] DINIZ, F.C. da C. B., *Implementação de um Codificador de Voz CELP em Tempo Real*, Projeto Final-DEL-POLI-UFRJ, Rio de Janeiro, RJ, Brasil, 2003
- [2] DINIZ, P. S. R., Da SILVA, E. A. B., NETTO, S. L., *Digital Signal Processing: System Analysis and Design*, Cambridge, UK, Cambridge, 2002
- [3] DELLER, J. R., PROAKIS, J. G., HANSEN, J. H. L., *Discrete Time Process of Speech Signals*, New York, NY, USA, MacMillan, 1993
- [4] MAIA, R. da S., *Codificação CELP e Análise Espectral de Voz*, Tese de M.Sc., PEE-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2000
- [5] HAYKIN, S., *Communication Systems*, McMaster University, USA, John Wiley & Sons, 2001
- [6] CISCO SYSTEMS, Inc. (<http://www.cisco.com>)
- [7] <http://www.javvin.com/protocolQ931.html>
- [8] <http://asn1.elibel.tm.fr/en/>
- [9] <http://www.asciitable.com>
- [10] <http://www.w3.org/Protocols/HTTP/1.1/spec.html>
- [11] THE INTERNET ENGINEERING TASK FORCE (<http://www.ietf.org>)
- [12] ITU-T Rec. P.800, “Methods for Subjective Determination of Transmission Quality”, Aug. 1996.



[13] ITU-T Rec. P.861, “Objective Quality Measurement of Telephone-band (300-3400 Hz) Speech Codecs”, Feb. 1998.

[14] FERNANDES, N. L. L., *Relação Entre a Qualidade das Respostas das Recomendações G.723.1 e G.729, e o Comportamento da Rede IP de Suporte*, Tese de M.Sc., PESC-COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2003

[15] CAMPOS, A.M.D., *Transmissão de Voz Sobre IP com Taxas Variáveis*, Projeto Final-DEL-POLI-UFRJ, Rio de Janeiro, RJ, Brasil, 2005

[16] ITU-T Rec. G.711, “General Aspects of Digital Transmission Systems Terminal Equipments - Pulse Code Modulation (PCM) of Voice Frequencies“, 1972.

[17] ITU-T Rec. G.726, “General Aspects of Digital Transmission Systems Terminal Equipments - 40, 32, 24, 16 Kbits/s Adaptive Differential Pulse Code Modulation (ADPCM)“, 1972.

[18] ITU-T Rec. G.728, “General Aspects of Digital Transmission Systems; Terminal Equipments - Coding of Speech at 16 Kbits/s Using Low-Delay Code Excited Linear Prediction”, 1992.

[19] ITU-T Rec. G.729, “General Aspects of Digital Transmission Systems - Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)”, 1996.

[20] SALAMI, R., LAFLAMME C., BESSETTE, B., et al, “ITU-T G.729 Annex A: Reduced Complexity 8 kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data”, *IEEE Communication Magazine*, 1997.

[21] ITU-T Rec. G.723.1, “General Aspects of Digital Transmission Systems - Dual

Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 Kbits/s”, 1996.

[22] BENYASSINE, A., SHLOMOT, E., SU, H., et al., “ITU-T Recommendation G.729 Annex B: Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications”, *IEEE Communications Magazine*, Sept. 1997.