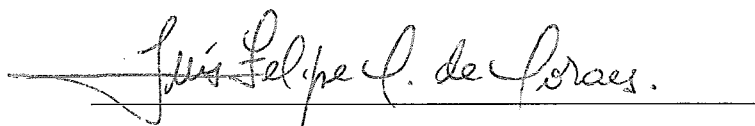


RELAÇÃO ENTRE A QUALIDADE DAS RESPOSTAS DAS RECOMENDAÇÕES  
G.723.1 E G.729, E O COMPORTAMENTO DA REDE IP DE SUPORTE

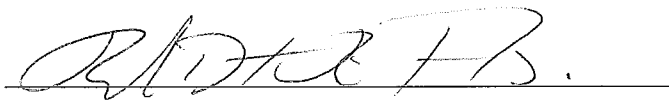
Nelson Luiz Leal Fernandes

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS  
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS  
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM  
ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

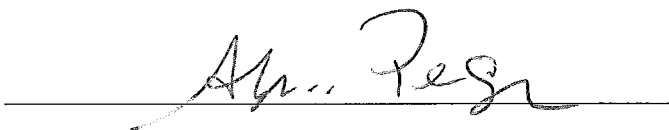
Aprovada por:



Prof. Luís Felipe Magalhães de Moraes, Ph.D.



Prof. Rubem Dutra R. Fagundes, Ph.D.



Prof. Aloysio de Castro Pinto Pedroza, Dr.



Prof. Paulo Roberto Rosa Lopes Nunes, Ph. D.

RIO DE JANEIRO, RJ - BRASIL  
MARÇO DE 2003

FERNANDES, NELSON LUIZ LEAL

Relação entre a Qualidade das Respostas das  
Recomendações G.723.1 e G.729, e o Compor-  
tamento da Rede IP de Suporte [Rio de Janeiro]  
2003

VIII, 161 p. 29,7 cm (COPPE/UFRJ, M.Sc.,  
Engenharia de Sistemas e Computação, 2003)

Tese - Universidade Federal do Rio de  
Janeiro, COPPE

1. Voz sobre IP
2. VoIP
3. Voz sobre Redes de Pacotes
4. Telefonia IP

I. COPPE/UFRJ II. Título (série)

## Agradecimentos

Aos meus pais, pelas oportunidades que me ofereceram toda vida, fazendo com que chagasse até aqui.

À minha esposa e filho, pela paciência e compreensão pelos muitos momentos de ausência da convivência familiar.

Ao Prof. Luís Felipe M. de Moraes, por ter me orientado de forma tão competente e profissional, harmonizando-se com minhas características e sempre presente quando necessário.

Ao Prof. Rubem Dutra Ribeiro Fagundes, por me honrar em tê-lo na banca examinadora. Figura representativa de minha primeira participação em simpósios da área, acolhendo meu trabalho de forma tão incentivadora.

Ao Prof. Aloysio de Castro Pinto Pedroza, por aceitar em fazer parte da banca examinadora desta tese. Representando todo o meu percurso acadêmico, por ter sido meu professor em disciplinas na graduação e na pós-graduação, desempenhou sua função de forma tão zelosa e respeitosa em relação aos seus alunos.

Ao Prof. Paulo Roberto Rosa Lopes Nunes, por me honrar em tê-lo na banca examinadora, aceitando o convite de forma tão receptiva.

À direção e corpo gerencial da DATAPREV, pelo apoio que me foi prestado, garantindo-me a oportunidade de conclusão do curso.

À Agilent *Technologies* do Brasil, pelo apoio e cooperação no empréstimo do equipamento VQT, usado para as medições executadas neste trabalho.

## Dedicatória

Ao meu filho querido, ofereço este trabalho como exemplo de dedicação, disciplina e persistência, por tudo que tive que abnegar em prol de um objetivo a ser alcançado.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

RELAÇÃO ENTRE A QUALIDADE DAS RESPOSTAS DAS RECOMENDAÇÕES  
G.723.1 E G.729, E O COMPORTAMENTO DA REDE IP DE SUPORTE

Nelson Luiz Leal Fernandes

Março/2003

Orientador: Luís Felipe Magalhães de Moraes

Programa: Engenharia de Sistemas e Computação

Esta tese estuda como se comportam os codificadores empregados nas Recomendações G.729 e G.723.1 do ITU-T (*Telecommunication Standardization Sector of International Telecommunication Union*), de ampla utilização no mercado, segundo variações no comportamento da rede IP de transporte dos dados.

Sua proposta é definir uma relação entre a qualidade de voz obtida nas duas Recomendações já citadas e os valores de perdas de pacotes e *jitter*, baseada em critério objetivo e reconhecidamente adequado. Para tal, é usado o método de avaliação descrito na Recomendação P.862 do ITU-T.

Inicialmente, este trabalho promove uma visão geral da tecnologia de voz em redes de pacotes IP, discorrendo sobre as características, problemas e suas soluções. Em seguida, são descritos os métodos adotados para alcançar os resultados pretendidos, além de apresentar os valores das medidas executadas. Por fim, são listadas as conclusões advindas das medições de qualidade feitas em laboratório, apontando claramente em quais situações cada codec deve ser empregado.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for degree of Master of Science (M.Sc.)

RELATIONSHIP AMONG THE QUALITY OF ANSWERS OF THE  
RECOMMENDATIONS G.723.1 AND G.729, AND THE BEHAVIOR OF THE  
SUPPORTING IP NETWORK

Nelson Luiz Leal Fernandes

March/2003

Advisors: Luís Felipe Magalhães de Moraes

Department: Computing and Systems Engineering

This thesis studies the behavior of employed encoders in the Recommendation G.729 and G.723.1 of ITU-T (*Telecommunication Standardization Sector of International Telecommunication Union*), of widely used in the market, according to variations in the behavior of the IP network of transport of data.

Its proposal is to define a relationship among the voice quality obtained in the two Recommendations already mentioned and the values of losses of packages and jitter, based on objective criterion and recognized appropriate. As such, it was used the evaluation method described in the Recommendation P.862 of ITU-T.

Initially, this work promotes a general view of the voice technology in packages IP networks, talking about the characteristics, problems and their solutions. Soon afterwards, the methods adopted to reach the intended results are described, besides presenting the values of the executed measures. Finally, it was listed the conclusions of the quality measurements done at laboratory, appearing clearly in which one situations each codec should be used.

# ÍNDICE

<b>1. INTRODUÇÃO .....</b>	<b>1</b>
<b>2. OBJETIVO .....</b>	<b>2</b>
<b>3. ESCOPO .....</b>	<b>3</b>
<b>4. VISÃO GERAL DE VOZ SOBRE REDES COM PROTOCOLO IP .....</b>	<b>4</b>
4.1. MOTIVAÇÃO .....	4
4.2. CODIFICAÇÃO DO SINAL DE VOZ .....	5
4.3. IMPLEMENTAÇÃO DE QUALIDADE DE SERVIÇO.....	7
4.3.1. Serviços Diferenciados .....	7
4.3.2. Serviços Integrados.....	7
4.4. RECOMENDAÇÃO H.323 .....	10
4.5. SIP .....	15
4.6. COMPARAÇÃO ENTRE H.323 E SIP .....	18
4.7. RTP / RTCP.....	20
<b>5. CODIFICADORES RELEVANTES AO TRABALHO.....</b>	<b>22</b>
5.1. G.711 .....	22
5.2. G.728 .....	24
5.3. G.729 .....	25
5.4. G.723.1 .....	28
<b>6. FATORES IMPACTANTES NA QUALIDADE DA TELEFONIA IP.....</b>	<b>31</b>
6.1. BANDA .....	31
6.2. ATRASO.....	36
6.3. <i>JITTER</i> .....	43
6.4. PERDAS.....	48
6.5. ECO .....	58
6.6. RESPOSTA DE COMPRESSÃO.....	60
<b>7. MÉTODOS PARA QUALIFICAÇÃO DO SINAL DE VOZ.....</b>	<b>62</b>
7.1. P.800 (MOS).....	62
7.2. P.861 (PSQM) .....	71
7.3. PSQM+ .....	77
7.4. PAMS .....	78
7.5. PESQ (P.862) .....	84
<b>8. METODOLOGIA DE TRABALHO.....</b>	<b>94</b>
8.1. PROCEDIMENTOS DE ABORDAGEM.....	94
8.2. MEDIÇÕES EM REDES <i>EM PRODUÇÃO</i> .....	94
8.3. MEDIÇÕES EM LABORATÓRIO .....	96
8.3.1. Descrição do ambiente.....	98
8.3.2. Material fonte.....	100
8.3.3. Execução das medidas .....	100

<b>9. RESULTADOS OBTIDOS .....</b>	<b>102</b>
9.1. DISTRIBUIÇÃO DAS PERDAS.....	102
9.2. DISTRIBUIÇÃO DO <i>JITTER</i> .....	106
9.2.1. Medidas entre RJ e AC, na rede Frame Relay da Previdência Social .....	107
9.2.2. Medidas entre RJ e SP, na rede Frame Relay da Previdência Social .....	108
9.2.3. Medidas entre a DATAPREV (RJ) e a UFRJ, pela Internet.....	110
9.3. VALORES PESQ OBTIDOS.....	112
9.3.1. G.711 lei A.....	113
9.3.2. G.729 (10 ms) .....	114
9.3.3. G.729 (20 ms) .....	115
9.3.4. G.729 (30 ms) .....	116
9.3.5. G.723.1 (6,3 Kbps).....	117
9.3.6. G.723.1 (5,3 Kbps).....	118
9.3.7. Comparativo entre codecs.....	119
9.3.8. Comparativo entre RJ/AC e RJ/SP .....	123
9.3.9. Comparativo entre G.729 e G.729A .....	125
<b>10. CONCLUSÕES.....</b>	<b>126</b>
<b>11. SUGESTÕES PARA TRABALHOS FUTUROS .....</b>	<b>129</b>
<b>12. BIBLIOGRAFIA.....</b>	<b>131</b>
<b>ANEXOS.....</b>	<b>136</b>
ANEXO A – MEDIDAS DE CONVERSACÕES TELEFÔNICAS REAIS.....	136
ANEXO B – MEDIDAS DE PERDAS.....	139
ANEXO C – MEDIDAS DE ATRASOS E <i>JITTER</i> .....	140
ANEXO D – MODELO DE UMA FONTE DE PACOTES DE VOZ.....	154
ANEXO E – DISTRIBUIÇÕES EXPONENCIAL E GEOMÉTRICA .....	160



## 1. INTRODUÇÃO

Nas últimas décadas do século XX temos presenciado mudanças significativas no serviço telefônico. A década de 50 foi marcada pela possibilidade de chamadas internacionais diretas, através da introdução dos cabos transatlânticos; nos anos 60, as centrais e transmissões digitais melhoraram em muito a qualidade do sinal de áudio; nos anos 70, as centrais programáveis viabilizaram serviços como as chamadas em espera e discagem por tons; na década de 80, o sistema de sinalização em canal comum *out-of-band* (como o *Signaling System 7 – SS7*) possibilitou serviços como os números 0800. Por fim, a década de 90, marca definitivamente a trajetória da transmissão e sinalização telefônica, rumo a uma infra-estrutura baseada em redes de pacotes. Contudo, os primeiros artigos sobre o assunto datam do início de 70, com a primeira experiência de transmissão de pacotes IP com áudio, entre a *University of Southern California* e o *Massachusetts Institute of Technology*, em agosto de 1974. A primeira RFC (*Request for Comments*) sobre pacotes de voz, RFC 741, foi publicada em 1977 [1].

O uso de voz sobre IP (*Internet Protocol*) – VoIP – vem sendo uma das grandes metas de investimentos por fornecedores de soluções e usuários de telecomunicações nos últimos anos. Esta tecnologia abre um novo horizonte para as possíveis aplicações, integrando-se voz e dados num mesmo equipamento terminal de usuário, aproximando pessoas geograficamente distantes, aumentando a interatividade de aplicativos e até diminuindo os custos de comunicação quando comparada às ligações telefônicas interurbanas dos sistemas convencionais. Apesar do grande interesse gerado pelo assunto, ainda não é uma tecnologia com larga utilização, onde muitos técnicos atuantes na área de redes não a conhecem o suficiente para definição segura de projetos.

É certo que a transição da infra-estrutura telefônica atual para um ambiente puro de voz sobre redes de pacotes será longa, haja visto o enorme investimento feito para implantação do legado. Contudo, parece ser incontestável que este é um caminho sem volta.

## 2. OBJETIVO

Como qualquer outra aplicação que use a rede de comunicação de dados como suporte de transmissão, a voz sobre IP também tem sua qualidade interferida pela alteração do comportamento da rede. O objetivo deste trabalho é de identificar a relação entre a qualidade da voz obtida por codecs (codificadores/decodificadores) de mercado, e a situação de alguns dos parâmetros que caracterizam o ambiente de rede.

Assim, conhecendo o funcionamento da rede e estipulando-se o nível de qualidade desejada, poder-se-á identificar o codec mais adequado para uso, já que os equipamentos de codificação de voz comercializados, implementam várias opções de fácil seleção.

Com este estudo, também poder-se-á concluir que, para determinada característica de um segmento de rede, não existe opção de codificação que garanta o nível de qualidade desejado pelo projetista, identificando-se apenas a melhor resposta possível, segundo os parâmetros observados para tal segmento.

Resumindo, a idéia básica é que um projetista tenha acesso a informações precisas e objetivas das respostas de qualidade de alguns codecs, em função do comportamento da rede.

### 3. ESCOPO

Vários são os agentes que interferem na qualidade do sinal de voz digitalizada em uma transmissão sobre redes com o protocolo IP. Para situações onde temos a rede de pacotes apenas como parte do trajeto entre o orador e o ouvinte, fatores como ruído, desvanecimento e diafonia degradam a fala. Mesmo nos ambientes totalmente digitais, o atraso fim-a-fim, *jitter*, perdas, erros, eco e distorções provenientes das respostas intrínsecas aos codecs, são responsáveis pelo decréscimo de qualidade. De todos os fatores mencionados, as perdas de pacotes e o *jitter* interferem diferentemente em cada tipo de codec, resultando em sensações de qualidades distintas para a mesma variação do fator. Em outras palavras, a mesma variação na quantidade de perdas, por exemplo, é mais danosa para um tipo de codec do que para outro. Assim, de todos os fatores existentes e mesmo os apresentados neste trabalho, apenas as perdas e o *jitter* serão alvo de estudos e medições.

Da mesma forma, vários codificadores têm emprego no mercado. Contudo, os mais usados atualmente em telefonia IP são os citados nas Recomendações G.729 e G.723.1 do ITU-T, os quais foram escolhidos para análise.

Por fim, a infra-estrutura de rede de comunicação de dados para suporte à aplicação de voz, poderá ser baseada em comutação de circuitos ou de pacotes/células, desde que, obrigatoriamente faça uso do protocolo IP no nível 3 do Modelo Referência OSI (*Open System Interconnection*), com os quadros de voz sendo transportados pelos datagramas IP. Assim, a capilaridade e abrangência da solução é garantida, abordando o que é entendido, do ponto de vista do autor, como a proposta de maior vanguarda e longevidade para o assunto.

### 4.1. Motivação

Independente da tecnologia adotada para transmissão de voz, as corporações já vêm num movimento de integração entre voz e dados na mesma infra-estrutura de rede há alguns anos. As vantagens são claras, pois os custos envolvidos na manutenção de equipes técnicas, infra-estruturas diferenciadas e ligações internacionais são reduzidos com a integração. Os aumentos do leque de novas aplicações, da disseminação de microcomputadores pessoais (para funcionamento como terminal multimídia), da capilaridade das redes IP e da banda de transmissão disponível para o usuário, contribuíram em muito para voz sobre IP tornar-se uma realidade.

Com este novo ambiente descortinado, é praticamente impossível prever limites às potenciais novas aplicações que podem ser oferecidas aos usuários de redes. A disseminação da informática hoje reconta a história ocorrida com os aparelhos de rádio e depois os televisores, e não é difícil imaginar que, em alguns anos, praticamente todas as pessoas tenham em suas residências microcomputadores ligados à Internet.

Certamente será bastante comum duas pessoas no meio de uma conversa aproveitarem a ligação para trocarem arquivos contendo fotos digitalizadas e comentarem sobre o que estão vendo.

A computação colaborativa ganhará um novo potencial, quando a atualização das agendas pessoais ou marcação de reuniões puderem ser feitas tendo a voz dos envolvidos como uma chancela informal.

As novas secretárias eletrônicas poderão identificar o originador da chamada e responder com uma mensagem apropriada, previamente programada pelo dono. Além disso, as aplicações desenvolvidas para os computadores pessoais, com agradável apelo visual, tornará fácil a identificação das ligações recebidas na ausência do usuário chamado e apenas com um *clique* sobre o ícone identificador da chamada escolhida, completará a ligação com o usuário chamador.

Uma conferência audiográfica (envolvendo conversação e interação com aplicações acessadas pelos envolvidos, por exemplo: uma apresentação) será uma excelente ferramenta para aumento de produtividade corporativa.

Outra grande arma para apoio aos usuários será o telessuporte interativo, onde o especialista poderá, junto com o usuário, resolver problemas remotos, orientando o cliente enquanto está trabalhando.

Ao acessar a caixa postal de uma pessoa, além de ouvir uma mensagem direcionada para o chamador, este também poderá deixar seu recado e/ou enviar um arquivo para posterior análise do chamado.

Acessar canais digitais na Internet, onde poderá se pagar para ouvir uma música sob demanda, sem a necessidade de comprar o disco.

As áreas de vendas também poderão se beneficiar com esta tecnologia, tentando convencer o potencial comprador com uma agradável voz enquanto demonstra o produto.

Por fim (não que seja impossível imaginar outras incríveis e práticas aplicações, mas porque acredito que a potencialidade do tema já tenha sido bastante explorada), o treinamento a distância já é usado por algumas empresas com sucesso sendo uma forma ágil e relativamente barata de disseminar informação.

#### **4.2. Codificação do Sinal de Voz**

A digitalização de um sinal de voz, permite que seu armazenamento e transmissão seja feito de forma mais eficiente. A primeira codificação digital da fala data do final da década de 20 (por Homer Dudley), mas apenas na década de 70 teve uso fora da área militar [2].

O modelo genérico de um sistema digitalizador, pode ser representado por um conjunto de três estágios seqüenciais [3]:

Amostrador: Nesse estágio, um sinal contínuo no tempo é transformado em um sinal discreto.

Quantizador: Processo de mapeamento do sinal discreto, para um número contável, representando os vários níveis amostrados.

Codificador: Responsável pela representação binária do sinal digital, com o compromisso de manter a menor taxa de codificação possível e a melhor qualidade do sinal sintetizado.

Existem três tipos de codificadores de voz:

**Forma de onda:** São codificadores que tentam reproduzir o sinal original, amostra por amostra, com base nas suas características estatísticas, temporais ou espectrais. Geralmente são codificadores de baixa complexidade, que introduzem um pequeno retardo na voz e produzem um sinal de voz com alta qualidade, no caso de taxas de transmissão superiores a 16 kbps. Quando trabalhamos com taxas inferior a essa, a qualidade do sinal reconstituído degrada rapidamente.

O codificador de forma de onda mais simples é o PCM (*Pulse Code Modulation*), seguido pelo DPCM (*Differential Pulse Code Modulation*) que trabalha apenas com a diferença dos valores entre amostras consecutivas do sinal de voz, e o ADPCM (*Adaptive Differential Pulse Code Modulation*) que possui um quantizador adaptável ao valor de diferença observada entre amostras consecutivas.

**Paramétricos:** Os codificadores paramétricos operam usando um modelo para a forma de geração da voz, extraindo os parâmetros que representam este modelo. Esses codificadores são chamados de *vocoders*. A informação efetivamente transmitida são os parâmetros obtidos do modelo, que são atualizados periodicamente e determinados com a segmentação do sinal origem em intervalos periódicos denominados quadros, onde o sinal de voz pode ser considerado estacionário. Os *vocoders* operam com baixas taxas de transmissão, normalmente inferiores a 4 kbps, com atrasos e complexidade elevados e baixa qualidade da voz digitalizada, soando de forma sintética. A aplicação militar é o principal uso desses codificadores, onde a fidelidade da voz não é tão importante quanto a obtenção de uma baixa taxa de transmissão, a fim de permitir criptografia forte e pouca necessidade de banda.

**Híbridos:** Os codificadores híbridos, tentam preencher a espaço entre os codificadores de forma de onda e os paramétricos. São de maior complexidade, em comparação com os *vocoders*, e trabalham com taxas de transmissão entre 4 e 16 kbps, mas a qualidade de voz se aproxima da obtida com os codificadores de forma de onda. Estes codificadores são baseados nos modelos de produção de voz e utilizam uma excitação mais apurada para o filtro de síntese.

### 4.3. Implementação de Qualidade de Serviço

Para a aplicação em questão (voz sobre IP), quando se fala em garantia de Qualidade de Serviço, implicitamente estamos nos referindo a um conjunto de mecanismos específicos que deve ser implementado, junto ao protocolo IP, para garantir os quesitos necessários ao bom desempenho da aplicação.

De forma mais abrangente, esforços com a finalidade de proverem serviços com qualidade na Internet, têm sido o foco de extensivas pesquisas nos últimos anos. Existem duas abordagens diferentes no IETF (*Internet Engineering Task Force*) com esta finalidade, mas que não são mutuamente exclusivas: a de serviços integrados (IntServ) e a de serviços diferenciados (DiffServ).

#### 4.3.1. Serviços Diferenciados

Tem como premissa básica o foco no tráfego agregado com requisitos de serviços similares, e não no fluxo individual de pacotes [4].

Na prática, esta classificação é implementada no IPv4 através do campo TOS (*Type of Service*) do cabeçalho IP [5]. Os três primeiros bits, dos oito contidos no campo TOS, formam o subcampo *PRECEDENCE*, usado para identificar a importância do pacote. Os pacotes mais importantes recebem números de maior valor, na faixa entre 0 e 7.

Mais recentemente, a RFC 2474 [6] de dezembro de 1998 definiu o campo DS (*Differentiated Services*), para a atribuição de prioridades. Compatível com as versões 4 e 6 do IP, usa os octetos TOS e *Traffic Class* respectivamente.

#### 4.3.2. Serviços Integrados

Focada no fluxo individual de pacotes, entre os mesmos pontos de origem e destino. Nesta abordagem cada fluxo pode requisitar níveis de serviços diferentes à rede, sendo tipicamente quantificados como banda mínima necessária ou tolerância a um atraso máximo definido. Os quatro componentes básicos do IntServ (figura 4.1) são: unidade de controle de admissão (que identifica se a rede pode suprir o serviço requisitado), unidade de classificação (inspeciona os campos dos pacotes para determinar suas classes e o nível de serviço acordado), unidade de *schedule* (aplica um ou mais mecanismos de gerência de tráfego para garantir que o pacote seja transmitido à rede, satisfazendo a banda e atraso adequados ao tipo de fluxo) e o protocolo de reserva

de recursos RSVP (*Resource Reservation Protocol*) (que é responsável pela sinalização dos níveis de requisitos aos nós da rede) [4],[7].

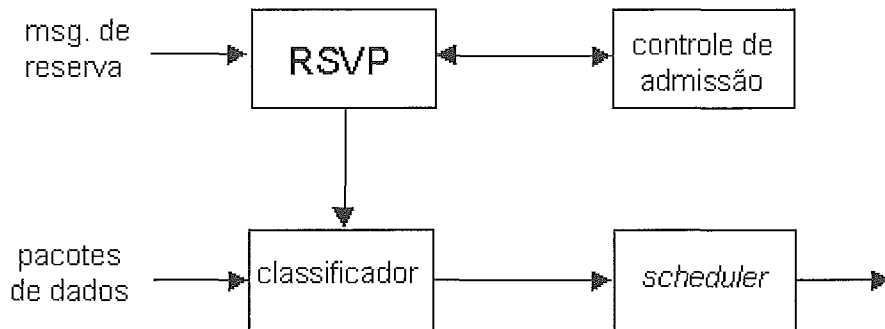


Figura 4.1 - Componentes básicos da arquitetura IntServ.

## RSVP

Definido na RFC 2205 [8], o protocolo RSVP não é um protocolo de roteamento, mas sim trabalha em conjunto com ele, desempenhando a função de sinalizar aos nós a necessidade de reserva de recursos de rede, permitindo que as aplicações possam usufruir da Qualidade de Serviço requerida para seus fluxos de dados, num ambiente de serviços integrados em redes IPv4 ou IPv6 [7],[9].

Suas funcionalidades principais são:

- Suporte a *unicast* e *multicast*;
- Reserva de recursos da rede apenas no sentido exigido;
- Transparência quanto ao tipo de dados;
- Atualização periódica quanto à capacidade e topologia da rede.

Caso um trecho intermediário ou sub-rede não suporte RSVP, é possível o uso de “tunelamento” entre os pontos que suportem o protocolo. Contudo, não é garantida a qualidade de serviço neste trecho, devendo ser dimensionada banda suficiente para o tráfego pretendido ou garantia do recurso por outro método.



Suporta três tipos de tráfegos diferentes:

- Melhor Esforço (*best-effort*) – tráfego IP tradicional, sem qualquer garantia de serviço, adequado para aplicações de transferência de arquivos ou correio eletrônico, por exemplo;
- Sensível a Velocidade (*rate-sensitive*) – são os que necessitam de garantia de banda mínima, como os gerados por tráfego de imagem;
- Sensível a Atraso (*delay-sensitive*) – são os que necessitam de garantia de atraso mínimo na rede, a exemplo de telefonia IP.

O RSVP trabalha com quatro tipos básicos de mensagens:

- *path message* – enviada pelo transmissor ao receptor, percorrendo todos os roteadores indicados como de melhor caminho pelo protocolo de roteamento. Esta mensagem contém informações sobre as características do caminho, a fim de subsidiar o receptor na solicitação de recursos;
- *reservation-request message* – enviada por cada receptor ao transmissor, percorrendo o caminho indicado pela *path message* e solicitando reserva dos recursos necessários;
- *error and confirmation message* – existem três tipos diferentes:
  - (a) *path-error message* – quando ocorre um erro no registro de caminho de uma *path message*, esta mensagem é retornada ao transmissor;
  - (b) *reservation-request error message* – informa ao receptor falha de admissão, banda indisponível, serviço não suportado ou caminho ambíguo, em retorno a uma *reservation-request message*;
  - (c) *reservation-request acknowledgment message* – enviada ao receptor como confirmação de recebimento da *reservation-request message*;
- *teardown message* – desfaz o caminho de reserva de recursos sem esperar por um *timeout*.

## Reserva de Recursos na Rede

Imaginemos duas máquinas, uma transmissora e outra receptora, onde determinada aplicação necessita reservar recursos da rede para funcionamento adequado. A transmissora envia à receptora uma *path message* contendo informações dos recursos solicitados pela aplicação, características do caminho roteado entre elas e o

caminho de volta à transmissora. Após manusear as informações contidas na *path message*, a receptora monta uma *reservation-request message* contendo efetivamente os recursos desejados para a aplicação em questão, segundo as possibilidades da rede. Esta mensagem é então enviada à transmissora, obedecendo o trajeto contrário do envio, e solicitando a reserva de recursos em cada roteador pelo caminho. Note que apesar da transmissora ter indicado os recursos necessários, quem solicita a reserva na prática é a receptora, pois é ela quem pode avaliar a capacidade da rede entre as duas máquinas.

### **Estado *soft* do RSVP**

Este estado permite que a rede RSVP suporte dinamicamente as alterações do número de nós envolvidos e de topologia da rede. O estado *soft* RSVP é criado e periodicamente atualizado pelas mensagens *path* e *reservation-request*. O término deste estado ocorre quando as mensagens citadas anteriormente não são trocadas dentro de um determinado período de tempo, ocorrendo *timeout*, ou por solicitação explícita de uma mensagem *teardown*.

### **4.4. Recomendação H.323**

O ITU-T (*Telecommunication Standardization Sector of International Telecommunication Union*) tem como proposta para a telefonia IP a Recomendação H.323 [10],[11], que descreve uma arquitetura contendo terminais, equipamentos e serviços para comunicação multimídia sobre LANs (*Local Area Networks*), sem garantia de qualidade de serviço. Na verdade a H.323 é uma recomendação “guarda-chuva”, referenciando outras recomendações ITU-T e até o protocolo RTP (*Real-Time Transport Protocol*) do IETF.

A primeira versão desta recomendação foi publicada em 1996 e a segunda em 1998, contendo melhorias no tocante à segurança, diminuição do tempo necessário para liberação do canal após a chamada ter sido atendida pelo destino, incorporação de serviços adicionais como transferência, redirecionamento, etc. e maior integração com T.120 (protocolo de transmissão de dados multimídia). A terceira versão foi publicada em setembro de 1999 sem introduzir alterações significativas no documento, mas com anexos tratando da reutilização de conexões, controle remoto de dispositivos, inclusão de um tipo simplificado de *endpoint*, possibilidade de estabelecimento de chamadas com UDP e outras características mais. A quarta versão foi aprovada em novembro de 2000, contendo várias alterações, dentre elas a possibilidade de multiplexação de áudio

e vídeo no mesmo fluxo, controle HTTP (*Hypertext Transfer Protocol*), uso de gatekeeper alternativo, gestão da capacidade de *endpoints*, URL (*Uniform Resource Locator*) designando usuários ou serviços e tunelamento de QSIG e ISUP [12].

## Escopo

Basicamente, destina-se a transmissão em ambiente com uma ou mais redes locais, sem garantia de qualidade de serviço.

Exemplos:

- Ethernet (IEEE 802.3)
- Fast Ethernet (IEEE 802.10)
- FDDI (no modo sem garantia de qualidade de serviço)
- Token Ring (IEEE 802.5)

## Componentes da Arquitetura

### (a) Terminal

Por definição é um ponto de rede que provê comunicação bidirecional, em tempo real, com outro equipamento terminal, *gateway* ou MCU (*Multipoint Control Unit*). Pode ser um equipamento telefônico com a funcionalidade de comunicação através do protocolo IP, um microcomputador com *software* específico e *hardware* multimídia para comunicação de voz ou qualquer outro tipo de equipamento que possa estar conectado em rede e permita, no mínimo, a comunicação de voz entre usuários.

Podemos ter transferência de:

- apenas voz;
- voz e dados;
- voz e vídeo;
- voz, dados e vídeo.

Na figura 4.2 estão ilustrados os elementos internos de um terminal H.323, seguida das descrições de cada um deles.

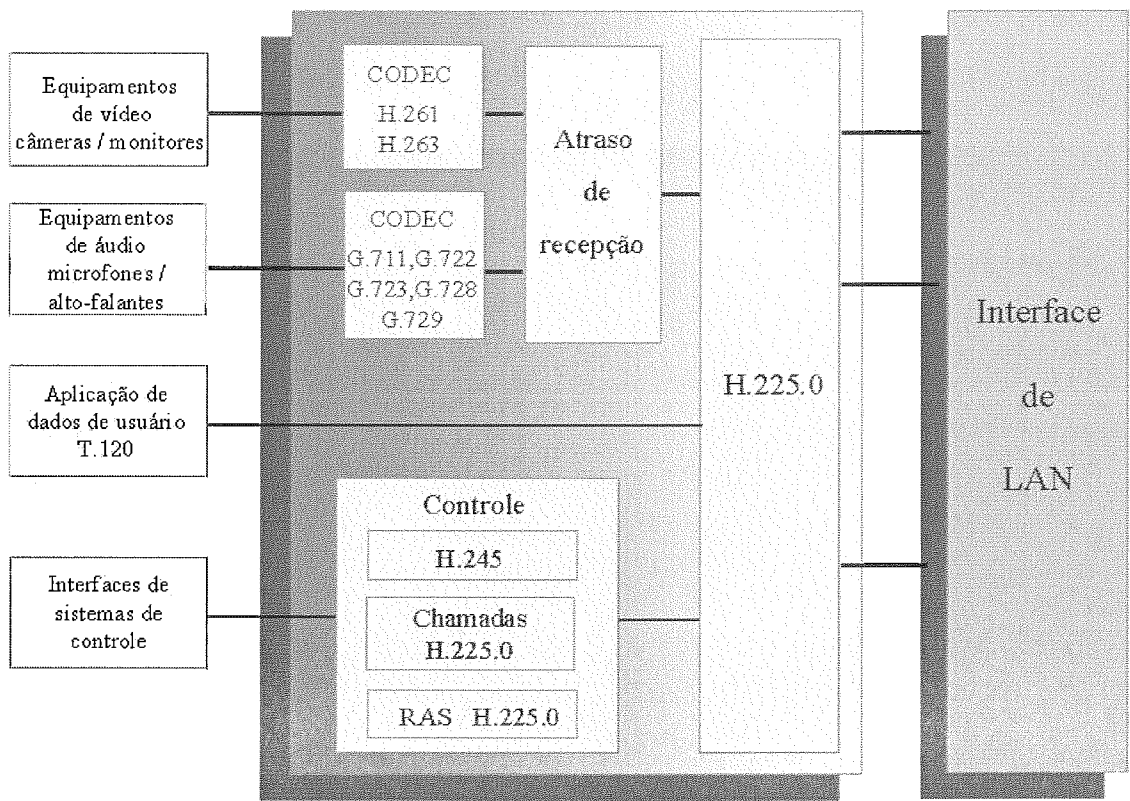


Figura 4.2 – Terminal H.323

- Codec de vídeo – suporta obrigatoriamente a Recomendação H.261 e opcionalmente a H.263, referentes à codificação de sinal de vídeo. A implementação deste elemento é opcional;
- Codec de áudio – suporta obrigatoriamente a Recomendação G.711 e opcionalmente as Recomendações G.722 (sinal de 7KHz codificado em 64 kbps), G.728, G.729, áudio MPEG1 e G.723.1. A implementação deste elemento é obrigatória;
- Atraso de recepção – responsável pelo controle do *jitter* nas recepções de áudio e/ou vídeo. A implementação deste elemento é obrigatória;
- Canal de dados – apesar de não estar explicitamente indicado na figura, suporta um ou mais canais de dados, uni ou bidirecionais. O T.120 é a base padrão para interoperabilidade entre terminais H.323 ou outro tipo suportado pelos ambientes externos ligados por *gateway*, no que se refere a aplicações de dados. A implementação deste elemento é opcional;
- Controle H.245 – responsável pelo manuseio de mensagens de controle fim-a-fim, que governam a operação da entidade H.323, no tocante à troca

do modo de operação de receptor para transmissor e vice-versa, abertura e fechamento de canais lógicos para troca de informações de um transmissor para um ou mais receptores, requisições de preferência de modos de operação (*unicast* ou *multicast*), mensagens de controle de fluxo, comandos gerais e indicações;

- Controle RAS (*Registration, Admission and Status*) – controle das mensagens de sinalização entre *endpoints* e *gatekeeper*, com as finalidades de: registro, admissão, troca de banda, pedido de informação sobre a situação atual e desvinculação, com o *gatekeeper*;
- Controle de Chamadas – usado para o estabelecimento de conexão entre dois terminais H.323. Utiliza a Recomendação Q.931 do ITU-T;
- H.225.0 – estabelece o formato das mensagens trocadas pelos canais lógicos de áudio, vídeo dados ou controle.

### **(b) Gateway**

Converte, apropriadamente, diferentes formatos de mensagens (exemplo: H.225.0/H.221 - estrutura de quadros em canais de 64 a 1920 kbps em telesserviços audiovisuais) e procedimentos de comunicação (exemplo: H.245/H.242 - estabelecimento de comunicação entre terminais audiovisuais em canais digitais até 2 Mbps). A conversão entre diferentes formatos de áudio, vídeo ou dados também pode ser feita pelo *gateway*.

### **(c) Gatekeeper**

É um elemento opcional no sistema H.323, provendo serviços de controle de chamadas para os terminais.

#### Funções obrigatórias:

Tradução de endereços – permite o uso de apelidos (*alias*) em lugar dos endereços de transporte;

Controle de admissão – autoriza o acesso de recursos usando por base critérios como: permissão de acesso em período predeterminado, banda requisitada, etc.;

Controle de banda – rege os pedidos de troca de banda em uso;

Gerência de zona – prove as funções acima para terminais, MCUs e *gateways* nele registrados.

#### Funções opcionais:

Controle de sinalização

de chamadas – pode controlar todo processo de sinalização de chamadas entre terminais, ou deixar que os próprios terminais tratem da sinalização;

Autorização de chamadas – através da sinalização prevista na H.225.0, o *gatekeeper* pode rejeitar chamadas de um terminal sem autorização adequada;

Gerência de Banda – controla o número de acessos simultâneos de terminais H.323 à LAN, podendo rejeitar uma chamada por falta de banda disponível;

Gerência de chamadas – controla quais terminais possuem chamadas estabelecidas, para o caso de uma nova chamada a um terminal ocupado ser informada aos módulos de direito.

Um conceito importante, introduzido pela segunda versão do H.323 é o de Zona. Por definição, Zona abrange a coleção de todos os terminais, *gateways*, e MCUs, gerenciados por um único *gatekeeper*. Cada Zona possui apenas um *gatekeeper* e não precisa estar confinada geograficamente a mesma LAN.

#### **(d) Multipoint Control Unit – MCU (Unidade de Controle de Multiponto)**

É um elemento para suporte a conferência com três ou mais pontos, podendo funcionar em três modos distintos: centralizado, descentralizado ou híbrido. No modo centralizado a comunicação entre a MCU e os terminais ou *gateway* é *unicast*. Dados, áudio, vídeo e controle passam obrigatoriamente pela MCU. Já no modo descentralizado, os terminais trocam informações de controle, e opcionalmente de dados, de forma centralizada com a MCU, mas trocam áudio e vídeo entre si por *multicast*. No modo híbrido, a comunicação de dados e controle sempre se dá de forma centralizada com a MCU. Contudo, podemos ter áudio também centralizado e *multicast* de vídeo ou vice-versa.

## Pilha de Protocolos

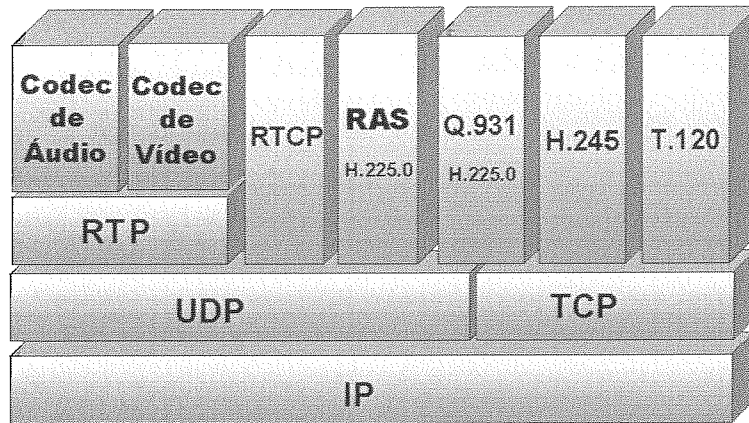


Figura 4.3 – Pilha de protocolos.

### 4.5. SIP

O IETF possui vários protocolos que compõem boa parte da solução para o complexo problema de integração de serviços na Internet [13]:

- SIP - *Session Initiation Protocol* - RFC 2543 (03/99) [14]  
Sinalização;
- SDP - *Session Description Protocol* - RFC 2327 (04/98)  
Definição de sessão multimídia;
- RTP/RTCP - RFC 1889 (01/96)  
Entrega fim-a-fim para dados interativos;
- RTSP - *Real Time Streaming Protocol* - RFC 2326 (04/98)  
Controle de servidores de mídia;
- MGCP - *Media Gateway Control Protocol* - RFC 2705 (10/99)  
Controle de *gateways*;

Neste item iremos abordar apenas o protocolo SIP, usado para sinalização de seções multimídia.

O SIP é um protocolo de controle referente à camada de aplicações do Modelo de Referência OSI, usado para iniciar, modificar ou terminar sessões ou chamadas multimídia entre usuários.

Suas funcionalidades são:

- Localização de usuários;
- Estabelecimento de chamadas;
- Suporte a *unicast* ou *multicast*;
- Administração na participação de chamadas (transferências, conferência, ...);
- Possibilidade de participação de um usuário em terminal H.323, via *gateway*.

O SIP é um protocolo cliente-servidor similar ao HTML (*Hypertext Markup Language*) no tocante à sintaxe e semântica das estruturas empregadas, com campos explicitamente descritos.

Durante seu funcionamento, requisições são geradas por um cliente e enviadas ao servidor (entidade receptora), que processa tal requisição e envia a resposta ao cliente.

O cliente SIP, chamado “agente usuário”, é formado por dois módulos: o agente usuário cliente, responsável em gerar requisições e o agente usuário servidor, responsável em responder as requisições. Os dois módulos são sempre obrigatórios para todo cliente.

### **Tipos de servidores SIP:**

- |                  |  |
|------------------|--|
| Registro         | - recebe atualizações da localização dos usuários na rede;   |
| <i>Proxy</i>     | - recebe requisições e as encaminha para outros servidores que contenham informação mais precisa sobre a localização do usuário chamado;                           |
| Redirecionamento | - também recebe requisições, mas no lugar de encaminhá-las para outros servidores, retorna o endereço do servidor para o qual a requisição deverá ser direcionada. |

### **Forma de endereçamento adotada:**

É usada uma SIP URL (*Uniform Resource Locators*), similar a um endereço de correio SMTP (*Simple Mail Transfer Protocol*), formada por uma parte referente ao nome do usuário ou número telefônico (*user*) e outra ao nome do domínio ou endereço numérico da rede (*host.domain*), isto é, *user@host.domain*.



## Comandos usados pelo protocolo:

- *REGISTER* ⇒ informa a localização de um usuário ao servidor de registro;
- *OPTIONS* ⇒ solicita informações de capacidade;
- *INVITE* ⇒ “convida” um usuário para uma chamada;
- *ACK* ⇒ confirma o recebimento de resposta a um *INVITE*;
- *BYE* ⇒ termina a conexão entre dois usuários;
- *CANCEL* ⇒ interrompe a busca por um usuário.

Como exemplo do formato das mensagens do protocolo, suponhamos que a usuária Maria queira saber os tipos de codificações suportados pelo usuário Lucas, antes de fazer uma chamada. Através de um *OPTIONS request*, a usuária solicita a informação desejada, utilizando o campo *Accept*. Esta mensagem é enviada do módulo cliente do agente usuário de Maria, para o módulo servidor do agente usuário de Lucas:

```
OPTIONS sip:lucas@example.com SIP/2.0
From: Maria <sip:maria@anywhere.org>
To: Lucas <sip:lucas@example.com>
Call-ID: 6378@host.anywhere.org
CSeq: 1 OPTIONS
Accept: application/sdp
```

Em resposta, o módulo cliente do agente usuário de Lucas envia um *OPTIONS response* indicando ser possível receber áudio nos padrões: PCM  $\mu$  law (*payload* tipo 0), 1016 (*payload* tipo 1), GSM (*payload* tipo 3), e SX7300/8000 (*payload* tipo 99); e vídeo H.261 (*payload* tipo 31) e H.263 (*payload* tipo 34):

```
SIP/2.0 200 OK
From: Maria <sip:maria@anywhere.org>
To: Lucas <sip:lucas@example.com> ;tag=376364382
Call-ID: 6378@host.anywhere.org
Content-Length: 81
Content-Type: application/sdp
v=0
m=audio 0 RTP/AVP 0 1 3 99
m=video 0 RTP/AVP 31 34
a=rtpmap:99 SX7300/8000
```

#### 4.6. Comparação entre H.323 e SIP

H.323 e SIP serão comparados, considerando-se os seguintes aspectos [15],[16]:

- ① Complexidade           ⇒ complexidade de implementação e/ou funcionamento;
- ② Expansão Funcional   ⇒ facilidade na inclusão de novas funcionalidades;
- ③ Escalabilidade         ⇒ facilidade para aumento da quantidade de elementos interligados;
- ④ Interoperabilidade   ⇒ possibilidade de cooperação entre diferentes versões, implementações e outros protocolos;
- ⑤ Serviços               ⇒ serviços oferecidos.

##### **Complexidade**

- Como o H.323 também inclui as recomendações H.225.0 para sinalização de chamadas, H.245 para controle, H.450 para serviços complementares e Q.931 para sinalização de chamada, o total de páginas do conjunto soma aproximadamente 1.000 páginas, em comparação com 195 páginas do SIP e SDP. Isto reflete maior complexidade para implementação do H.323, devido ao maior esforço do desenvolvedor para entendimento completa das especificações;
- O H.323 faz uso de centenas de elementos diferentes para comunicação. Já o SIP trabalha com apenas 37 tipos de cabeçalhos diferentes;
- As mensagens do H.323 possuem representação binária para cada campo, até sua versão 3. Na versão 4, o Anexo K descreve a possibilidade de controle por HTTP. O SIP utiliza texto para representação, facilitando em muito o rápido entendimento visual do protocolo;
- O H.323 usa vários protocolos sem uma separação clara, isto é, muitos serviços usam componentes dos vários protocolos. O SIP coloca todas as informações necessárias em uma mesma requisição.

## Expansão Funcional

- A estrutura textual dos campos no SIP, permite que novas características sejam incluídas de forma fácil e compatível com as versões anteriores. Os novos campos ou parâmetros podem ser colocados em qualquer parte da mensagem. Já no H.323, existem alguns locais predefinidos para inclusões futuras;
- O suporte a novos padrões de codificadores de áudio e vídeo é livre no SIP, basta que este novo codec seja registrado com uma identificação em órgão competente. No caso do H.323, os codecs devem ser padronizados pelo ITU, dificultando o processo de inclusões dos codecs de terceiros.

## Interoperabilidade

- A completa interoperabilidade com o H.323 exige a definição de perfis funcionais. O SIP trabalha de forma única para tratamento de uma mesma tarefa.
- O H.323 apresenta completa compatibilidade entre versões. No SIP, uma versão mais recente pode não mais suportar algumas características de antigas versões.

## Escalabilidade

- Os *gateways* ou servidores SIP podem trabalhar nos modos *stateful* ou *stateless*. No segundo caso, o servidor recebe uma requisição de chamada, executa algumas operações, encaminha a requisição e depois não se preocupa mais com este processo. As mensagens contêm informações de estado suficientes para a resposta ser encaminhada corretamente. Os *gateways* e *gatekeepers* H.323v2 são *stateful*, mantendo controle do estado da chamada durante toda a sua duração. Quando tratamos de um ambiente com uma grande quantidade de chamadas simultâneas, isto pode implicar em um sério problema de desempenho. A partir da versão 3, o H.323 também pode trabalhar no modo *stateless*.
- Conferências com H.323 obrigatoriamente necessitam da MCU para realização, que centraliza toda sinalização, por menor que seja o número de participantes. No caso de conferências com muitos participantes, este equipamento pode ser um “gargalo”. O SIP trabalha com o controle da conferência de forma distribuída pelos participantes, sem a necessidade de um equipamento centralizador.

## Serviços

- O conjunto de facilidades de telefonia oferecido pelos dois protocolos é bastante similar. Entende-se como facilidades a possibilidade de transferências, conferências, encaminhamento de chamadas, etc.
- A partir da versão 3 do H.323, é possível oferecer negociação de parâmetros de qualidade de serviço baseado em RSVP. O SIP não suporta tal funcionalidade, apenas recomenda o uso do RSVP.
- Até a versão 2, o H.323 usava apenas o TCP para estabelecimento de conexão. Isto implicava em mais tempo para o término do processo. Atualmente o SIP e H.323 consomem aproximadamente o mesmo tempo para conexão, pois o primeiro usa inicialmente o UDP, enquanto o segundo pode usar tanto o UDP como o TCP. A vantagem do H.323 sobre o SIP é que abre uma conexão UDP e outra TCP simultaneamente. Caso a conexão UDP falhe, a conexão TCP assume imediatamente. Se não houver falha na conexão UDP, existe um mecanismo eficiente para fechar a conexão TCP. Já o SIP espera ocorrer a falha de conexão UDP para iniciar a TCP.
- Em comparação com o SIP, a atual versão do H.323 provê maior garantia de funcionamento em caso de falhas de rede, através da possibilidade de registro em *gatekeepers* alternativos e *endpoints* com múltiplos endereços de transporte.
- A troca de informações de capacidade de um elemento SIP é feita pelo protocolo SDP, o qual não possui tanta flexibilidade como o H.245 usado no H.323. Um exemplo típico é que o SIP não suporta capacidade assimétrica, isto é, apenas recepção ou transmissão.

## 4.7. RTP / RTCP

Já que a transmissão de voz sobre IP usa o UDP como protocolo de transporte, os datagramas podem tomar caminhos diferentes na rede, resultando em diferentes tempos de propagação, sendo possível inclusive que um pacote transmitido posteriormente, chegue ao destino primeiro que seu antecessor. Esse é apenas um dos problemas que o RTP (*Real-Time Transport Protocol*) [17] se propõe a contornar.

## RTP

Este protocolo provê um serviço de entrega fim-a-fim, para dados com características de tempo real, tais como: áudio e vídeo interativos (RFC 1889). Suporta transferência de dados para múltiplos destinos, usando distribuição *multicast*.

### Serviços:

- Identificação do tipo de *payload*;
- Numeração de seqüência;
- Indicação do tempo da amostragem (*timestamping*);
- Monitorização de entrega.

### Não implementa:

- Entrega em prazo máximo;
- Garantia de qualidade de serviço;
- Garantia de entrega.

## **RTCP (RTP *Control Protocol*)**

É baseado na transmissão periódica de pacotes de controle por todos os participantes de uma sessão, a fim de monitorizar a qualidade de serviço e transportar informações destes participantes (RFC 1889).

### Funções básicas:

- Retorno sobre o congestionamento e fluxo dos dados;
- Transporte do “CNAME” de um elemento originador de chamada, para evitar conflitos de identificação em casos anômalos;
- Controle da taxa de transmissão dos pacotes, para o caso de muitos participantes.

## 5. CODIFICADORES RELEVANTES AO TRABALHO

Apesar deste trabalho ter escopo restrito aos codificadores mencionados nas Recomendações G.729 e G.723.1 do ITU-T, este capítulo discorrerá também sobre as Recomendações G.711 e G.728, com o intuito de complementar o assunto, dando maior base teórica ao documento. Desta forma, serão apresentadas informações gerais e as características das Recomendações G.711, G.728, G.729 e G.723.1, nesta ordem. Por fim, é apresentado um quadro resumo comparando as taxas de transmissão, atrasos e necessidades de recursos da G.723.1, G.729 e G.729 Anexo A, sendo esta última uma variação descrita no item 5.3.

### 5.1. G.711

#### Dados Gerais [18]

Data de Aprovação: 1972 (sem registro mais preciso)

Consumo de Banda: 64 kbps

Tipo de Codificação: PCM (*Pulse Code Modulation*)

Escopo: Não existe um escopo preciso definido pela norma, mas foi concebida como uma forma de digitalização do sinal de voz para ser tratado de maneira mais eficiente pelos sistemas digitais de comunicação.

#### Atraso

Como são realizadas 8.000 amostras do sinal analógico por segundo, o atraso de algoritmo para cada informação de voz digital (um octeto) é de 0,125 ms. Na prática, um pacote com voz no padrão G.711 contém mais de um octeto de informação.

#### Princípio de Codificação

Em um processo de amostragem, um sinal de voz contínuo no tempo é transformado em um sinal discreto no tempo. Para tanto, tomam-se amostras do sinal a intervalos periódicos. Para que o sinal original possa ser recuperado, a partir do sinal amostrado, é preciso que a frequência de amostragem obedeça ao critério de Nyquist, isto é, seja maior ou igual ao dobro da maior frequência formadora do sinal. Desta forma, para um sinal com máxima frequência de  $f_{\max}$ , a frequência de amostragem deve ser  $f_a = 2 \times f_{\max}$ .

Para aplicações em sistemas telefônicos em PCM, a frequência de amostragem adotada é de 8.000 amostras da amplitude do sinal por segundo. Após feitas as amostragens, estas são quantizadas em 256 níveis, sendo utilizado 8 bits para representação. Este tipo de codificação necessita de um canal com banda de 64 kbps para transmissão do sinal digitalizado, já que são geradas 8.000 amostras de 8 bits por segundo.

O procedimento de quantização nada mais é do que a atribuição de um valor discreto contável, aos níveis de amplitude do sinal contínuo amostrado. Esse processo gera um erro ou ruído de quantização que consiste na diferença entre o sinal na entrada do quantizador e o sinal discreto na saída. Para que esse erro seja percentualmente próximo, independente da amplitude, os níveis de quantização têm um espaçamento exponencial.

### Informações Complementares

Essa Recomendação, historicamente com ampla utilização no mercado, prevê o uso de duas curvas para tratamento dos erros de quantização, denominadas por Lei A (figura 5.1) e Lei  $\mu$ . A primeira foi adotada inicialmente na Europa e no Brasil, enquanto a segunda foi desenvolvida nos Estados Unidos e depois incorporada como padrão.

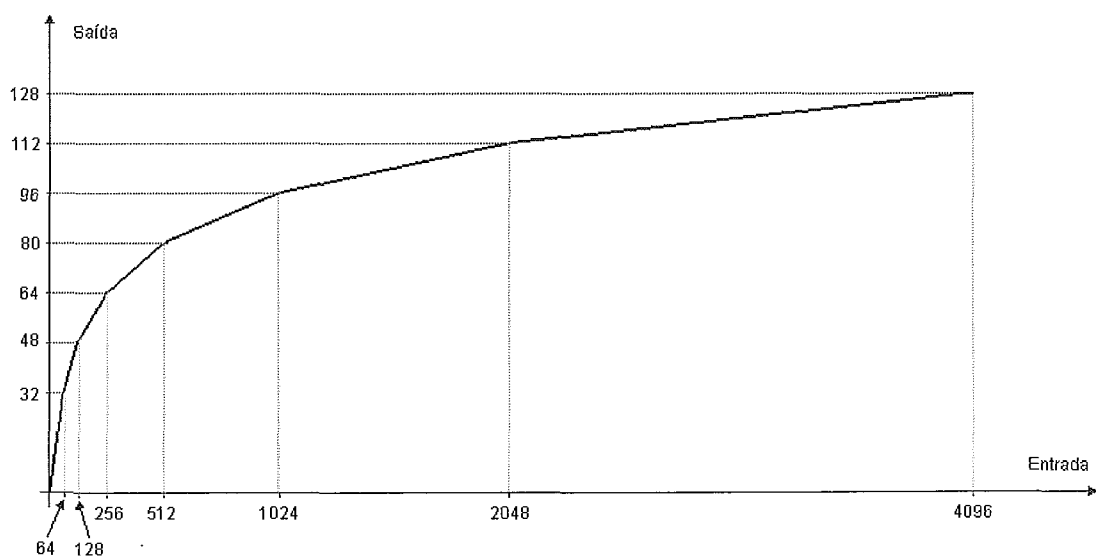


Figura 5.1 – Distribuição dos níveis de entrada e saída pela Lei A.

## 5.2. G.728

### Dados Gerais [19]

Data de Aprovação: 1º de setembro de 1992

Consumo de Banda: 16 kbps

Tipo de Codificação: LD-CELP (*Low-Delay Code Excited Linear Prediction*)

Escopo: Não foi encontrada referência na literatura.

### Atraso

Como o codificador trabalha com blocos de cinco amostras PCM e cada amostra tem um atraso de 0,125 ms, o atraso de algoritmo dessa Recomendação é de 0,625 ms.

### Princípio de Codificação

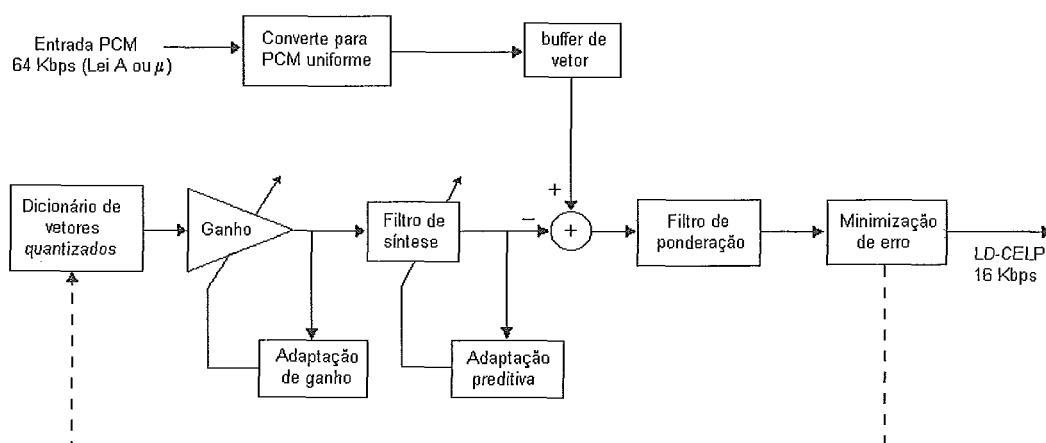


Figura 5.2 - Diagrama de blocos simplificado da codificação LD-CELP.

Após a conversão de PCM Lei A ou  $\mu$  para PCM uniforme, o sinal de entrada é agrupado em blocos de cinco amostragens consecutivas do sinal de entrada, no módulo *buffer* de vetor. Para cada bloco é feita uma comparação com todos os 1.024 vetores armazenados no dicionário de vetores quantizados, após esses terem passado pelas unidades de ganho e o filtro de síntese. A comparação que resultar na medida mais próxima apurada pelo módulo minimização de erro, indica o índice do dicionário que deverá ser transmitido pelo codificador. Esse índice é formado por 10 bits. Daí a banda



de 16 kbps consumida por essa técnica, já que das 8.000 amostras por segundo, a cada 5 temos uma seqüência de 10 bits a ser transmitida (figura 5.2).

Três parâmetros são atualizados periodicamente: o ganho e os coeficientes dos filtros de ponderação e síntese. Esses parâmetros são derivados do retorno de sinal do vetor precedente ao corrente. O ganho é atualizado a cada vetor, mas os coeficientes dos filtros são atualizados a cada 4 vetores, correspondendo a 20 amostras PCM ou 2,5 ms.

### Informações Complementares

Em 13 de novembro de 1995, foi publicado o Apêndice II dessa Recomendação onde é abordado o desempenho da fala em LD-CELP. Segundo ele, sob condições de transmissão sem erros, a qualidade da saída de um único codificador LD-CELP 16 kbps é inferior ao do PCM 64 kbps, mas equivalente ao ADPCM (*Adaptive Differential Pulse Code Modulation*) 32 kbps. Em meios com taxa de erro até  $1 \times 10^{-3}$ , o desempenho do LD-CELP é equivalente ao do ADPCM 32 kbps sob taxa de erro de  $1 \times 10^{-2}$ .

### **5.3. G.729**

#### Dados Gerais [20], [21]

Data de Aprovação: 19 de março de 1996

Consumo de Banda: 8 kbps

Tipo de Codificação: CS-ACELP (Conjugate-Structure Algebraic-Code-Excited Linear-Prediction)

Escopo: Sua concepção foi de codificar um sinal de voz com qualidade total em 8 kbps, para ser usado em primeiro lugar por aplicações com comunicação sem fio e em segundo lugar por redes com fio que necessitem compressão da banda usada pelo sinal codificado, tais como circuitos transoceânicos.

#### Atraso

Esta Recomendação codifica a fala e outros sinais de áudio em quadros de 10 ms cada. Adicionalmente, temos um tempo de *look-ahead* de 5 ms, resultando em 15 ms o tempo de atraso do algoritmo.

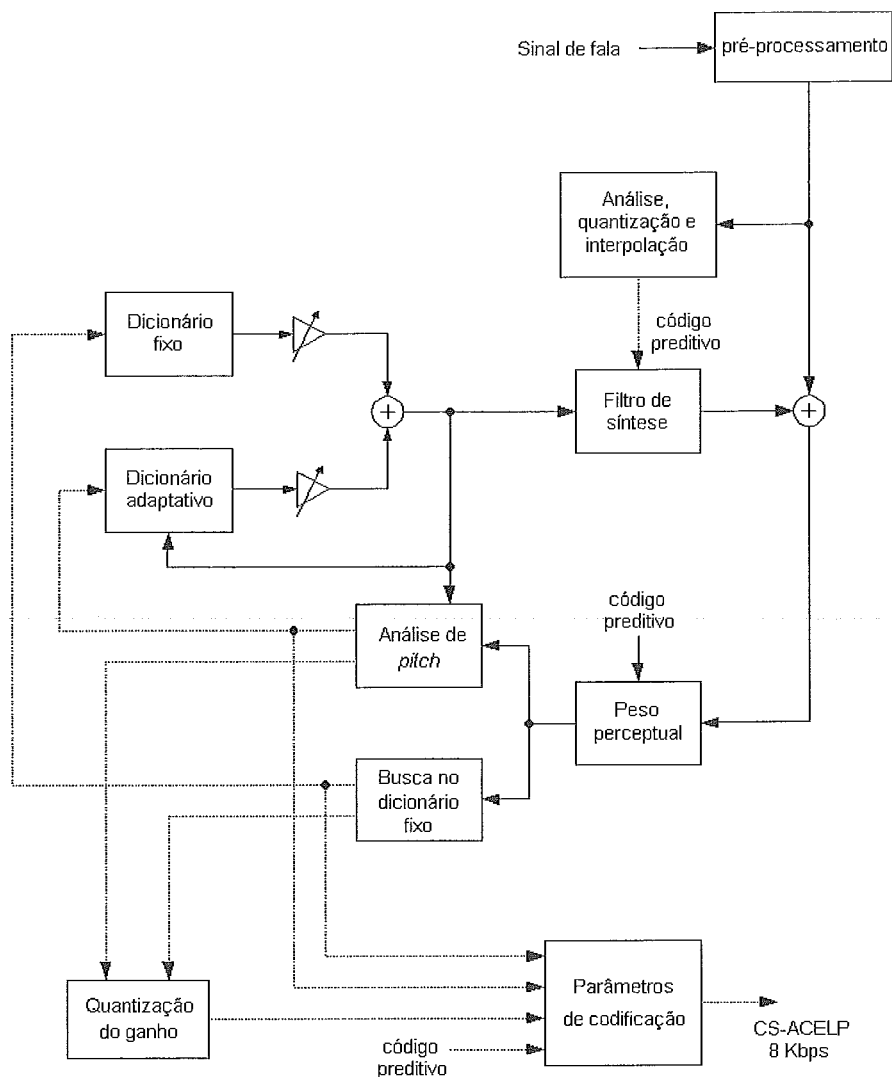


Figura 5.3 - Diagrama de blocos simplificado da codificação CS-ACELP.

Este codificador foi desenhado para operar com um sinal digital obtido de um sinal analógico resultante de filtragem, com largura de banda empregada em sistemas telefônicos, conforme Recomendação G.712. Após 8.000 amostragens por segundo, o sinal é convertido em um PCM linear de 16 bits, servindo de entrada para o codificador. A saída do decodificador é convertida para analógica de forma similar. Outra forma de entrada e saída, é a descrita pela Recomendação G.711 para dados PCM a 64 kbps, sendo convertida em PCM linear de 16 bits antes da codificação e de PCM linear de 16 bits para o formato apropriado depois da decodificação.

Este codificador é baseado no modelo de codificação CELP, isto é, codificação por excitação linear preditiva, base da Recomendação G.728. Opera com quadros de 10 ms, correspondentes a 80 amostras das 8.000 do PCM. Para cada um dos quadros, o sinal de voz é analisado para retirada dos parâmetros do modelo CELP (ganho, índices dos dicionários fixo e adaptativo e coeficientes de filtro). Esses parâmetros são codificados e transmitidos ao meio, num total de 80 bits por cada quadro amostrado (figura 5.3) [22], [23].

Os coeficientes utilizados por seus filtros são gerados através do método de autocorrelação com janelas de observação de 30 ms. A cada 80 amostragens (10 ms) os coeficientes são computados e é feito o *deslizamento* da janela. Dessa forma, a análise desses valores leva em conta as 120 amostras dos quadros passados, as 80 amostras do quadro atual e 40 amostras do próximo quadro (revelando aqui os 5 ms de *look-ahead* da codificação).

### Informações Complementares

#### ▪ Anexo A [24]

Apesar de contemporânea da Recomendação G.723.1, seus requisitos computacional e de memória eram superiores. Em maio de 1996, foi apresentado o Anexo A da Norma, reduzindo sua complexidade e mantendo a interoperabilidade com a G.729 original, conforme visto na tabela 5.1. O funcionamento básico do algoritmo na G.729 Anexo A é o mesmo da G.729. As principais simplificações feitas foram com relação à operação dos filtros e forma de busca nos dicionários de vetores.

#### ▪ Anexo B

Aprovado em outubro de 1996, o Anexo B descreve o detetor de voz ativa e gerador de ruído de conforto, ambos usados na compressão de silêncio, tanto na G.729 como na G.729 Anexo A.

## 5.4. G.723.1

### Dados Gerais [25]

Data de Aprovação: 19 de março de 1996

Consumo de Banda: 5,3 e 6,3 kbps

Tipos de Codificação:

ACELP (*Algebraic-Code-Excited Linear-Prediction*) para 5,3 kbps

MP-MLQ (*Multipulse Maximum Likelihood Quantization*) para 6,3 kbps

Escopo: Esta Recomendação especifica uma codificação usada para compressão de voz ou sinal de áudio de um serviço multimídia para meios de muito baixa velocidade de transmissão. No projeto deste codificador, a principal aplicação considerada foi de videofone, como parte dos padrões da família da Recomendação H.324.

### Atraso

Independente de qual das duas velocidades em uso, são necessários 30 ms para a formação de um quadro. Adicionalmente, temos um tempo de *look-ahead* de 7,5 ms, resultando em 37,5 ms o tempo de atraso do algoritmo.

### Princípio de Codificação

Na implementação desta Recomendação é obrigatório que as duas velocidades sejam disponíveis. É possível trocar de velocidade de transmissão, dentre as duas previstas, de um quadro para outro, durante uma conversação. Uma outra opção para alteração da velocidade de operação, é a utilização dos períodos de descontinuidade de transmissão nos intervalos sem fala para realizar a troca.

Este codificador foi desenhado para operar com um sinal digital obtido de um sinal analógico resultante de filtragem com largura de banda empregada em sistemas telefônicos, conforme Recomendação G.712. Após 8.000 amostragens por segundo, o sinal é convertido em um PCM linear de 16 bits, servindo de entrada para o codificador. A saída do decodificador é convertida para analógica de forma similar. Outra forma de entrada e saída, é a descrita pela Recomendação G.711 para dados PCM a 64 kbps, sendo convertida em PCM linear de 16 bits antes da codificação e de PCM linear de 16 bits para o formato apropriado depois da decodificação.

É baseado no princípio de análise do sinal para síntese do resultado, tentando minimizar o peso percentual do erro do sinal. Opera com blocos de 240 amostras cada, obtidos da observação em 30 ms, das 8.000 amostras por segundo. Cada bloco é inicialmente submetido a um filtro passa alta, para remoção da componente DC e em seguida é dividido em quatro subquadros com 60 amostras cada.

O processo de codificação é similar ao descrito na G.729. Nesta Recomendação uma janela *deslizante* de observação, com tamanho de 180 amostras é centrada em cada subquadro, gerando quatro conjuntos de parâmetros que serão agrupados, codificados e transmitidos. Estes parâmetros, como na Recomendação anterior, também são referentes a valores de ganho, índices de dicionários e coeficientes de filtros usados no modelo.

Quando a janela está centrada no último subquadro de cada uma das 240 amostras, deverá considerar também o primeiro subquadro do próximo quadro, para a obtenção do último grupo de parâmetros. Como cada amostra leva 0,125 ms e um subquadro tem 60 amostras, o atraso de *look-ahead* é de 7,5 ms.

No caso da codificação de maior banda (MP-MLQ), o total de parâmetros transmitidos a cada quadro de 240 amostras é composto por 189 bits. Já na codificação de menor banda (ACELP) o conjunto de parâmetros soma 158 bits. Assim,  $(189 \times 8000 / 240) = 6,3$  kbps e  $(158 \times 8000 / 240) = 5,3$  kbps, respectivamente.

### Informações Complementares

#### ▪ Histórico

Em 1992 foi anunciado o primeiro sistema de videofone do mercado, com velocidade de transmissão muito baixa. Em novembro de 1992 o Grupo de Estudos 15 (SG15 – Study Group 15) do ITU-T (CCITT – *Comité Consultatif International Téléphonique et Télégraphique*, na época) autorizou que um grupo de especialistas se reunissem a fim de definirem os requisitos para um padrão de codificação de fala para esta aplicação. Em setembro de 1993, o SG15 autorizou o início dos esforços para tal padrão. Em dezembro de 1993, 10 codificadores foram submetidos como candidatos. Já que este esforço supunha a escolha de um único codec, foi criado um programa de testes. Em março de 1994, o número de candidatos foi resumido a cinco. Nos testes feitos nos laboratórios da AT&T e France Telecom, dois codecs foram indicados como os de melhor qualidade, sem ser conclusivos entre os dois. Em outubro de 1994 foi acordada a criação do codec com dupla velocidade. A primeira versão do esboço da

Recomendação (*Draft Recommendation*) foi disponível em fevereiro de 1995, mas só foi encaminhado para decisão pelo SG15 em novembro de 1995, após algumas alterações.

▪ Anexo A

Este anexo descreve o esquema de compressão de silêncio com um único usuário, que permite a redução da banda passante, a fim de liberar uma parte maior do canal digital para outros usuários, sinais de vídeo codificados ou transmissão de dados. Aqui são descritos tanto o detetor de voz ativa (VAD – *Voice Activity Detector*) como o gerador de ruído de conforto.

Tabela 5.1 – Comparação dos atributos da G.723.1, G.729 e G.729 Anexo A.

	G.723.1	G.729	G.729A
Taxa (kbps)	5,3 e 6,3	8	8
quadro	30 ms	10 ms	10 ms
Atrasos <i>look-ahead</i>	7,5 ms	5 ms	5 ms
processamento	30 ms	10 ms	10 ms
Total	67,5 ms	25 ms	25 ms
MIPS	16 / 14,6	20	10,5
RAM ( <i>words</i> )	2.200	3.000	2.000

## 6. FATORES IMPACTANTES NA QUALIDADE DA TELEFONIA IP

Num ambiente de telefonia IP, a voz sofre a interferência de vários fatores inerentes aos sistemas convencional e digital. Dentre todos os possíveis, são destacados neste capítulo os de maior relevância e contribuição para a degradação da qualidade da fala.

### 6.1. Banda

A banda de comunicação de dados é um recurso oneroso e que não é infinito, dentro de uma rede em operação. Cada tipo de aplicação demanda uma quantidade específica de banda, para que apresente um bom desempenho. Assim também é com voz sobre IP, que dependendo do tipo de codificador usado, consumirá mais ou menos banda por cada canal de voz ativo, conforme foi comentado no capítulo 5.

Desta forma, para que se consiga trafegar voz no mesmo meio de comunicação, simultaneamente com outras aplicações de dados, faz-se necessário um mínimo de banda disponível. Tomando como exemplo um canal de voz que utilize a Recomendação G.729, onde cada pacote IP carregue um quadro de voz codificada, é consumida uma banda de 40 kbps só para a carga útil da voz e os cabeçalhos IP, UDP e RTP. É fácil perceber que:

- 1) Em casos como os antigos acessos discados com velocidades de 33,6 kbps era impossível o uso de VoIP sem a utilização de alguns recursos para diminuição do consumo de banda;
- 2) Mesmo com acessos de maior velocidade, continua representado um percentual significativo da banda total dos circuitos, levando à necessidade de emprego de alguns artifícios para diminuição desta banda por canal.

Os aumentos de banda dos meios de comunicação usados tanto por acessos residenciais, como por empresas, foi um dos fatores que viabilizou o uso de VoIP. Segundo [26][27], em 1997 54,68% dos acessos à Internet eram inferiores a 56 kbps com apenas 6,13% nesta velocidade. Já no ano seguinte, o primeiro percentual baixou para algo inferior a 40% e o segundo elevou-se para algo em torno de 30%, caracterizando o aumento de banda tipicamente dos acessos discados à rede. Atualmente, este público residencial conta com opções como *cable modem*, RDSI (Rede

Digital de Serviços Integrados), ADSL (*Asymmetric Digital Subscriber Line*) e outros mais, garantindo velocidades de acesso superiores a 128 kbps.

Apesar dos aumentos de disponibilidade de banda, a otimização do consumo de recursos continua sendo necessária. Neste enfoque, a supressão de transmissão dos pacotes de silêncio e a compressão dos cabeçalhos relativos aos protocolos IP, UDP e RTP, são duas técnicas muito úteis e praticadas pelo mercado.

### **Supressão de Silêncio**

Conforme [28], entre 40 e 50% do tempo de uma conversação telefônica, não existe voz trafegando em nenhum dos sentidos. Tomando proveito desta característica da aplicação de voz, não existe necessidade na transmissão de pacotes IP durante os períodos de silêncio. Atualmente, todos os produtos que implementam voz sobre IP possuem a possibilidade de ativação de uma funcionalidade conhecida por VAD (*Voice Activity Detector*) que é responsável em identificar os períodos de silêncio para suspensão da transmissão de pacotes.

Cada quadro de fala é inspecionado pelo VAD, que decide se deve ser usado o codificador de voz ativa ou o codificador de voz inativa. No primeiro caso um fluxo e bits resultante da voz codificada é transmitido. No segundo caso, pode-se ficar sem transmitir nada ao meio de comunicação ou enviar uma informação de atualização ao decodificador remoto, denominada SID (*Silence Insertion Descriptor*), conforme ilustrado na figura 6.1. Esta técnica é chamada de transmissão descontínua (DTX – *Discontinuous Transmission*).

Durante o período de fala ativa, o decodificador reconstrói o sinal original. Quando o VAD declara um quadro sem fala ativa, o sinal de saída é gerado pelo decodificador de voz inativa. A opção em deixar a saída completamente muda, durante o período de silêncio, cria ao ouvinte uma desconfortável sensação de súbita perda de informação, advinda do corte do nível de energia. Dessa forma, para preencher as fases de silêncio com algum sinal sonoro, uma descrição do ruído ambiente (SID) é transmitida do codificador de voz inativa para o decodificador de voz inativa. Isto é necessário porque as características do ruído de fundo podem variar muito durante o desenrolar de uma conversa, principalmente no caso do uso de telefonia móvel. Assim, um sinal sonoro próximo ao ruído de fundo percebido pelo codificador (ruído de



conforto) é emitido pelo decodificador de voz inativa, também chamado de gerador de ruído de conforto.

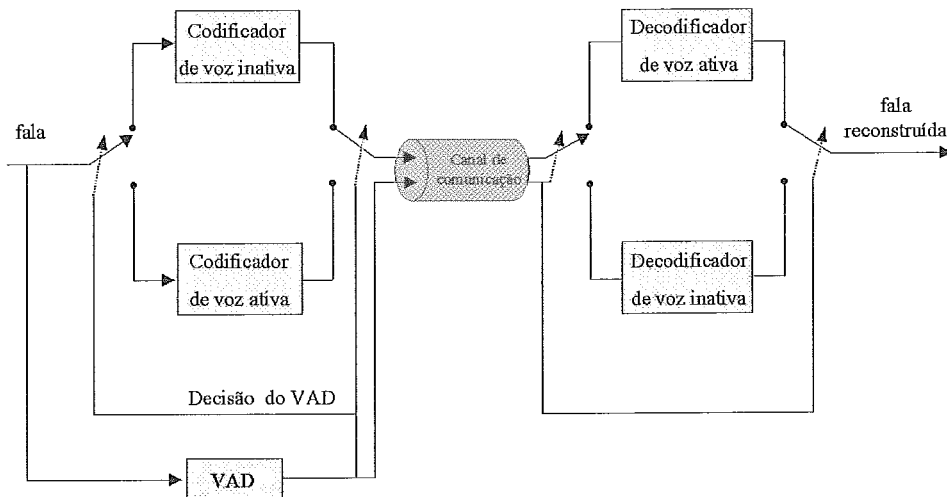


Figura 6.1 – Sistema de comunicação de voz com VAD.

O algoritmo de DTX determina, para cada quadro de voz inativa, a necessidade de transmitir um parâmetro de atualização do codificador ao decodificador. Durante a transição de atividade de fala para a inatividade, um quadro de SID é sempre transmitido, atualizando os parâmetros do gerador de ruído de conforto. Durante os quadros subsequentes, o DTX avalia a necessidade de novas transmissões de SID, avaliando a variação das características do ruído ambiente [29].

### Compressão de Cabeçalhos IP/UDP/RTP

Como já abordado no item 4.7, o RTP é o protocolo para transporte de mídia normalmente usado nas implementações de VoIP. Considerando os tamanhos de cabeçalhos dos protocolos IP, UDP e RTP, somamos 40 octetos ao todo, isso ainda sem a consideração do protocolo de enlace a ser obrigatoriamente utilizado. No caso de uma conversação com G.729, apenas 10 octetos são empregados para carga de voz codificada, o que é uma desproporção comparando-se com o tamanho dos cabeçalhos.

Para contornar este problema de desperdício de recursos, existem métodos para compressão dos cabeçalhos IP/UDP/RTP. Algumas implementações são proprietárias de cada fabricante, mas um método muito empregado é o descrito na RFC 2508 [30].

A idéia básica desta RFC consiste na observação de que o conteúdo de alguns campos não variam de pacote para pacote. Outros ainda variam, mas de forma predefinida e conhecida na mesma sessão. Assim, são trocadas informações entre os nós contendo os valores iniciais de todos os campos dos protocolos a serem comprimidos, bem como quais campos terão seus valores alterados durante a sessão e de que forma essa variação se dá. O contexto de cada sessão é armazenado numa tabela no nó de rede responsável pela compressão, indexada por uma função de *hash*, definida pela combinação dos endereços IP de origem e destino, portas UDP origem e destino e campo SSRC do RTP. O pacote comprimido carrega um campo de identificação do contexto de cada sessão denominado CID (*Session Context Identifier*) que pode ser formado por 8 ou 16 bits, dependendo da quantidade de sessões simultâneas que serão gerenciadas pelos nós de rede responsáveis pela compressão e descompressão dos cabeçalhos. Este tamanho é previamente combinado entre as partes envolvidas.

É previsto nesta RFC a transferência de 5 formatos diferentes de pacotes, dependendo da necessidade de informação entre os nós de rede e a abrangência de compressão usada. Para tanto, o protocolo de enlace deve ser capaz de indicar o transporte desses novos formatos, além dos pacotes IPv4 e IPv6 padrões.

Para que seja possível a identificação de perda de pacotes ou alteração de ordem, um campo de seqüenciamento com 4 bits é adicionado ao pacote comprimido. No caso de uso do campo de *checksum* do UDP com valor diferente de zero, esses 2 octetos também são inseridos no pacote, sem qualquer alteração, para serem utilizados pelos envolvidos na comunicação fim-a-fim. Em regime normal de conversação, no caso mais simples, onde não se tem a necessidade de informar variações nos campos dos pacotes descomprimidos e com CID igual a 1 octeto, os cabeçalhos dos pacotes comprimidos serão de 4 octetos quando do uso de *checksum*, e de 2 octetos caso o *checksum* do UDP esteja zerado (figura 6.2).

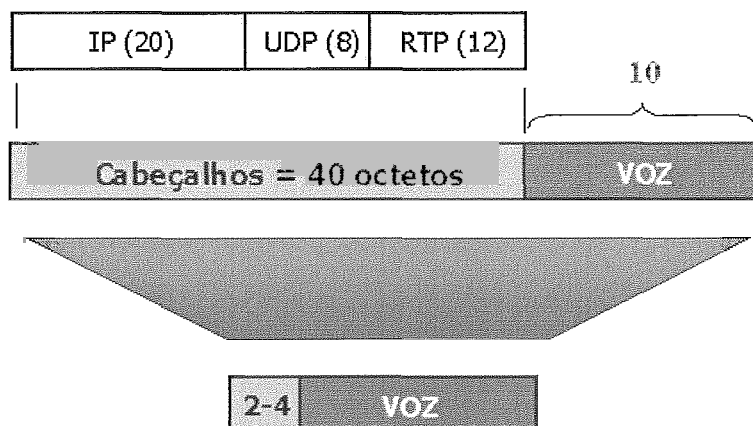


Figura 6.2 – Compressão de cabeçalhos.

Como o roteamento na rede IP se dá pela análise do endereço IP de destino, a compressão e descompressão dos 40 octetos inicialmente mencionados terá que ser feita enlace a enlace pelo par de nós envolvidos, até o destino final do pacote. É certo que este procedimento consome recurso de processamento e tempo para execução. Contudo, dependendo da velocidade e custo do meio de comunicação e da disponibilidade de processamento dos nós de rede, é mais vantajoso o seu uso do que o consumo de recurso de banda. Para enlaces *Frame Relay* operando a 64 kbps, um quadro com voz usando G.729 não comprimida, com tamanho de 54 octetos, consumiria 6,6 ms para serialização do quadro, enquanto que se fosse transmitido em um enlace de 1 Mbps esse tempo seria de 0,4 ms. Fica claro que quanto maior a velocidade do meio de comunicação, menor a parcela de tempo necessária a serialização dos octetos. Normalmente, a compressão de cabeçalhos é usada em enlaces inferiores a um E1 (2 Mbps), mas cada caso deve ser avaliado individualmente, pois o desempenho e ocupação de cada nó depende do seu tipo e do ambiente de produção.

Para ilustrar o ganho que se pode ter em um único canal de comunicação de voz, a tabela 6.1 mostra o comparativo do consumo de banda sem a compressão de cabeçalhos e seu uso com 2 octetos de cabeçalho.

Tabela 6.1 – Comparativo de consumo de banda com o RTP.

Recomendação	Banda útil (kbps)	Tempo entre pacotes (ms)	Tamanho da carga útil (octetos)	Banda sem compressão (kbps)	Banda com compressão (kbps)
G.729	8	10	10	40	9,6
G.729 <sup>1</sup>	8	20	20	24	8,8
G.723.1	6,4 <sup>2</sup>	30	24	17	6,9
G.723.1	5,3	30	20	16	5,9

Notas:

- 1 – Neste caso temos dois quadros de voz no mesmo pacotes IP;
- 2 – Segundo a Recomendação G723.1, a banda necessária é de 6,3 kbps, referentes aos 189 bits gerados pelo codec. Contudo, como a transmissão se dá efetivamente com octetos, isto é, múltiplos de 8 bits, quando se busca o menor múltiplo mais próximo (192 bits), tem-se um acréscimo na banda para 6,4 kbps.

## 6.2. Atraso

### Definição

O termo atraso, em comunicação de dados, normalmente vem associado ao tempo necessário para uma informação atravessar a rede. No nosso contexto de telefonia IP, este termo estará se referenciando ao atraso fim-a-fim observado pela aplicação. Desta forma, pode ser definido como o tempo gasto entre a geração do sinal de voz pelo orador e sua percepção pelo ouvinte.

### Efeitos

O atraso causa dois tipos de efeitos que contribuem para a degradação da qualidade de uma conversação telefônica. A partir de determinados valores de atraso, a reflexão do sinal emitido pelo orador é notada por ele. Este efeito é conhecido como eco e será minuciosamente tratado no item 6.5.

O outro efeito, que é de enfoque principal neste item, é responsável pela interferência na dinâmica da conversação, que será prejudicada por valores tão menores de atraso, quanto maior for o grau de interação entre os participantes.

Longos atrasos dificultam a comunicação entre oradores, onde um participante fala e só após algum tempo o outro consegue escutar, demorando para retornar com a resposta ao primeiro orador, interrompendo o fluxo normal da conversação. Se o atraso cresce muito, a pausa entre o final da fala e o início da resposta aumenta muito, podendo até inviabilizar a conversação.

O tempo necessário para que um pacote de áudio gerado chegue até seu destino não deve ultrapassar um patamar adequado, sob a pena de degradar a qualidade da aplicação. O patamar ideal depende fundamentalmente de três aspectos: o tipo de interatividade entre os usuários da aplicação, o nível de exigência dos usuários da aplicação e o quanto se está disposto em gastar para viabilizar uma solução que reflita pequenos atrasos.

Segundo [31], o atraso máximo em um sentido, para usuários mais críticos e exigentes não deve ultrapassar 100 ms, enquanto que para usuários mais tolerantes, este valor pode chegar até a 400 ms.

Dependendo da aplicação em questão, o grau de interação entre os usuários pode ser grande ou não. Considerando uma conversa entre duas ou mais pessoas, o tempo entre a geração do pacote de voz e a entrega no destino deve ser muito menor que o de uma aplicação de *stream* de áudio, onde a informação é transmitida em um único sentido, podendo ter valores consideravelmente relaxados.

A Recomendação G.114 do ITU-T, de maio de 2000, discorre sobre especificações do tempo de transmissão, incluindo o processamento em equipamentos e o tempo de propagação de rede [32].

Os limites para os tempos de transmissão fim-a-fim indicados na G.114 são divididos em três classes:

- De 0 até 150 ms – Aceitável para a maioria das aplicações;
- De 150 até 400 ms – Deve-se estar atento ao impacto do tempo de transmissão na qualidade da aplicação. A comunicação entre dois pontos que dependa de um enlace satélite se encontra contida nesta faixa.

- Acima de 400 ms – De forma geral é inaceitável para aplicações em rede. Entretanto, casos excepcionais com a necessidade de dois enlaces satélites podem estar nesta faixa.

Nesta mesma Recomendação, é mostrado um teste subjetivo para avaliação do efeito do atraso em conversações telefônicas feito em 1991, envolvendo usuários voluntários desse serviço. Os resultados estão registrados na figura 6.3.

Também são comentados testes que demonstram a diferença de percepção do efeito do atraso na qualidade, segundo o tipo de público usado para os testes subjetivos e segundo o tipo de teste feito. Foi observado que o grau de exigência e percepção de variações no atraso de pessoas treinadas, isto é, envolvidas em tarefas afins a telecomunicações, é maior do que o de outras pessoas, por exemplo donas de casa ou estudantes. Em testes dirigidos (por exemplo: a leitura de números aleatórios, o mais rápido possível, alternando a fala entre as duas pessoas envolvidas no teste ou a tentativa de identificação do nome de uma cidade, o mais rápido possível, também com a alternância entre orador e ouvinte), os envolvidos identificam perda de qualidade com atrasos menores do que em teste de conversação livre.

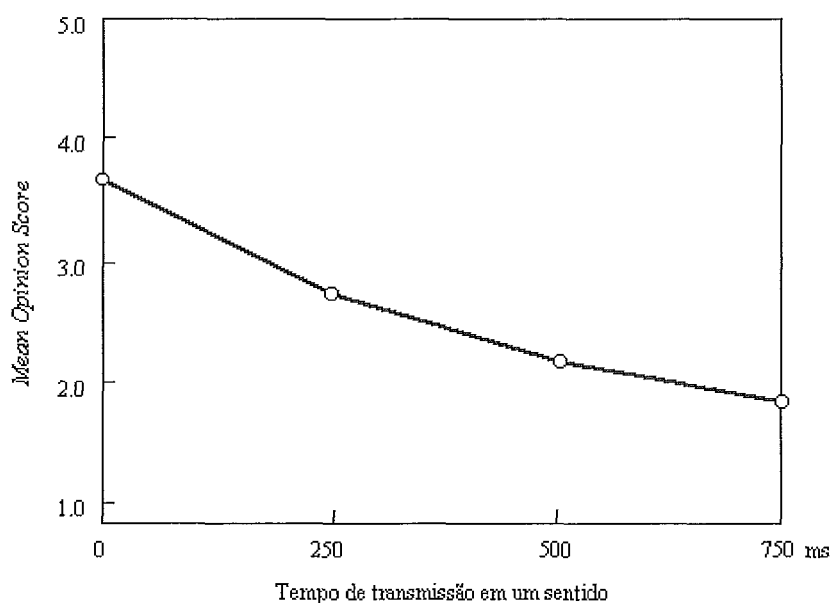


Figura 6.3 – Variação do MOS segundo o tempo de atraso.

No Anexo B da G.114 foi definida uma medida de “limiar de detecção”, que representa o valor máximo de *round-trip delay*, expresso em milissegundos, no qual 50% dos envolvidos nos testes classificam o serviço de voz como aceitável. Para o teste de leitura de números, comentado na parágrafo anterior, o limiar de detecção foi de 90 ms, enquanto que o de conversação livre chegou a 740 ms.

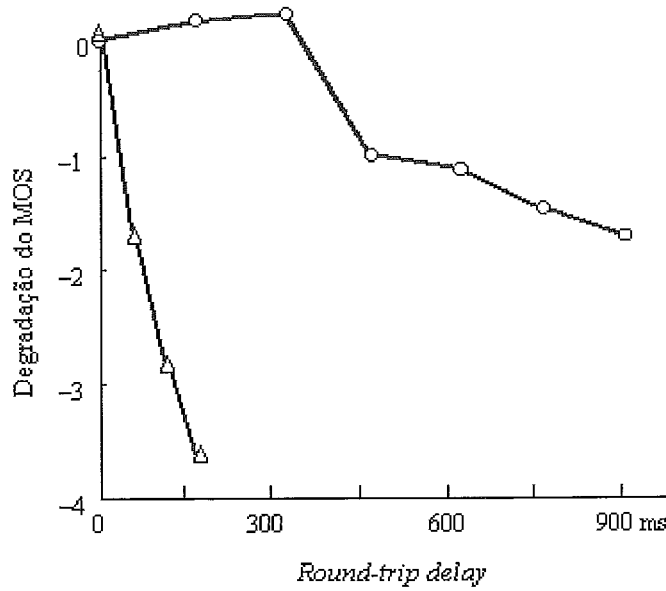


Figura 6.4 – Degradação do MOS segundo o *round-trip delay*, onde o círculo representa a conversação telefônica e o triângulo a leitura de números.

Em outros testes, também relatados no mesmo Anexo B, obteve-se a degradação do MOS em um teste de leitura rápida de números com alternância de oradores e em uma conversação telefônica, segundo a variação do *round-trip delay*, conforme mostrado na figura 6.4.

Como conclusão dos testes, a Recomendação G.114, sugere que:

- Os efeitos do atraso em conversações dinâmicas podem ser claramente identificados para valores inferiores a 400 ms em um único sentido e
- Os efeitos do atraso na qualidade da fala aumentam de forma moderada, conforme aumenta o valor do atraso.

## Fatores causadores

Compondo todo o sistema de um ambiente de comunicação de voz sobre IP, temos várias parcelas que contribuem no atraso total fim-a-fim. Algumas dessas parcelas são elementos fixos e de fácil determinação, sendo elas: o atraso gerado pela formação de um quadro de voz pelo codec em uso; tempo de serialização do pacote IP, dada a velocidade do circuito contratado e atraso gerado pelo *buffer* de *jitter*. Outras parcelas são variáveis ou de determinação menos precisa, sendo elas: o processamento de codificação e decodificação da voz e o tempo de propagação dos pacotes através da rede IP.

Estes elementos de atraso podem ser agrupados em três grupos, conforme a figura 6.5.

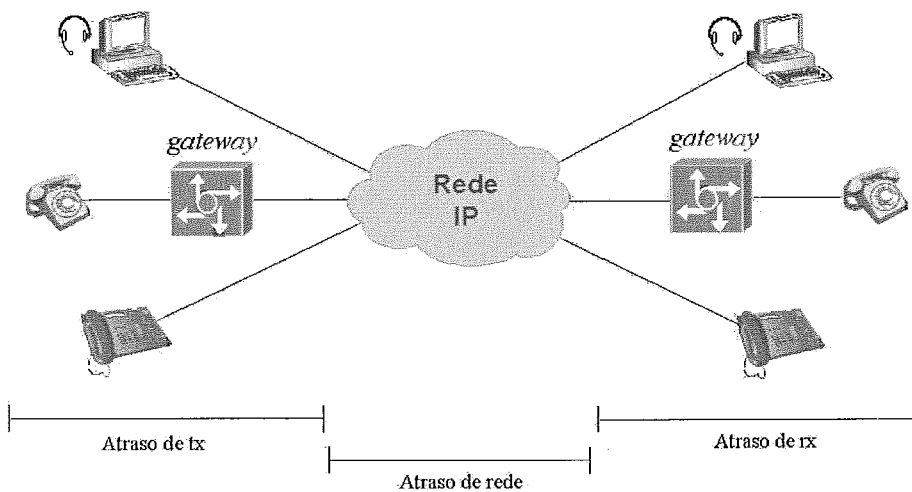


Figura 6.5 – Atrasos em um ambiente de voz sobre redes IP.

- Atraso de transmissão (tx): Formação dos quadros de voz e processamento para codificação dos quadros de voz.
- Atraso de rede: Serialização dos pacotes IP e propagação na rede (formada pela propagação através dos meios de comunicação que formam a rede e pelos tempos de enfileiramentos nos nós de rede).
- Atraso de recepção (rx): Atraso gerado pelo *buffer* de *jitter* e tempo de processamento para decodificação dos quadros de voz.



Os tempos necessários para a formação de cada quadro de voz e de processamento para codificação, dependem intrinsecamente do tipo de codec escolhido. No caso das duas Recomendações de interesse para este trabalho, G.729 e G.723.1, conforme itens 5.3 e 5.4 respectivamente, esses tempos são consideravelmente diferentes e de magnitudes significativas na contabilização do atraso total fim-a-fim no sistema. Considerando os tempos apresentados na tabela 1 e como exemplo um atraso total pretendido de 200 ms, os tempos de formação de quadro e processamento para codificação e decodificação, são responsáveis por uma parcela de 13% a 34% do global, dependendo do codec escolhido.

A serialização dos pacotes IP também é um fator a ser considerado como formador do atraso total. Dependendo da velocidade do canal de comunicação utilizado e do tamanho do pacote, este tempo poderá trazer sérios danos à qualidade do sistema. Por exemplo: podemos pensar em pacote de 1024 octetos sendo transmitido por um canal de 64 kbps. Neste caso, seriam necessários 128 ms apenas para a serialização do pacote, representando 64% dos 200 ms pretendidos para atraso total, tornando impraticável a obtenção deste tempo total.

A propagação de um pacote através de uma rede de longa distância, pode demandar desde poucas dezenas até algumas centenas de milissegundos. Excepcionalidades a parte, percursos formados exclusivamente por meios terrestres (fibra óptica, cabo de pares, enlace de microondas, etc.) consomem muito menos tempo de propagação do que duplos saltos em sistemas satélites geoestacionários.

Por cada nó de rede em que o datagrama de voz passe, este é armazenado em uma fila de transmissão para aguardar a sua vez de ser enviado até o próximo nó, pelo meio de comunicação. Dependendo do algoritmo para tratamento de prioridade e dos outros tipos de fluxo que compartilham a mesma fila, o pacote de voz pode demorar mais ou menos tempo até ser atendido, isto é, chegar à cabeça da fila e ser transmitido. Atualmente no Brasil, a grande maioria de redes públicas não tratam fluxos prioritários de dados, ficando esta tarefa restrita aos nós internos das redes das empresas que se interessam em fazer esta diferenciação.

O enfileiramento de pacotes para amenizar o efeito do *jitter* é abordado com detalhes no item 6.3. Este tratamento consome em média meia centena de milissegundos.

A latência dos equipamentos envolvidos no sistema de transmissão de voz sobre redes de pacotes (roteadores, comutadores, *gateways*, terminais telefônicos) é um dos componentes do atraso total, mas que não contribui significativamente para tal. Cada vez mais comuns, os programas de computador que emulam terminais telefônicos ou multimídia, oferecem funcionalidades similares aos dos equipamentos especificamente projetados para tal. Como usam uma plataforma de *hardware* genérica, obviamente não apresentam o mesmo desempenho, já que muitas funções são desenvolvidas por programas em lugar dos processadores de sinais especializados, resultando em maior latência.

### **Medição do atraso de rede**

Sendo o fator isolado de grande responsabilidade pelo atraso fim-a-fim total, o atraso de rede em cada sentido da conversação é de difícil medição. Para uma medição correta e acurada seria necessário o sincronismo, em milissegundos, dos relógios dos dois equipamentos usados na medição, estando um em cada extremidade da rede observada. Na prática, isto não é possível quando tratamos de pontos geograficamente distantes, que é a situação mais comum.

Para contornar este problema prático e conseguir-se uma medida aproximada da real, a única forma de medição é usando-se o ICMP (*Internet Control Message Protocol*) para obtenção do atraso nos dois sentidos somados (*round-trip delay*), e a partir daí, dividir-se o valor por dois para ter-se o valor médio em cada um dos sentidos. Assim, o relógio base para a medida é sempre o mesmo e com garantia da acurácia necessária. Apesar de não ser um método ideal, é o usado por produtos de mercado e estudos publicados [33], [34].

O comportamento distinto nos atrasos observados para cada sentido de uma conversação é conhecido [35] e muito comum nas redes de longa distância, principalmente em grandes redes IP. Normalmente os fatores que influenciam no atraso fim-a-fim, agem de forma diferente e independente em cada um dos sentidos da comunicação.

## Classificação

De forma mais sistemática, pode-se classificar o atraso fim-a-fim em três grupos:

- atraso de algoritmo – gerado pelo módulo de codificação do sinal, é causado pelo tempo necessário para a formação de um quadro, mais o tempo de *look-ahead*;
- atraso total de codec – refere-se à soma do atraso de algoritmo ao tempo de processamento necessário para codificar o sinal na origem e decodificar no destino (ambos somados). O tempo para processamento depende do potencial do equipamento usado, mas é próximo ao tempo de codificação;
- atraso total de sistema – refere-se à soma do atraso total de codec aos tempos necessários à serialização e propagação do sinal digital, da origem ao destino, e ao tempo despendido para enfileiramento dos quadros (*buffer* para amenizar o efeito do *jitter*).

### 6.3. Jitter

O termo jitter, no nosso escopo de transferência de pacotes de voz digitalizada, pode ser encontrado na literatura em citações com duas nuanças: uma relacionada ao atraso fim-a-fim na rede e outra aos intervalos de tempo entre chegadas de pacotes no receptor. O resultado prático é o mesmo, contudo, a definição mais rigorosa depende da forma abordada.

Como exemplos podem ser apontados alguns textos:

a. Relacionado com o atraso fim-a-fim na rede

a.1. “Queue Monitoring. A Delay Jitter Management Policy” – NOSSDAV 1993 [36]

*“Variance in end-to-end delay is called delay jitter.”*

a.2. “Real-Time Voice Over Packet-Switched Networks” – *IEEE Network*, 1998 [31]

*“If network condition are poor, average packet delay and packet delay variance (jitter) will be high (on the order of 75-300 ms).”*

b. Relacionado com os intervalos de tempo entre chegadas

b.1. "RTP: A Transport Protocol for Real-Time Applications" – RFC 1889 [17]

*"interarrival jitter: 32 bits*

*An estimate of the statistical variance of the RTP data packet interarrival time, measured in timestamp units and expressed as an unsigned integer. The interarrival jitter J is defined to be the mean deviation (smoothed absolute value) of the difference D in packet spacing at the receiver compared to the sender for a pair of packets. As shown in the equation below, this is equivalent to the difference in the "relative transit time" for the two packets; the relative transit time is the difference between a packet's RTP timestamp and the receiver's clock at the time of arrival, measured in the same units.*

*If  $S_i$  is the RTP timestamp from packet  $i$ , and  $R_i$  is the time of arrival in RTP timestamp units for packet  $i$ , then for two packets  $i$  and  $j$ ,  $D$  may be expressed as*

$$D(i,j)=(R_j-R_i)-(S_j-S_i)=(R_j-S_j)-(R_i-S_i)$$

*The interarrival jitter is calculated continuously as each data packet  $i$  is received from source  $SSRC\_n$ , using this difference  $D$  for that packet and the previous packet  $i-1$  in order of arrival (not necessarily in sequence), according to the formula*

$$J=J+(|D(i-1,i)|-J)/16$$

*Whenever a reception report is issued, the current value of  $J$  is sampled.*

*The jitter calculation is prescribed here to allow profile-independent monitors to make valid interpretations of reports coming from different implementations. This algorithm is the optimal first-order estimator and the gain parameter  $1/16$  gives a good noise reduction ratio while maintaining a reasonable rate of convergence [11]. A sample implementation is shown in Appendix A.8."*

b.2. "Internet Telephony: Services, Technical Challenges, and Products." - IEEE Communications Magazine, 2000 [37]

*"Variance in interframe arrival times at the receiver is called jitter and is potentially more disruptive for IP telephony than delays discussed above."*

b.3. “*Diagnosing CiscoCallManager Problems*” – Cisco Systems Inc. [38]

“*Jitter is the variation in the arrival times of packets.*”

b.4. “*Fine-tuning Voice over Packet services*” – RADCOM Ltd. [33]

“*Jitter is calculated based on the inter-arrival time of successive packets.*”

b.5. “*Measuring Internet Telephony Quality: Where Are We Today ?*” – Globecom’99 [39]

“*Interarrival times can be obtained by the difference in the local arrival time and a timestamp in the packet itself with the previous packet, this is continually calculated and a mean variation is obtained by multiplying the mean variation by a smoothing constant. This mean variance or jitter gives a good indication of the running condition of a connection.*”

Como o ponto focal deste trabalho é a qualidade observada pelo prisma da aplicação usuária dos recursos de voz, inicialmente poder-se-ia aceitar como melhor definição aquela que referencia a variação dos intervalos de tempo entre chegadas. Entretanto, o diagrama “f” na figura 6.6 mostra uma situação onde o *jitter* existe, mas o intervalo de tempo entre chegadas de pacotes consecutivos não varia. Desta forma, será adotada como definição de *jitter*:

**A existência de variação entre os intervalos de tempo necessários para que pacotes consecutivamente transmitidos, percorram a distância entre os pontos origem e destino de um sistema de comunicação.**

Contudo, o método para se quantificar o *jitter*, não consegue levar em conta apenas a medição do tempo de rede fim-a-fim, de forma absoluta, já que não se tem como garantir o sincronismo em milissegundos, dos relógios dos equipamentos remotos envolvidos. Desta forma, a quantificação do *jitter* terá que considerar o tempo relativo na comparação entre dois pacotes.

Um método muito interessante é o utilizado na RFC 1889, conforme subitem b1. Apesar de não explicitar a diferença entre as situações representadas nos diagramas “d” e “f”, da figura 6.6, dá uma boa idéia sobre o comportamento da rede ou sistema observado. Os diagramas “b” e “c” são bons exemplos disso, já que no caso “c” apesar

da diferença de tempo do segundo pacote ter sido apenas de 2 milissegundos, em relação ao mesmo pacote do caso “b”, ocorreu inversão de ordem entre pacotes consecutivos. No caso do sistema observado não possuir a capacidade de detecção e correção de ordem, dois pacotes serão “tocados” com suas posições trocadas, ocasionando uma percepção de qualidade pior do que no caso “b”. Outra vantagem é a identificação se um problema isolado ocorreu há muito tempo atrás ou recentemente. Caso fosse optado por uma outra forma de medida, que considerasse a média e desvio padrão de um conjunto de intervalos de tempo entre chegadas de pacotes consecutivos, a percepção de quando o problema ocorreu não seria possível.

Desta forma, a grandeza de quantificação do *jitter* será obtida da seguinte forma:

$$J_i = J_{i-1} + ( |D_{i-1,i}| - J_{i-1} ) / 16 \quad (6.1)$$

onde,

$J_i$  é o valor do *jitter* até o  $i$ -ésimo pacote;

$J_{i-1}$  é o valor do *jitter* até o pacote anterior ao  $i$ -ésimo;

$D_{i-1,i} = ( R_i - R_{i-1} ) - ( T_i - T_{i-1} )$ ;

$R_i$  é o horário de recebimento do  $i$ -ésimo pacote;

$R_{i-1}$  é o horário de recebimento do pacote anterior ao  $i$ -ésimo;

$T_i$  é o horário de transmissão do  $i$ -ésimo pacote (*timestamp*);

$T_{i-1}$  é o horário de transmissão do pacote anterior ao  $i$ -ésimo (*timestamp*);

$|D_{i-1,i}|$  é o módulo de  $D_{i-1,i}$ .

O *jitter* é potencialmente mais danoso à qualidade do serviço de voz do que o atraso total fim-a-fim, que cria um intervalo entre o final da fala de um orador e o início da fala do outro.

É causado pela variação do tamanho das filas de transmissão nos nós intermediários de rede, alteração das rotas entre origem e destino ou uso de sistemas de comunicação de satélites de órbitas baixas (LEO), resultando em diferentes tempos de propagação.

A variação do tamanho das filas de transmissão nos nós de rede, pode ser minimizada priorizando-se o tráfego de voz em relação aos outros tipos e impedindo a

grande variação dos tamanhos de pacotes, usando-se a técnica de fragmentação dos datagramas. Neste caso, os pacotes IP são divididos em datagramas com um tamanho máximo especificado, minimizando a variação do tempo necessário para serialização dos mesmos.

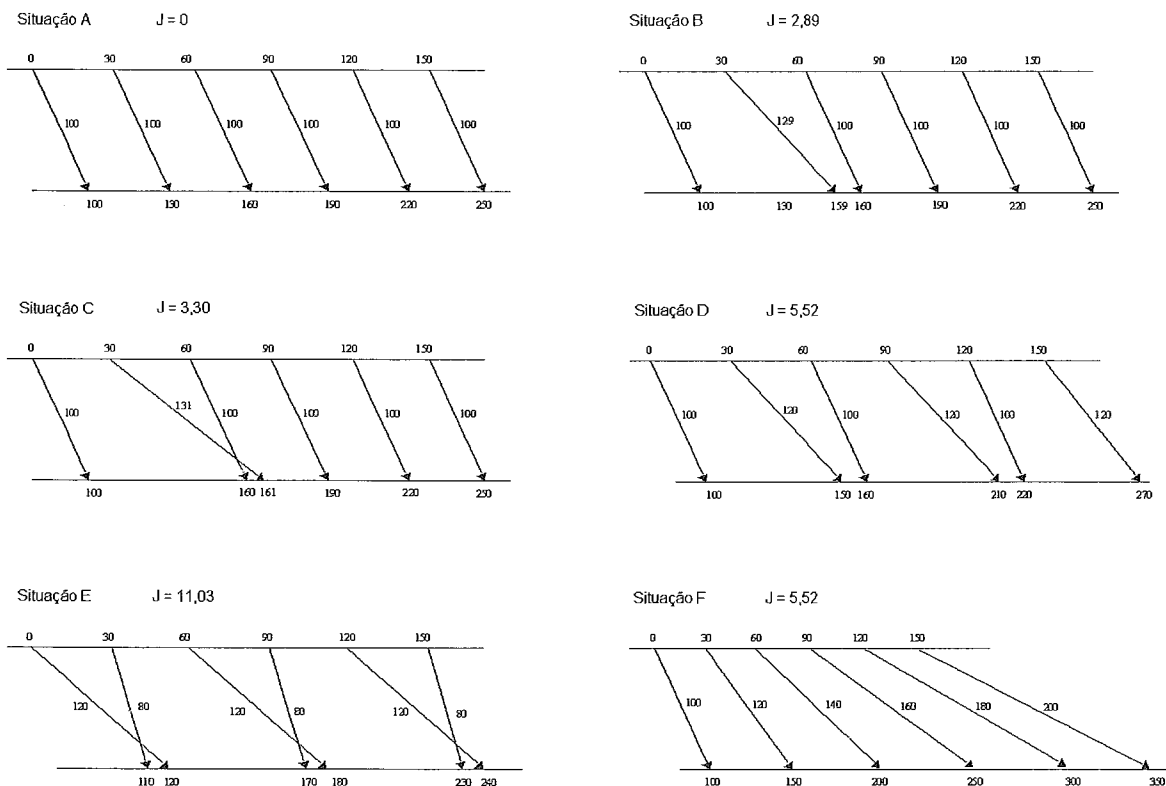


Figura 6.6 – Diagramas de tempo para representação de algumas situações no estudo do comportamento da medida do *jitter*. O valor de J equivale ao *jitter* no final de cada situação independente.

Para amenizar sua conseqüência, os equipamentos de conversão para interconexão ao sistema telefônico convencional (*gateways* de voz), no lado receptor, possuem um *buffer* para armazenamento de alguns pacotes e posterior envio ao sistema telefônico (seja uma central telefônica ou mesmo um telefone analógico ou RDSI – Rede Digital de Serviços Integrados) na cadência originalmente transmitida à rede de dados, no lado do orador. Assim, ambos os lados envolvidos na conversação possuem seus *buffers* de *jitter*, entretanto, cada um deles só tem função em cada um dos lados receptores do fluxo de voz, vide figura 6.7.

A definição ideal do tamanho do *buffer* de *jitter* é muito importante para a qualidade também. Se o *buffer* é pequeno demais, pode não conseguir abrigar uma longa rajada de pacotes, devido a uma anterior retenção na rede ou em um servidor de mídia, ocasionando perda por transbordo. Também por ser de tamanho insuficiente, não será capaz de armazenar um número de pacotes necessário para anular o efeito do *jitter*. Se o *buffer* é grande demais, acentua o problema de interrupção entre as falas dos oradores, devido ao aumento de tempo total fim-a-fim. Por isso, os principais fabricantes implementam seus *buffers* de *jitter* nas opções de tamanho fixo ou dinamicamente variável, de forma automaticamente adaptativa, segundo às condições apresentadas pela rede. O tamanho característico desses *buffers* variam na casa de dezenas de milissegundos, de acordo com os retardos de rede e codificador de voz utilizado. Outra funcionalidade implementada por alguns fabricantes é a reordenação dos pacotes dentro do *buffer*, quando necessário, valendo-se do número de seqüência contido no cabeçalho RTP [40].

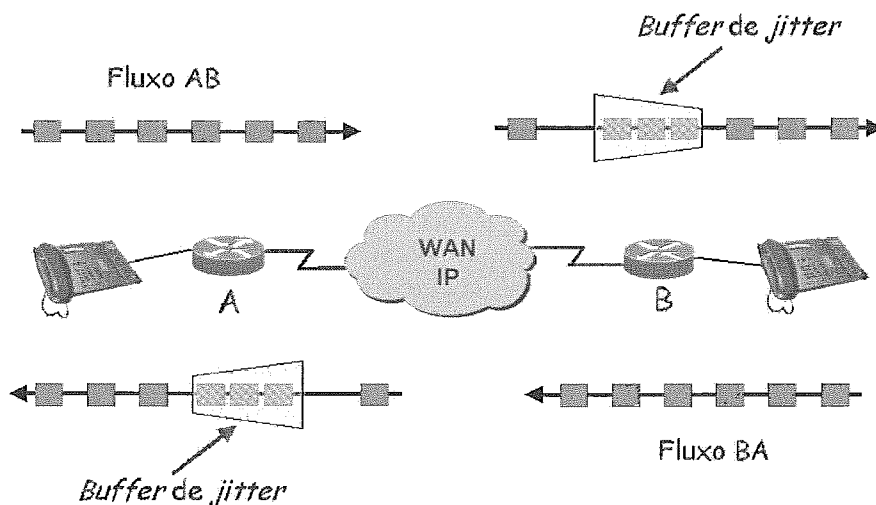


Figura 6.7 – Esquema de utilização do *buffer* de *jitter*.

#### 6.4. Perdas

É comum ter-se perdas de pacotes em redes IP, geralmente atribuídas ao descarte por congestionamento dos *buffers* nos nós de rede ou por substituições por pacotes de maior prioridade. As aplicações que são pouco sensíveis ao atraso, como transferência de arquivos, correio eletrônico e acesso Web, podem dar-se ao luxo do uso de protocolos do tipo ARQ (*Automatic Repeat Request*), para recuperação da informação perdida, por intermédio de sua retransmissão. Contudo, em voz sobre IP



este recurso está completamente descartado, já que o atraso fim-a-fim assumiria valores inaceitáveis para uma boa qualidade de serviço. Desta forma, problemas como erros de transmissão e alterações de seqüência dos pacotes com grande distância (fora da possibilidade do tratamento oferecido pelo *buffer* de *jitter*, conforme mencionado no item 6.3), são tratados na prática também como perdas de pacotes.

### Modelo

Em [41], são apresentadas medições de perdas em uma conexão Internet entre INRIA (*Institut National de Recherche en Informatique et en Automatique*) *Sophia Antipolis*, no sul da França e UCL (*University College London*), utilizando-se tráfegos de voz com 30.000 pacotes em 45 minutos de duração às 08:00h e às 16:00h. A figura 6.8 ilustra a observação feita às 16:00h, onde 16% dos pacotes transmitidos foram perdidos. Sendo  $L$  uma variável aleatória que represente o número de pacotes consecutivos perdidos entre duas chegadas de pacotes corretos, sua média é,  $E(L) = 1,85$ . Cabe ressaltar a atenção devida para a consideração de que pacotes com erro ou fora de seqüência também foram tomados como perdidos.

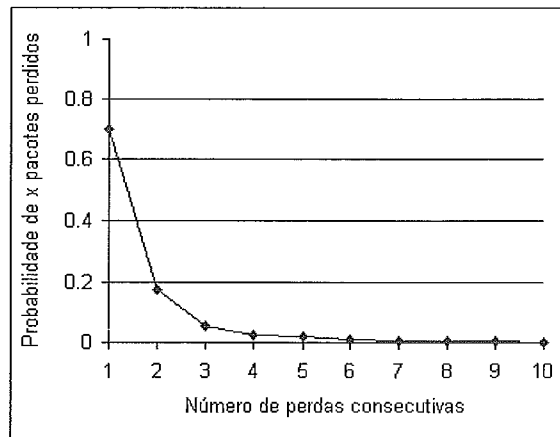


Figura 6.8 – Comportamento da perda de pacotes entre INRIA e UCL.

Comportamento muito semelhante também foi identificado em [42], onde em ambos os casos, aproximadamente 90% das rajadas de perdas são formadas por 1 ou 2 pacotes.

Em acordo com [41], pode-se modelar o processo de perdas de pacotes por um modelo de Gilbert de 2 estados (Cadeia de Markov), conforme a figura 6.9, onde o estado “0” representa o sucesso na recepção de pacotes e o estado “1” a perda de pacotes. Assim, temos que:

$P_{00}$  – a probabilidade de, estando em “0”, continuar-se neste estado;  
 $P_{01}$  – a probabilidade de, estando em “0”, migrar-se para o estado “1”;  
 $P_{10}$  – a probabilidade de, estando em “1”, migrar-se para o estado “0”;  
 $P_{11}$  – a probabilidade de, estando em “1”, continuar-se neste estado.

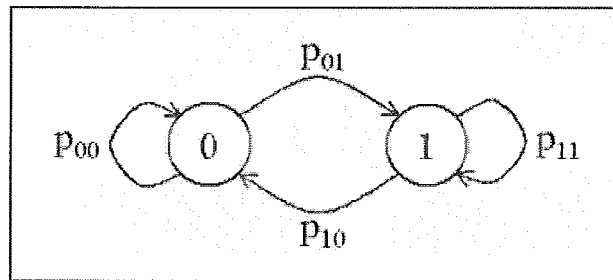


Figura 6.9 – Cadeia de Markov de 2 estados.

Seja  $N$  uma variável aleatória que expresse a quantidade de pacotes consecutivos perdidos e  $p$  uma probabilidade igual a  $P_{11}$ . A probabilidade de ter-se  $n$  pacotes consecutivos perdidos é  $(1-p).p^{n-1}$ , caracterizando uma distribuição geométrica para  $n$  maior ou igual a 1.

Observando a figura 6.8, também poder-se-ia chegar a mesma distribuição, lembrando a relação entre as distribuições exponencial e geométrica, por se tratar de uma abordagem discreta. É visualmente perceptível que a curva da figura 6.8 é semelhante à obtida pela função densidade de probabilidade de uma variável aleatória contínua, com distribuição exponencial.

### Técnicas de contorno

Para abordar o tema de forma mais didática e organizada, serão apresentadas as diversas técnicas para tentativa de contorno do problema em questão. Pode-se dividir essas técnicas em dois grandes grupos: Reparo Baseado no Transmissor e Disfarce de Erro [43].

#### I – Reparo Baseado no Transmissor

Neste grupo de técnicas (figura 6.10), o transmissor do fluxo de áudio participa ativa ou passivamente para a recuperação dos pacotes perdidos.

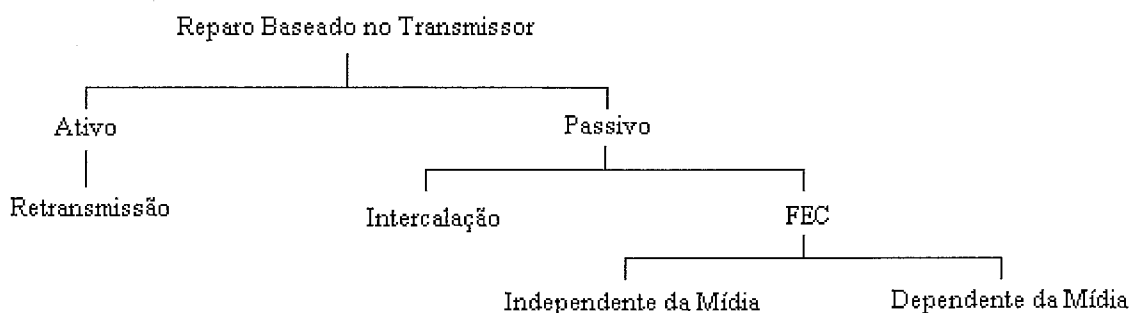


Figura 6.10 – Taxinomia das técnicas de Reparo Baseado no Transmissor.

### I.1 – FEC (*Forward Error Correction*)

Um grande número de variações desta técnica vem sendo desenvolvidas ao longo dos anos. Fundamentalmente, todas trabalham adicionando dados extras ao fluxo original para que o receptor possa recuperar dados perdidos na transmissão. Existem duas classes de FEC, sendo ambas independentes do conteúdo do fluxo, mas valendo-se de conhece-lo para aperfeiçoar o processo.

#### I.1.1 – FEC Independente da Mídia

Esta técnica pode ser implementada usando-se código de bloco ou algébrico, produzindo pacotes adicionais que auxiliam na correção em caso de perdas. Para cada  $n$  pacotes transmitidos à rede, os códigos trabalham com  $k$  pacotes de dados e  $n - k$  pacotes adicionais de conferência.

Existe uma grande variedade de códigos de bloco, mas como exemplos, serão comentados apenas os códigos de paridade e *Reed-Salomon*, propostos para uso no campo de dados úteis do RTP pela RFC 2733. Originalmente estes métodos foram desenvolvidos para detecção e correção de erros na transmissão de um fluxo de bits. No caso de fluxo de pacotes, trabalha-se com o conceito de perda de todo um pacote, aplicando-se o método aos bits correspondentes dentro de cada pacote. Assim, o  $i$ -ésimo bit do pacote de conferência é gerado pelos  $i$ -ésimos bits de cada pacote de dados associado.

No código de paridade, a operação lógica ou-exclusive (XOR) é aplicada no grupo de pacotes de dados, gerando um pacote de paridade. A figura 6.11 ilustra esse método, onde um pacote de paridade é transmitido a cada  $n - 1$  pacotes de dados. Aqui,

apenas um dos  $n$  pacotes pode ser perdido para que ainda seja possível sua recuperação pelo método.

*Reed-Solomon* é reconhecido pela sua excelente característica de recuperação de erros, sendo baseado nas propriedades polinomiais sobre bases numéricas particulares.

Existem várias vantagens neste método. Além de ser independente do tipo de mídia carregada pelos pacotes de dados, também não exige um esforço computacional intenso para a correção de erros. Contudo, o atraso na transmissão de um fluxo de pacotes é grande, aumenta a banda de comunicação necessária e a decodificação é de implementação relativamente difícil.

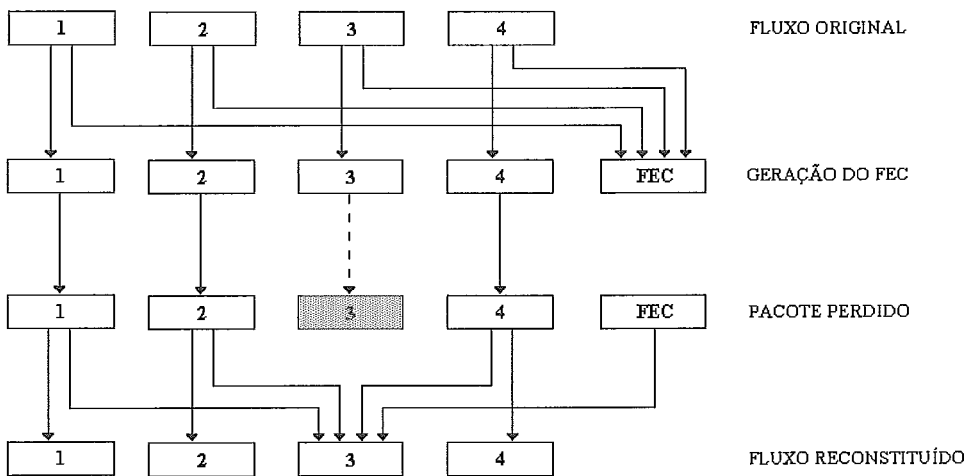


Figura 6.11 – Código de paridade.

### I.1.2 – FEC Dependente da Mídia

Uma forma simples de recuperar pacotes de voz perdidos é transmitindo cada unidade de áudio em múltiplos pacotes. Se um pacote é perdido, outro contém a mesma unidade para tratamento da perda. O princípio é ilustrado na figura 6.12. O transmissor pode decidir em copiar a informação de uma unidade de voz em outro pacote usando o mesmo tipo de codificação que no pacote principal, ou copiar usando uma codificação que consuma menos banda e de baixa qualidade. Esta técnica tem como vantagem o pequeno atraso, quando a cópia é feita sempre no pacote seguinte, muito cômodo para uso em aplicações interativas. No caso da tolerância de grandes atrasos fim-a-fim, a

cópia pode se dar em pacotes mais a frente, permitindo o tratamento de rajadas de perdas.

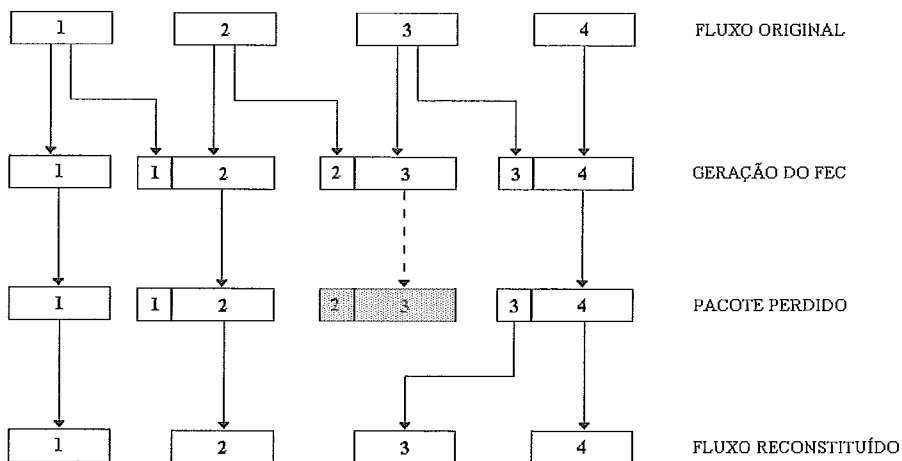


Figura 6.12 – FEC Dependente da Mídia no caso de cópia com menor consumo de banda.

## I.2 – Intercalação

Quando o tamanho da unidade de mídia é menor que o do pacote, e o atraso fim-a-fim é irrelevante, a técnica de intercalação é útil para redução dos efeitos de perdas. Nesta técnica, as unidades de mídia são ressequencializadas antes de serem transmitidas, separando-se as unidades adjacentes para garantir que estarão distantes dentro do fluxo de pacotes, dispersando o efeito de perda. O receptor terá como atribuição a reordenação das unidades, retornando para a forma original, conforme ilustra a figura 6.13.

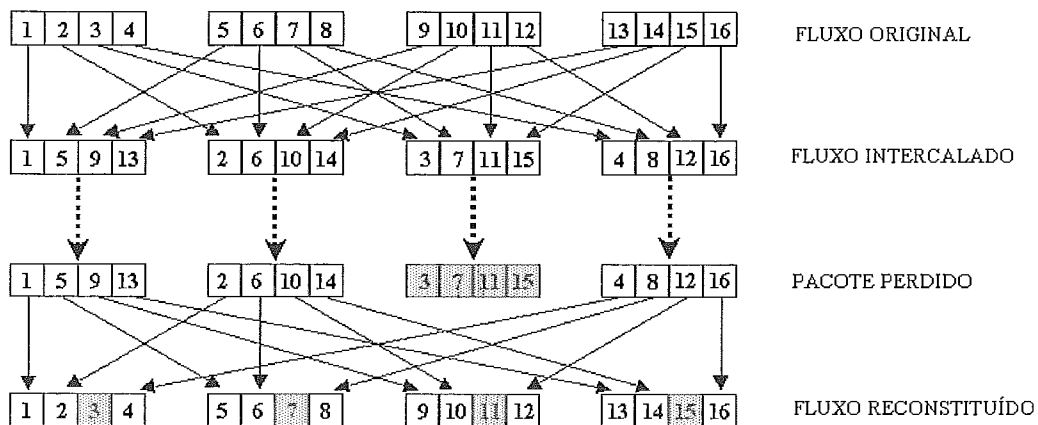


Figura 6.13 – Técnica de Intercalação.

Como vantagem, esta técnica não acarreta aumento no consumo de banda para transmissão do fluxo de pacotes. Contudo, os valores de atraso fim-a-fim são ampliados, limitando seu uso em aplicações com grande interatividade.

### I.3 – Retransmissão

O método de retransmissão do pacote perdido para recuperação da informação, implica tipicamente no aumento do atraso fim-a-fim, em níveis inapropriados a aplicações interativas como VoIP. É uma técnica de fácil implementação e consumo computacional mínimo.

## II – Reparo Baseado no Receptor

Existem várias técnicas para disfarce de erros e perdas, cuja responsabilidade de processamento é do receptor do fluxo de dados, não sendo necessário o envolvimento do transmissor. Estas técnicas são usadas nos casos de falha do esquema de recuperação de perdas do transmissor, ou quando o transmissor do fluxo não tem condições de participar da recuperação.

Os métodos de reparo do receptor baseiam-se na possibilidade de produção de pacotes similares aos originais para substituição dos perdidos. Isto é possível em sinais de áudio e em particular na fala humana. Desta forma, trabalham com valores de perdas relativamente pequenos (<15%) e também com pacotes pequenos que representem uma duração de voz curta (de 4 até 40 ms). Quando o comprimento do pacote perdido se aproxima do tamanho de um fonema (até 100 ms), estas técnicas não funcionam corretamente.

A técnica de reparo baseado no receptor não substitui a baseada no transmissor, mas sim pode trabalhar alinhadas.

Pode ser dividida em três categorias, conforme a figura 6.14.

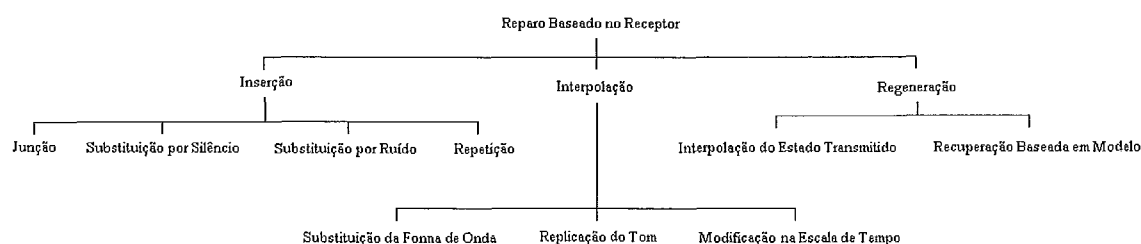


Figura 6.14 – Taxinomia das técnicas de Reparo Baseado no Receptor.

## II.1 – Inserção

Este método tem como princípio básico o preenchimento do tempo respectivo ao(s) pacote(s) perdido(s), dentro do fluxo de dados. Existem variações de técnicas que vão desde a mais simples abordagem com a inserção de nada, isto é, os pacotes vizinhos à perda são unidos, até a substituição da perda por ruído ou por pacotes idênticos aos previamente recebidos.

### II.1.1 – Junção

Unidades de dados perdidos podem ser tratadas com a simples junção dos pacotes vizinhos à perda. No caso de perdas que representam uma curta faixa de tempo (4 – 16 ms) apresentam um desempenho melhor. Contudo, os resultados são intoleráveis para perdas que representem mais de 3% do total de pacotes transferidos.

Esta técnica pode interferir no desempenho do *buffer* de *jitter* adaptativo (ou dinamicamente variável), já que reduz a quantidade de dados disponíveis no *buffer*, induzindo-o a uma interpretação errônea sobre o comportamento do fluxo de dados.

Desta forma, a junção não é uma técnica recomendável.

### II.1.2 – Substituição por Silêncio

Esta substituição preenche a lacuna de uma perda com pacotes de silêncio, mantendo a relação de tempo entre os pacotes circunvizinhos. Só apresenta bom desempenho com pacotes de comprimentos curtos (menor que 5 ms) e baixa taxa de perda (2%), mostrando um resultado mais satisfatório quando comparado com a técnica de intercalação, em caminhos de baixa taxa de perda, para o tráfego de voz.

Seu desempenho degrada rapidamente com o aumento do tamanho dos pacotes, e sua qualidade é normalmente inaceitável para um tamanho de pacote de 40 ms, o qual é usado por alguns produtos de conferências auditivas em redes de dados. A substituição por silêncio tem uma pequena aceitação no mercado, apesar de sua implementação ser muito simples.

### II.1.3 – Substituição por Ruído

Considerando que a substituição por silêncio não apresenta bom desempenho, uma próxima escolha óbvia é substituição por ruído, onde, ao invés de preencher-se a lacuna formada pela perda com silêncio, é inserido ruído de fundo.

Vários estudos da percepção humana em fala com interrupções, mostraram maior habilidade do cérebro humano para a restauração fonética subconsciente do segmento de fala perdido por som correto, quando usada a substituição por ruído. O mesmo não ocorrendo na substituição por silêncio.

#### II.1.4 – Repetição

A técnica de repetição substitui unidades perdidas por cópias da unidade que chegou imediatamente antes da perda. Graças à baixa complexidade computacional e desempenho bem razoável, possui ampla aceitação no mercado.

No caso específico dos equipamentos CISCO, a repetição é adotada até períodos de perda correspondentes a 50 ms, após isso utiliza-se a substituição por silêncio [40].

#### II.2 – Interpolação

Este método baseia-se no uso da similaridade de padrão entre pacotes e interpolação de unidades de dados, para obter um pacote semelhante ao perdido. As técnicas de interpolação são mais difíceis de implementar e requerem maior processamento, quando comparadas com o método de inserção. Tipicamente seu desempenho é superior aos anteriores.

##### II.2.1 – Substituição da Forma de Onda

Esta técnica usa a informação de áudio antes, e opcionalmente depois da perda na tentativa de substituí-la satisfatoriamente por um sinal de formato próximo ao perdido.

##### II.2.2 – Replicação do Tom

Esta técnica representa um refinamento da anterior, usando um algoritmo de detecção do tom da fala, com informações retiradas de ambos os lados da perda, isto é, através dos pacotes anterior e posterior à perda.

##### II.2.3 – Modificação na Escala de Tempo

A modificação na escala de tempo faz com que os áudios de ambos os lados da perda sejam estirados sobre esta.



Embora tenha uma demanda computacional maior, esta técnica apresenta um resultado melhor que o da substituição de forma de onda e da replicação de tom.

### II.3 – Regeneração

Técnicas regenerativas se valem em conhecer o algoritmo de compressão de voz para derivar parâmetros de codec, sintetizando o áudio de um pacote perdido. Estas técnicas são necessariamente dependentes do codec, mas apresentam bom desempenho devido à grande quantidade de informações de estado usada para a recuperação. Tipicamente, apresentam um consumo computacional intenso.

#### II.3.1 – Interpolação do Estado Transmitido

Para codecs baseados em predição linear, é possível que o decodificador possa interpolar entre os estados. Por exemplo, no caso de perdas de curta duração, o codec de fala ITU G.723.1 interpola o estado dos coeficientes dos preditores lineares de ambos os lados da perda e usa uma excitação periódica igual ao do quadro anterior ou ganho obtido de um gerador de número aleatório, dependendo se o sinal era de voz ou silêncio. Para perdas mais longas, o sinal reproduzido é gradualmente esmaecido.

As vantagens de codecs que podem interpolar estados em lugar de recodificar o áudio de ambos os lados da perda é que não é notada qualquer diferença nas fronteiras do tratamento e o esforço computacional permanece aproximadamente constante. Porém, deve ser notado que codecs onde a interpolação possa ser aplicada demandam alto processamento.

#### II.3.2 – Recuperação Baseada em Modelo

Em recuperação baseada em modelo, a fala em um ou ambos os lados da perda é ajustada a um modelo, que é usado para gerar a fala em substituição ao período da perda. A maioria dos codecs de fala com baixa taxa de transmissão usam um modelo auto-regressivo em conjunto com um sinal de excitação, aumentando a correlação entre blocos consecutivos de voz. Assim, para blocos curtos (menores que 20 ms), as características da fala entre blocos consecutivos possuem alta probabilidade de serem próximas.

### III – Resumo

É difícil fazer uma comparação precisa entre todas as técnicas acima comentadas. Contudo, a figura 6.15 retrata de forma geral, a relação entre complexidade e qualidade no panorama apresentado.

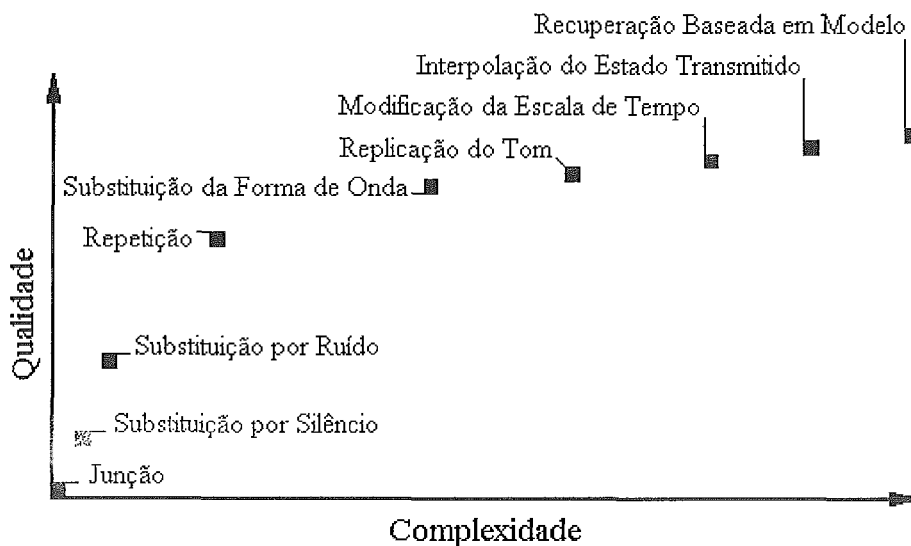


Figura 6.15 – Relação qualidade/complexidade entre as técnicas de tratamento de perdas.

### 6.5. Eco

Em uma conversação de telefone, por vezes um orador pode ouvir sua própria voz com um atraso no tempo. Este fenômeno é chamado de eco, ou também conhecido como eco do orador.

Sua existência pode ter causa elétrica e/ou acústica. No primeiro caso, o problema ocorre apenas em ambientes analógicos. Caso todo o percurso do sinal fosse digital, não ocorreria o eco por motivo elétrico. Contudo, durante todo o percurso da voz, desde sua origem no orador até a chegada ao ouvinte, várias podem ser as causas do eco. A principal delas é causada por um dispositivo encontrado nas centrais telefônicas denominado híbrida. Sua função é a de separação dos sinais de transmissão e recepção no acoplamento entre trechos com 2 e 4 fios. No caso de uso de um tronco telefônico com conexão analógica de 4 fios E&M, a central deve acoplar este circuito ao cabo local do aparelho telefônico com 2 fios apenas. Este acoplamento nem sempre é perfeito, havendo retorno do sinal recebido. A híbrida também é encontrada dentro dos

aparelhos telefônicos analógicos, para acoplar a linha ao sistema de cápsulas para fala e escuta. Outras fontes de eco por causa elétrica, podem ser a indução entre os condutores internos do fio que liga o aparelho telefônico ao fone ou o inadequado casamento de impedâncias entre a interface analógica da central telefônica e o aparelho. A ocorrência de eco por causa acústica tem como origem, o retorno de sinal de um alto-falante ou fone de ouvido para um microfone, através do ambiente.

Para que o eco seja percebido pelo orador e interfira na qualidade da comunicação, faz-se necessário que ocorram três fatores:

- A existência de uma fonte de eco;
- A potência do eco em relação ao do sinal original seja relevante e
- O deslocamento do sinal do eco em relação ao original, na escala do tempo, seja superior ao limiar sensível à percepção humana.

Segundo a Recomendação G.131 do ITU-T [44], o tempo limite médio para defasagem entre o sinal original e o eco é de 25 ms. Entretanto, para tratar o assunto com maior rigor, a figura 6.16 mostra a relação entre o tempo de transmissão em um único sentido da comunicação, inerente ao sistema observado, e a intensidade do eco em relação ao sinal original.

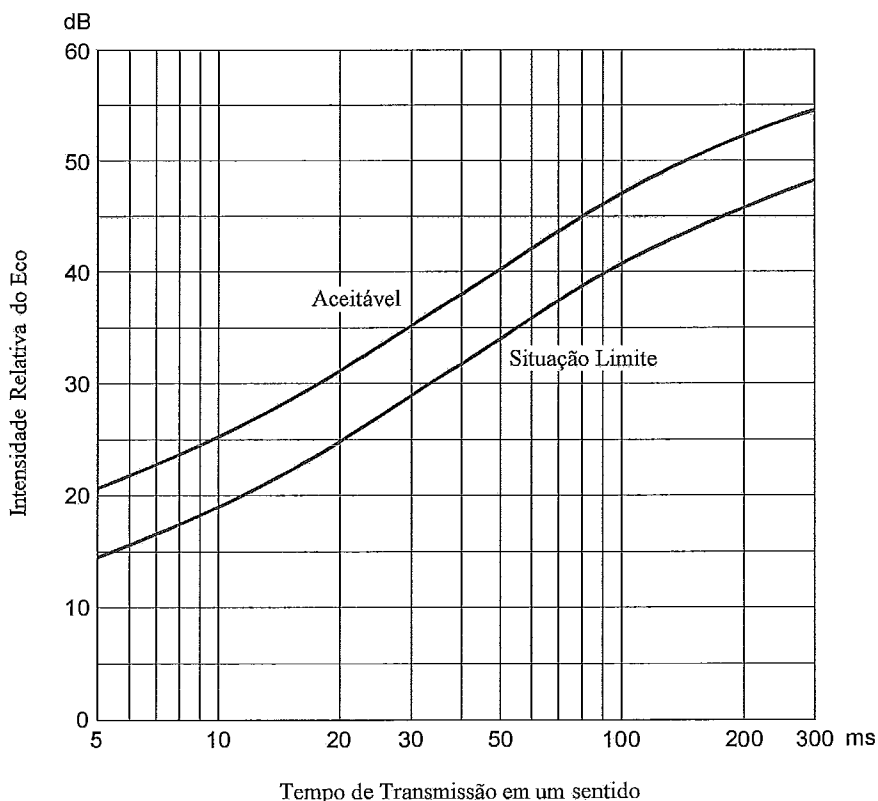


Figura 6.16 – Curvas de Tolerância do Eco.

A curva “Aceitável”, mostra o limite geralmente adotado entre as qualificações tolerável (acima da curva) e intolerável (abaixo da curva). Em circunstâncias excepcionais, a curva “Situação Limite” pode ser adotada.

Para tentar controlar o problema do eco, existem dois dispositivos que podem ser utilizados. Um é o supressor de eco, com uso restrito atualmente. Este dispositivo tem como princípio básico atenuar drasticamente o sinal de retorno. Quando a fala é detectada no sentido de transmissão, o circuito de recepção recebe um forte atenuador, praticamente inibindo a percepção do sinal de eco. Este tipo de operação pode trazer alguns inconvenientes à conversação telefônica com o corte de parte da conversa, durante os períodos de transição no sentido da conversação ou no caso das pessoas envolvidas tentarem falar ao mesmo tempo. Outro problema grave é o uso de equipamentos de comunicação de dados na rede telefônica, já que existe a necessidade de transmissão simultânea em ambos os sentidos. Para atender a esta necessidade, a Recomendação G.164 do ITU-T [45] descreve a funcionalidade de um “tom para desativação do supressor de eco”, onde um sinal de  $2100 \text{ Hz} \pm 15 \text{ Hz}$  e  $-12 \pm 6 \text{ dB}$ , suspende seu funcionamento até que seja observado um silêncio de  $300 \pm 100 \text{ ms}$ . Mesmo assim, o problema ainda persiste para o caso de aparelhos fax, pois pode-se ter períodos de silêncio superiores aos 400 ms especificados.

O outro dispositivo é o cancelador de eco. Tem um funcionamento mais sofisticado que o supressor, sendo normalmente aplicado na porção digital da rede. Sua idéia básica é de subtrair do sinal recebido o sinal coincidente que foi transmitido anteriormente.

Ambos os dispositivos podem ter implementações analógicas ou digitais, e devem estar presentes nas duas extremidades do trajeto da conversação.

## **6.6. Resposta de compressão**

Como foi visto no capítulo 5, os vários codecs referentes às Recomendações do ITU-T apresentam características próprias, no que tange ao consumo de banda, esforço computacional, atraso e, como não poderia deixar de ser, na qualidade do sinal. É esperado que quanto mais afastando do consumo de banda do G.711, esta economia de recurso seja refletida na degradação da resposta obtida.

O item 4.2, comenta as diferenças entre os codificadores de forma de onda, paramétricos e híbridos, sendo estes últimos os de maior uso no mercado atualmente, apresentando uma boa relação entre consumo de recursos e qualidade.

Na figura 6.17, retirada de [2], são apresentados os valores de qualidade MOS para as Recomendações citadas neste trabalho. O método MOS para pontuação da qualidade de voz é um procedimento de avaliação subjetiva, que é detalhado no item 7.1 deste trabalho. Pode-se notar claramente a relação comentada entre consumo de banda e qualidade de resposta, para cada um deles.

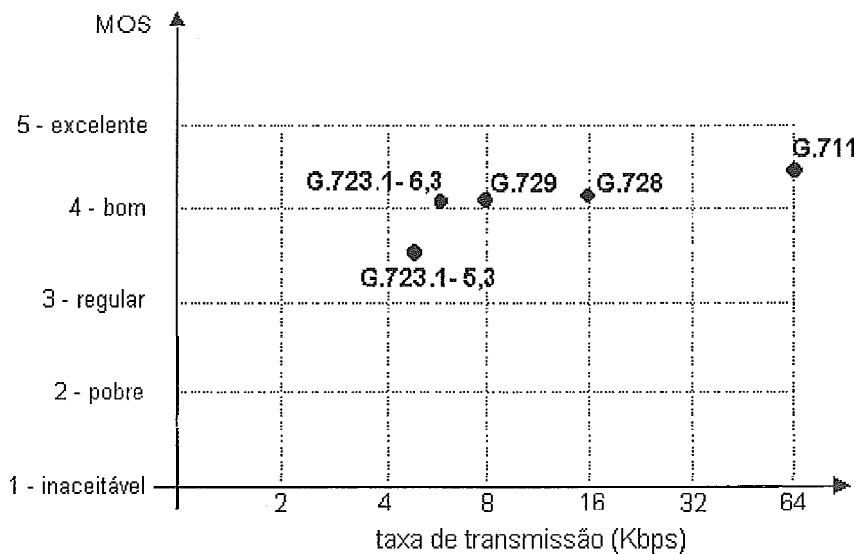


Figura 6.17 – Comparativo entre codificadores, usando MOS.

Usando-se o produto VQT, fabricado pela Agilent *Technologies* do Brasil, que é um equipamento específico para medição objetiva de qualidade de voz, foram feitas medidas comparativas em laboratório para as mesmas Recomendações apresentadas na figura 7.17, mas agora com o método PESQ, que é um método de avaliação subjetiva detalhado no item 7.5. Os resultados obtidos possuem um grau de confiança mínimo de 95% e intervalo de confiança de no máximo 2% [46] dos valores médios apresentados na tabela 6.2.

Tabela 6.2 - Comparativo entre codificadores, usando PESQ.

Recomendação	Valor PESQ
G.711	4,01
G.728	3,60
G.729	3,53
G.723.1 (6,3)	3,41
G.723.1 (5,3)	3,27

## 7. MÉTODOS PARA QUALIFICAÇÃO DO SINAL DE VOZ

Em se tratando de sistemas de voz, seja no uso de telefonia convencional ou de voz digitalizada trafegando por uma rede de IP, a identificação da qualidade da fala é importante para mensurar o quão satisfatório é o sistema. Quanto mais próximos os sinais de entrada e saída do sistema, melhor a qualidade do serviço observado. Outra necessidade para o uso destes métodos é a possibilidade de comparação da resposta dos codificadores (codecs).

Diversos métodos para tal propósito foram utilizados pelo mercado, com alguns deles sendo Recomendações ITU-T. O de maior disseminação é o MOS (*Mean Opinion Score*), empregado em trabalhos acadêmicos, artigos e publicações sobre o tema. Contudo, é um procedimento subjetivo e muito custoso, já que exige um grande número de ouvintes para que a pontuação seja o mais independente possível do grupo de pessoas.

Os itens a seguir apresentam os principais métodos de mercado, comentando suas características e modo de funcionamento.

### 7.1. P.800 (MOS)

#### Sumário

Esta Recomendação descreve métodos e procedimentos que permitem uma avaliação subjetiva da qualidade de transmissão telefônica. De forma corriqueira, é conhecida como MOS – *Mean Opinion Score* (Pontuação de Opinião Média), mas na realidade essa é apenas uma das formas de pontuação citadas nesta Recomendação [47].

#### Origem

A Recomendação P.800 foi revisada pelo Grupo de Estudos 12 do ITU-T, entre 1993 e 1996, sendo aprovada em 30 de agosto de 1996 pela resolução nº 1 do WTSC (*World Telecommunication Standardization Conference*).

#### Escopo

A P.800 contém sugestões para condução de testes subjetivos de qualidade de transmissão em laboratório. Provê métodos considerados convenientes para determinarem o quão satisfatório é o desempenho de dada conexão telefônica.

Entende-se como avaliação subjetiva, aquela que usa procedimentos de conversação ou apenas escuta como métodos para testes, com o fim de aferir a qualidade da transmissão por equipamentos ou serviços de telecomunicações.

Os métodos indicados têm aplicações genéricas, qualquer que sejam os fatores presentes de degradação. Exemplos desses fatores são: perda, ruído de circuito, erros de transmissão, ruído de ambiente, eco, distorção não linear, tempo de propagação, etc. Combinações de dois ou mais desses fatores também são considerados.

## **Métodos Recomendados**

### **a) Teste de opinião por conversação**

Testes de conversação em laboratório pretendem, na medida do possível, reproduzir as condições dos atuais serviços de telefonia, experimentados pelos usuários.

É importante que as condições simuladas nos testes sejam corretamente especificadas, configuradas e medidas de forma acurada, antes e depois de cada experimento. Facilidades de discagem e toque da campainha do telefone devem ser providenciadas, além de registros de saída fidedignos, para cada teste.

#### Cabines de teste

Para execução desse tipo de teste, duas pessoas são colocadas em cabines distintas a prova de som, próximas do ponto onde o experimento é controlado. O volume do local não deve ser menor que 20 m<sup>3</sup>, com um tempo de reverberação menor que 500 ms (normalmente na faixa de 200-300 ms), para os casos de uso de sistemas do tipo aparelho telefônico. Para sistemas *handsfree*, a sala não deve ser menor que 30 m<sup>3</sup>. As cabines devem ser decoradas recriando um ambiente natural.

#### Características de ruído

Ruído mínimo: O nível de ruído deve ser mantido tão baixo quanto possível, com valores próximos aos níveis de ruído observados em hospitais e bibliotecas, isto é, áreas de silêncio.

Ruído ambiente: O ruído ambiente deve obedecer a um nível rigidamente controlado, com equipamentos apropriados e medidos, pelo menos, no início e final de cada experimento. Caso

necessário, as transições entre diferentes ruídos ambiente, devem ser executadas sem a presença dos ouvintes na sala, ou com variações máximas de 4 dB por segundo.

Ruído de sala e

interno em veículos: Esses ruídos são cuidadosamente especificados na Recomendação P.800.

### Escolha dos participantes

Os participantes desse tipo de teste devem ser pessoas usuárias de telefone, escolhidas ao acaso. Não podem estar envolvidas com atividades de medida de desempenho de sistemas telefônicos, ou trabalharem com assuntos relacionados a codificação de voz. Além disso, essas pessoas não podem ter participado de testes subjetivos há pelo menos 6 meses, e em testes de conversação há pelo menos um ano. Caso não seja possível atender ao descrito, isto deve ser registrado na conclusão dos resultados.

Não existe qualquer critério para formação dos pares de conversação. A escolha deve ser aleatória entre os homens e mulheres presentes no grupo.

### Escala de opinião da conversação

A opinião sobre a conexão em curso é classificada em cinco níveis: Excelente, Bom, Regular, Pobre e Ruim.

Os participantes pontuam a experiência seguindo os valores: Excelente = 5; Bom = 4; Regular = 3; Pobre = 2 e Ruim = 1. A média aritmética dos valores atribuídos é denominada pontuação de opinião de conversação, e representada pela simbologia MOSc.

Nota: No passado a equivalência Excelente = 4; Bom = 3; Regular = 2; Pobre = 1 e Ruim = 0 era empregada. Um exemplo desse uso é o anexo B da Recomendação G.114, na versão de 1993, anterior à mais recente, publicada pelo ITU-T no ano de 2000.

### Escala de dificuldade

É uma resposta binária obtida de cada participante, no final de uma conversação. É feita a seguinte pergunta: “Você ou seu parceiro tiveram alguma



dificuldade enquanto falava ou ouvia, durante a conexão?”. São atribuídos então os valores 1 para SIM e 0 para NÃO. A proporção de respostas SIM é representada pelo símbolo %D.

#### **b) Teste de opinião por escuta**

Neste tipo de teste não é esperado obter o mesmo padrão de realismo como o alcançado no teste em conversação. O método de teste mais recomendado para opinião em escuta é o ACR (*Absolute Category Rating*), sendo bastante estável e com aplicação em conexões telefônicas analógicas e digitais. Outros métodos como DCR (*Degradation Category Rating*), *Quantal-Response Detectability Method*, CCR (*Comparison Category Rating*) e *Threshold Method*, também são usados.

O teste de escuta tem uso direto na qualificação de sistemas de transmissão que sejam essencialmente unidirecionais. Os resultados obtidos por esse tipo de teste podem ser usados, com alguma reserva, na qualificação de conversações sobre sistemas bidirecionais, como a rede pública telefônica.

#### Método de teste ACR

Serão descritos os diversos procedimentos para gravação e escuta, que devem ser seguidos neste método.

A fim de eliminar variações indesejáveis na fonte da fala, as amostras devem ser obtidas respeitando-se alguns critérios:

##### i. Ambiente de gravação

A pessoa que irá falar deve estar sentada dentro de uma sala silenciosa, de volume entre 30 e 120 m<sup>3</sup> e com tempo de reverberação menor que 500 ms (preferencialmente entre 200 e 300 ms). O nível de ruído na sala deve estar abaixo de 30 dBA, sem picos dominantes no espectro.

##### ii. Sistema de envio

O telefone local ou o Sistema de Referência Intermediário (*Intermediate Reference System – IRS*) deve ser escolhido conforme especificado na Recomendação P.48 e calibrados de acordo com a Recomendação P.64.

É recomendado que as características de sensibilidade em frequência da conexão sejam medidas pelo menos duas vezes, no início e no fim de cada experimento. Qualquer variação na comparação entre essas duas medidas, deve ser avaliada pelo responsável, que pode duvidar da validade do experimento.

### iii. Sistema de gravação

O sistema de gravação deve ser de alta qualidade (semelhante aos de estúdios de gravação) e possuir um dos itens a seguir:

- Um gravador de fita convencional com duas trilhas e equalização fixa;
- Um processador digital de áudio de dois canais, com um gravador de vídeo cassete de alta qualidade ou DTA (*Digital Audio Tape*);
- Um sistema de armazenamento digital controlado por computador.

### iv. Fala

O tipo de fala a ser usada consiste de sentenças curtas e simples, escolhidas ao acaso e que sejam de fácil entendimento. Podem ser retiradas de jornais ou literatura não técnica, por exemplo. Deve ser formada uma lista contendo sentenças, sem conexão óbvia entre elas. Cada sentença não pode ser muito curta ou longa em demasia, com tempo de pronúncia ideal entre dois e três segundos cada.

O responsável pelo experimento decide quantas sentenças são necessárias para formar cada grupo para amostra de fala. O mínimo de dois e máximo de cinco é o recomendado.

Grupos são combinados em listas com cinco ou dez grupos cada, de modo que a lista completa é usada como uma série de amostras sujeitas ao mesmo tratamento, mas com nível de escuta ou outros parâmetros alterados, enquanto a lista é pronunciada.

Exemplo de material de fala:

Não existe nada para ser visto;

Eu quero um minuto com o inspetor;

Ele precisa de dinheiro ?.

#### v. Procedimentos de gravação

A fala deve ser gravada utilizando-se um microfone linear e um amplificador de baixo ruído, conforme especificado na publicação IEC 581-5. O microfone deve estar posicionado a uma distância entre 140 e 200 mm dos lábios. Em alguns casos faz-se necessário o uso de um anteparo “corta-sopro”, para que a respiração do orador não seja notada.

Dois sistemas de gravação são usados simultaneamente: um grava toda a faixa de frequência da fala em um canal e o outro a fala em resposta do telefone, no canal correspondente. Este procedimento é necessário para o caso de ser preciso comparar as duas versões.

O nível da fala é definido pela Recomendação P.56 e observado durante toda a gravação.

Para reduzir o risco de resultados dependentes de peculiaridades das vozes escolhidas para as falas, é essencial que mais de uma voz masculina e feminina sejam usadas de forma balanceada.

Da mesma forma que foram listados critérios relativos à fala, também deve-se observar procedimentos rígidos para a escuta, conforme listados a seguir:

##### i) Ambiente de escuta

Deve obedecer às mesmas condições da sala de gravação, atendendo também aos critérios de ruído ambiente mencionados no teste de opinião por conversação (item a).

##### ii) Sistema de escuta

O sistema de telefone local, sistema de alto-falante, etc., deve ser calibrado de acordo com a Recomendação P.64. É recomendado que as características de sensibilidade de frequência de recepção sejam medidas pelo menos duas vezes, no início e no fim de cada experimento. Qualquer variação significativa entre as medidas pode invalidar o experimento.

##### iii) Ouvintes

As pessoas escolhidas para os testes devem ser usuários de telefone, escolhidas ao acaso. Não podem estar envolvidas com atividades de medida de

desempenho de sistemas telefônicos, ou trabalharem com assuntos relacionados a codificação de voz. Além disso, essas pessoas não podem ter participado de testes subjetivos há pelo menos 6 meses, não podem ter participado de testes de opinião por escuta há pelo menos um ano e por fim, nunca devam ter escutado a mesma lista de sentenças antes. Caso não seja possível atender ao descrito, isto deve ser registrado na conclusão dos resultados.

iv) Escalas de opinião recomendadas pelo ITU-T

Várias escalas de julgamento com cinco níveis podem ser empregadas, dependendo do seu propósito. A forma de apresentação e as palavras usadas nos experimentos subjetivos têm grande importância. As escalas de opinião mostradas a seguir são freqüentemente adotadas pelo ITU-T:

- Escala de qualidade de escuta

Tabela 7.1 - MOS

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

A avaliação qualitativa dessa escala (tabela 7.1) é representada pelo símbolo MOS (pontuação de opinião média da qualidade de escuta, ou simplesmente pontuação de opinião média).

- Escala de esforço de escuta

Tabela 7.2 -  $MOS_{LE}$

<b>Esforço necessário para entender o significado das sentenças</b>	<b>Pontos</b>
Completamente relaxado; sem necessidade de esforço.	5
Necessidade de atenção; pequeno esforço.	4
Necessidade de esforço moderado.	3
Necessidade de esforço considerável.	2
Não existe entendimento, mesmo com todo esforço possível.	1

A avaliação qualitativa dessa escala (tabela 7.2) é representada pelo símbolo  $MOS_{LE}$ . Quando não é possível o uso da notação com texto subscrito, pode ser adotado o símbolo  $MOS_{le}$ .

#### v) Instruções aos ouvintes

Um exemplo típico de instruções é apresentado na tabela 7.3. Elas devem ser apresentadas antes do início do experimento, podendo ser verbais, caso necessário. Após serem completamente entendidas, o ouvinte deve escutar algumas sentenças para praticar o emprego da pontuação. Nenhuma opinião que direcione serem exemplos com boa ou má qualidade, deve ser sugerida a ele. Nem deve ser submetido a uma grande quantidade de exemplos, que cubra toda a faixa de pontuação. Depois das sentenças preliminares exemplificadoras, o ouvinte deve ter tempo suficiente para tirar qualquer tipo de dúvida, desde que não seja de cunho técnico. Perguntas técnicas só poderão ser respondidas após o término do experimento.

Tabela 7.3 – Exemplo de instruções

Neste experimento você irá ouvir um grupo de sentenças curtas, através do telefone.	
Na mesa em frente a você, existe uma caixa com cinco botões iluminados. Quando todas as lâmpadas se acenderem, você deve escutar ... sentenças, e após todos os botões se apagarem, pressione o botão que indica sua opinião segundo a escala abaixo:	
<b>Esforço necessário para entender o significado das sentenças</b>	<b>pontos</b>
Completamente relaxado; sem necessidade de esforço	5
Necessidade de atenção; pequeno esforço	4
Necessidade de esforço moderado	3
Necessidade de esforço considerável	2
Não existe entendimento, mesmo com todo esforço possível	1
O botão que for pressionado se iluminará por alguns segundos. Então, a lâmpada se apagará e terá uma breve pausa antes que todas as lâmpadas se acendam novamente, para o novo grupo de ... sentenças.	
Haverá uma longa pausa a cada ... grupos. Sendo pronunciado ao todo ... grupos neste experimento.	
<b>OBRIGADO PELA SUA AJUDA NESTA EXPERIÊNCIA.</b>	

#### vi) Análise estatística e resultados

A média numérica deve ser calculada para cada condição de nível de escuta, e devem ser listadas para uma inspeção inicial, de modo que se possa observar os efeitos das vozes masculinas e femininas.

O cálculo do desvio padrão para cada condição em separado não é recomendado. O limite de confiança deve ser avaliado com cuidado.

Para auxílio na análise dos dados, pode-se desenhar gráficos mostrando a pontuação em função de parâmetros do teste, por exemplo: MOS x Atenuação do circuito. Sendo que o MOS sempre deve ser mostrado no eixo vertical.

#### c) Testes por entrevistas e inspeção

A qualidade de transmissão pode ser determinada por *observação de serviço*, quando o grande esforço imputado a este tipo de teste é viável, face a importância do caso considerado. Recomendações de como proceder, incluindo os tipos de perguntas aos usuários, são indicadas na Recomendação P.82. Para obter um alto grau de precisão, no mínimo 100 entrevistas são necessárias.

A desvantagem desse método é o pequeno controle que é possível ter, por vários aspectos, sobre detalhes das características das conexões telefônicas testadas. Entretanto, consegue-se ter uma apreciação global sobre o desempenho do sistema num ambiente real.

## **Conclusão**

Este método é bastante útil para a identificação da qualidade da fala tanto em escuta simples, como em conversação. Possibilita ser notada a interferência causada pela perda de pacotes, *jitter* e atraso por propagação na rede, no caso do teste em conversação. Contudo, é um método extremamente complexo de ser colocado em prática, devido ao tipo de ambiente necessário para os experimentos e grande demanda de tempo para coleta de um número significativo de amostras de pontuação, além de ser um método de medida subjetiva de quantificação, o que depõem contra seu uso em um documentos acadêmicos mais rigorosos, já que seus resultados apresentam íntima dependência com o grupo de pessoas que estará julgando a qualidade da voz. Por tudo isso, não será empregado como principal medida neste trabalho.

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

### **Sumário**

Esta Recomendação descreve um método objetivo para estimativa de qualidade de codecs de voz, operando na banda telefônica (300 – 3400 Hz). O cálculo da qualidade é baseado em uma medida denominada PSQM (*Perceptual Speech Quality Measure*), que foi escolhida dentre várias outras estudadas pelo ITU-T, tais como CD (*LPC Cepstrum Distance Measure*), IT (*Information Index*), CHF (*Coherence Function*) e EPR (*Expert Patter Recognition*), por apresentar o melhor desempenho na correlação com a qualidade subjetiva de codecs de voz [48].

### **Origem**

Originalmente este método foi desenvolvido na Holanda, pela empresa KPN *Research* [49], com base nos trabalhos de Beerends e Stemerding [50].

A Recomendação P.861 foi revisada pelo Grupo de Estudos 12 do ITU-T, sendo aprovada em 27 de fevereiro de 1998 pela resolução nº 1 do WTSC.

## Escopo

A medida de qualidade objetiva descrita nesta Recomendação é estimada em testes apenas de escuta, ao contrário da P.800 que pode ser aplicada tanto em testes de escuta como nos de conversação (nos dois sentidos da comunicação).

Esta medida apresenta precisão aceitável na avaliação dos seguintes fatores:

- Resposta de codecs de forma de onda, ou híbridos com taxas superiores a 4 kbps;
- Nível de fala na entrada dos codecs e
- Dependência das características do sinal de origem.

Por outro lado, não existe informação disponível suficiente sobre a precisão na avaliação dos seguintes fatores:

- Erros nos canais de transmissão;
- Ruído ambiente na geração do sinal;
- Música como sinal de entrada e
- Corte (*clipping*) temporal ou de amplitude na fala.

É sabido que este método não possui precisão em experimentos com variação no atraso entre os sinais de entrada e saída do sistema testado.

## Processo de medida

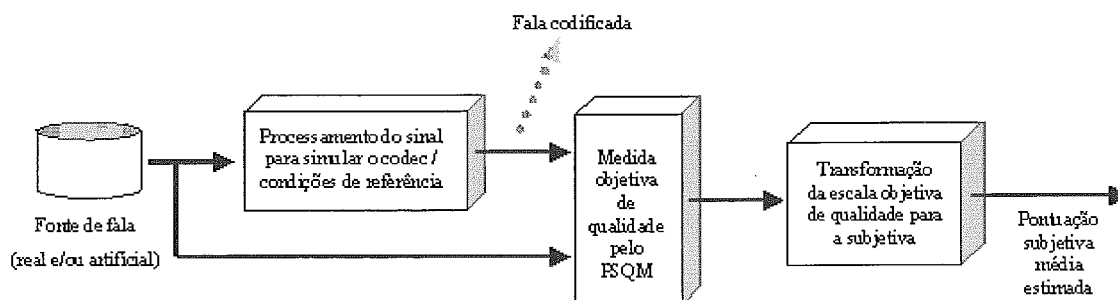


Figura 7.1 – Esquema do procedimento de medida objetiva de qualidade.

A figura 7.1 apresenta um diagrama contendo os principais componentes do processo de medição objetiva de qualidade de voz, segundo esta Recomendação.

### i. Material fonte

O sinal fonte pode ser gerado por vozes reais ou artificiais, como especificado na Recomendação P.50, dependendo do objetivo do experimento. Já que as vozes artificiais da P.50 reproduzem as principais características da fala humana das várias linguagens, elas são úteis para a estimativa da qualidade



de um codec sob as diversas línguas. Para esse caso, é recomendado o uso de voz artificial tanto masculina como feminina. Quando existe o interesse em avaliar o desempenho de um codec para uma linguagem particular, é indicado o uso da voz real, com no mínimo duas vozes masculinas e duas femininas. Se a dependência do tipo de voz for um fator de teste relevante, é recomendado o uso de no mínimo oito homens, oito mulheres e oito crianças.

Nos casos de vozes reais ou artificiais, não se deve adicionar ruído ambiente no material fonte do experimento.

## ii. Seleção dos parâmetros do experimento

Para avaliar o desempenho de um codec, devemos examinar os efeitos de vários fatores de qualidade. A Recomendação P.830 [51] provê um guia para a avaliação subjetiva dos seguintes fatores:

- 1) Nível do sinal de entrada no codec;
- 2) Nível de escuta;
- 3) Oradores (incluindo oradores múltiplos simultâneos);
- 4) Erros no canal de transmissão entre o codificador e o decodificador;
- 5) Taxa de bit, se o codec tiver mais de um modo de operação com taxas diferenciadas;
- 6) Transcodificação;
- 7) Taxas de bit diferentes entre o codificador e o decodificador devido a um engano, caso o codec possa operar com diferentes taxas;
- 8) Ruído ambiente no transmissor;
- 9) Sinalização de rede como entrada no codec e
- 10) Música como entrada no codec.

Nesta Recomendação P.861 emprega-se apenas os fatores de qualidade relativos ao nível de entrada, oradores (excluindo oradores múltiplos simultâneos), taxa de bit e transcodificação, conforme os itens 1, 3, 5 e 6 anteriormente referenciados.

Para todos os efeitos, assume-se que:

- ✓ A fala tomada como fonte é limpa, isto é, sem a presença de ruído ambiente e
- ✓ Não existe degradação no canal, devido a transmissão de bits errados ou perdas de pacotes/células.

### iii. Cálculo da qualidade

Este item descreve o método para medida da qualidade de um codec de voz, para a banda telefônica (300 – 3400 Hz), usando o PSQM. O seu objetivo é reproduzir a percepção dos sons experimentada por um indivíduo, em situações da vida real, simulando experimentos onde indivíduos julgam a qualidade de voz. Isto é possível através da comparação entre os sinais de fonte e o resultante da saída do codec, conforme a figura 7.2 mostra. Como este método é uma representação da percepção e julgamento humano, diferenças inaudíveis entre os dois sinais comparados recebem o mesmo valor PSQM.

Internamente, o PSQM mapeia o sinal físico da fonte e a fala codificada em uma representação psicofísica comparável, tão próxima quanto possível, à representação interna da cabeça humana.

A diferença entre as representações internas dos sinais de entrada e saída, é usada para calcular a perturbação de ruído, como função do tempo e da frequência. A perturbação média do ruído é diretamente relacionada à qualidade da voz codificada, sendo na prática o valor que temos como resposta de uma medida PSQM. Esse valor pode variar de 0 (zero), melhor medida de qualidade, a 6,5 (seis e meio), medida de qualidade indicando a maior degradação do sinal.

A transformação do domínio físico (externo) para o psicofísico (interno), é desempenhado por três operações:

- Mapeamento tempo-frequência;
- Conversão não linear da escala de frequência;
- Conversão não linear da escala de intensidade (compressão).

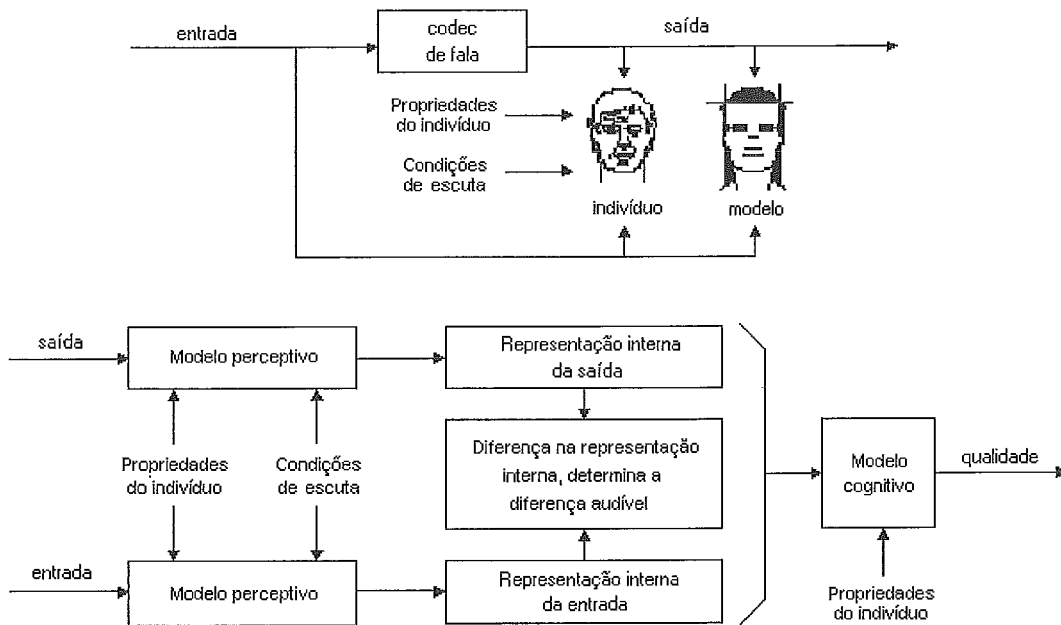


Figura 7.2 – Filosofia básica de funcionamento do PSQM.

Além da modelagem perceptiva, o método PSQM também usa a modelagem cognitiva, a fim de obter alta correlação entre as medidas objetiva e subjetiva.

#### iv. Transformação da escala objetiva para a subjetiva

Em casos onde faz-se necessária apenas a simples comparação de desempenhos de codecs ou em trabalhos de otimização dos parâmetros internos de um codec específico, o valor PSQM é suficiente. Contudo, podem existir circunstâncias onde a comparação do valor PSQM com valores subjetivos seja interessante. Para estimar-se a qualidade subjetiva em escalas de qualidade como MOS ou Q-equivalente, o valor PSQM deve ser transformado considerando-se o descrito a seguir:

##### a) Valor MOS

O método ACR, conforme a Recomendação P.800 é freqüentemente usado para expressar medidas subjetivas de qualidade de voz. Como o relacionamento entre os valores MOS e PSQM não é necessariamente o mesmo para diferentes línguas ou ainda para diferentes testes subjetivos na mesma linguagem, é difícil determinar uma função única na qual pode-se transformar valores PSQM em valores MOS. Na prática, quando resultados de medições

forem expressos no domínio de estimativas MOS, a função de transformação de PSQM para MOS deverá ser apresentada.

Existe uma implementação em equipamento de mercado, que estabelece a correlação entre esses dois valores através da fórmula [50]:

$$\text{MOS} = \frac{4}{(1 + e^{0,66 \times \text{PSQM} - 2,2})} + 1 \quad (7.1)$$

#### b) Valor Q-equivalente

O Q-equivalente é definido como o valor Q do método MNRU (*Modulated Noise Reference Unit*), conforme Recomendação P.810, para o qual o valor MOS é equivalente a ele, dada uma voz codificada.

Na medida objetiva, segundo esta Recomendação P.861, o Q-equivalente pode ser estimado diretamente dos valores PSQM da fala e condições MNRU, sem transformar o valor PSQM em valor MOS, conforme figura 7.3. Contudo, a curva apresentada nesta mesma figura não possui valores numéricos para que seja possível um mapeamento prático entre os valores PSQM e Q.

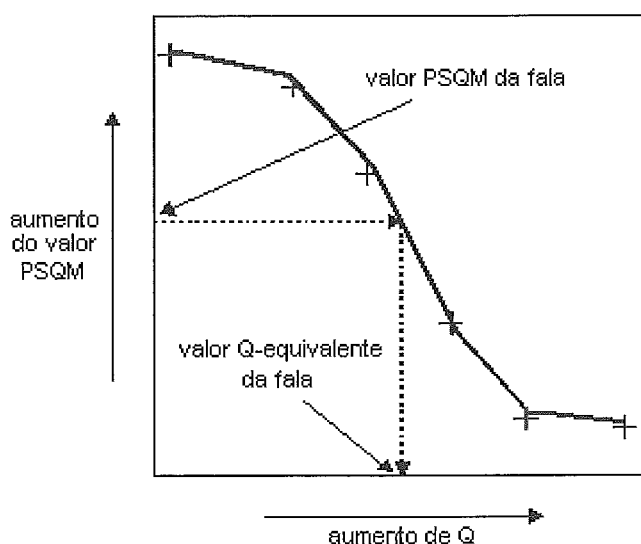


Figura 7.3 – Determinação do valor Q-equivalente em função do valor PSQM, segundo a Recomendação P.861.

## **Conclusão**

Esta Recomendação tem grande relevância por ser o primeiro método objetivo para medição de qualidade, com larga utilização no mercado. Mas, como foi visto, possui sérias restrições na precisão obtida em medições onde são considerados erros nos canais de transmissão e perdas de pacotes, problemas tão comuns nas redes IP que servem com infra-estrutura para o transporte de voz, que é o nosso interesse. Por esse motivo, não será empregada neste trabalho.

### **7.3. PSQM+**

#### **Sumário**

Não existe muita informação disponível na literatura divulgada ao público sobre este método. É dita como uma *evolução* do método PSQM, descrito na Recomendação P.861 do ITU-T, tentando corrigir o reflexo inadequado das interferências da rede de transmissão de dados (por exemplo: perdas de pacotes e taxa de erro de bit), na pontuação resultante do método original. De forma geral, o PSQM+ possui a mesma filosofia do PSQM e mantém o mesmo tipo de pontuação.

O PSQM+, como seu antecessor, também foi desenvolvido pela empresa KPN *Research* e divulgado em 1997, porém não foi adotado como padrão por nenhum organismo internacional de normatização, mas se encontra implementado na maioria dos equipamentos disponíveis no mercado, para medição de qualidade de voz na faixa de 300 a 3.400 Hz.

#### **Descrição das melhorias**

No processo de modelagem cognitiva do PSQM, é notado o efeito de assimetria, apresentando um impacto maior na representação da clareza nos casos de ganho de sinal (adição de energia) do que quando existe subtração de sinal, por conta de perdas de células. Para pequenas distorções que são semelhantes as atribuídas ao codec, PSQM provê excelente correlação com os resultados dos testes subjetivos. Para grandes distorções devido ao corte no tempo ou perda de pacotes onde todas as células de um mesmo quadro são atingidas, resultando em grande perda de energia do sinal, o PSQM produz uma pontuação bem mais baixa que a apresentada pelo teste subjetivo.

Para contornar este problema, foi adicionado ao PSQM+ um segundo fator de pontuação que se opõe ao fator do PSQM, sob condições de distorções severas

representadas por grande perda de energia em uma célula tempo-frequência. Este novo fator é aplicado a cada quadro. Quando as potências dos sinais origem e degradados são próximas no mesmo quadro, o novo fator tem o valor 1,0 e o PSQM+ gera uma pontuação idêntica ao PSQM. Quando uma grande distorção é introduzida, o algoritmo do PSQM+ aplica outra escala de fator que tem efeito contrário, ampliando a sensibilidade da perturbação. Isto resulta no aumento da pontuação PSQM+, refletindo numa melhor correlação com a pontuação obtida nos testes subjetivos [52].

## Conclusão

Como não é um padrão formalmente estabelecido e apresenta documentação com divulgação muito restrita, este método não será adotado neste trabalho.

## 7.4. PAMS

### Sumário

O PAMS (*Perceptual Analysis Measurement System*) é um método de medida objetiva da qualidade de voz em sistemas telefônicos, que procura prever os resultados obtidos nos testes subjetivos. Para obter esta medida, tenta reproduzir a percepção humana usando técnicas de processamento de sinal e um modelo sensitivo diferentes dos empregados no PSQM+, mas com o mesmo objetivo: comparar o sinal original da voz à versão degradada da saída do sistema de comunicação, conforme ilustrado na figura 7.4 [52] e [53].

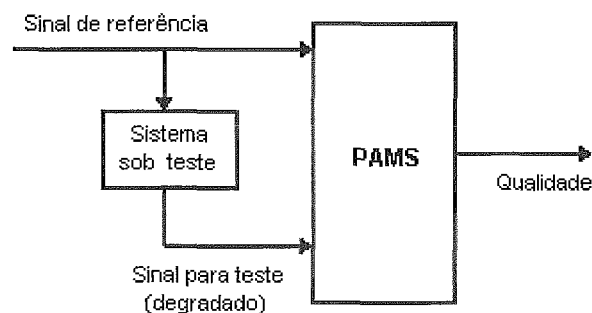


Figura 7.4 – Esquema simplificado de funcionamento do PAMS.

## Origem

O PAMS foi desenvolvido em 1998 pela empresa *British Telecommunications* e teve grande aceitação na Europa, porém não foi adotado como padrão por nenhum organismo internacional de normatização, mas se encontra implementado na maioria dos equipamentos disponíveis no mercado, para medição de qualidade de voz na faixa de 300 a 3.400 Hz. Em fevereiro de 2000 foi lançada sua terceira versão.

## Escopo

A medida de qualidade objetiva do PAMS é obtida em testes apenas de escuta, da mesma forma que no PSQM.

Este método apresenta precisão em seus resultados, quando a clareza da voz é afetada pelos seguintes processos ou parâmetros:

- Codecs de forma de onda (exemplos: G.711 e G.726);
- Codecs paramétricos e híbridos;
- Transcodificações (conversão de um formato digital para outro);
- Nível do sinal de entrada no codec;
- Dependência de oradores (exemplos: linguagem, frases e entonação);
- Rápidas variações de atraso;
- Corte (*clipping*) temporal ou de amplitude e
- Inserção de ruído.

Por outro lado, não é pretendido mensurar o impacto dos seguintes processos ou parâmetros na medição da clareza da voz:

- Atraso;
- Lentas variações de atraso;
- Ganho/atenuação total no sistema;
- Filtragem analógica em telefone;
- Ruído de fundo presente na entrada do sinal de voz e
- Música como sinal de entrada.

## Processo de medida

### i. material fonte

#### a) escolha do material

Gravações de fala natural ou artificial podem ser usadas como sinais de teste. Gravações de fala natural devem conter uma gama representativa e equilibrada de classes de palavras. No caso de gravações diferentes, devem ser concatenadas dentro dos períodos de silêncio, para evitar descontinuidades.

Sinais que não sejam semelhantes à fala não devem ser usados como entrada por várias razões: eles podem fazer a rede se comportar de um modo atípico, não podem testar a qualidade de codecs de fala por completo, e não reproduzem a estrutura temporal de fala que pode ser explorada por elementos como detectores de atividade de voz.

#### b) duração de uma gravação individual

O PAMS é otimizado para gravações com duração de 8 segundos contendo 4 segundos, pelo menos, de fala ativa. Como orientação, o comprimento mínimo para uma medida PAMS ser representativa é aproximadamente de 6 segundos, contendo pelo menos 3 segundos de fala ativa. Gravações com duração de 16 segundos ou mais devem ser divididas em seções mais curtas e cada trecho medido separadamente. Não há nenhuma necessidade dos sinais de referência e degradados serem do mesmo comprimento, pois o PAMS alinha e processa apenas as seções para as quais houver dados disponível de ambos aos sinais.

#### c) múltiplas medidas

Sempre que possível, mais de uma medida, em uma dada condição, deve ser feita para permitir avaliar a qualidade com a variação do tempo ou dependência do material. Pode-se calcular a média das pontuações PAMS de todas as gravações de determinada condição para ter-se uma visão da qualidade global, e a pontuação individual mostra a variação da qualidade do material ou dependência do tempo. No caso de uso da fala, a medida deve ter duração mínima de 28 segundos, e é recomendado que isto seja dividido em três ou quatro arquivos. Se a fala natural for usada como um sinal de teste, devem ser considerados 32 segundos como um mínimo (8 segundos para cada um de quatro oradores) e, se possível, até 2 minutos (16 gravações com 8 segundos de duração) como tempo máximo.



## ii. descrição do modelo

Tanto o sinal de voz original, que serve como referência, como o sinal degradado pelo sistema de rede sob teste, servem de entrada para o modelo PAMS, que executa os alinhamentos de tempo e de nível e a equalização para remoção dos efeitos de atraso, ganho/atenuação globais do sistema e filtragem analógica do telefone. O alinhamento de tempo é executado dentro de segmentos de tempo, de forma que os efeitos negativos de variações lentas do atraso são removidas. Entretanto, os efeitos perceptíveis das variações de atraso são preservados e refletidos na pontuação PAMS.

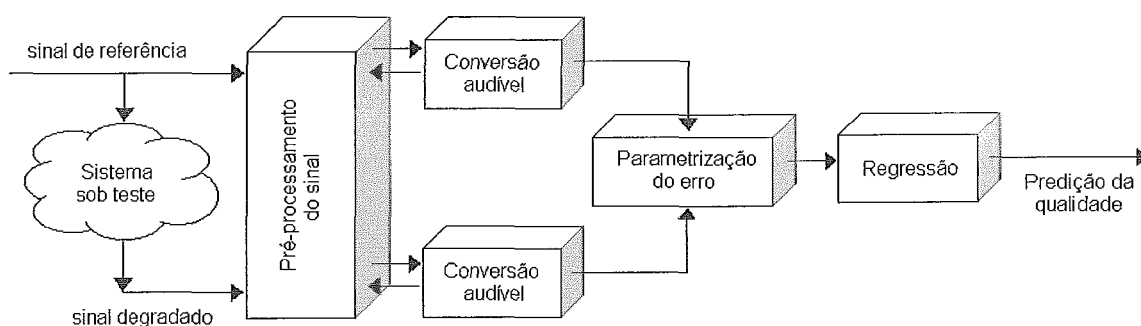


Figura 7.5 – Processo de medida do PAMS.

O processo de medição pelo PAMS, ilustrado na figura 7.5, possui os seguintes estágios:

**Pré-processamento do sinal** – O PAMS pré-processa os sinais de referência e degradado, executando as operações de:

- Alinhamento de tempo - O PAMS é um modelo de escuta que não tem conhecimento do atraso inserido pelo sistema. Para comparar os sinais de referência e degradado, eles precisam ser alinhados entre si. O alinhamento de tempo faz isto dividindo o sinal em segmentos de tamanhos variados conhecidos como expressões vocais, e achando o atraso de cada segmento. Isto permite cancelar qualquer tamanho de atraso e também a maioria das mudanças de atraso que poderiam ser causadas, por exemplo, por transmissão baseada em pacotes.
- Alinhamento de nível - Da mesma forma, o PAMS não tem nenhum conhecimento de mudanças de nível no sistema. Ele então assume um

nível constante de escuta de aproximadamente 79dB SPL, conforme a Recomendação P.830, aplicando um ganho aos sinais de referência e degradado para trazê-los a este nível.

- Equalização - Conexões analógicas introduzem freqüentemente algum grau de filtragem. Por exemplo, o circuito de transmissão de um fone normalmente filtra a fala. Geralmente é aceito que isto tem menos efeito na qualidade que distorções de codificação. O PAMS identifica qualquer filtragem ocorrida na rede e cancela seu efeito.

**Conversão audível** - O PAMS executa um processo de modelagem perceptual para transformar os sinais de referência e degradado no domínio de tempo-freqüência. Usa bancos de filtro *Sekey* para filtrar sinais audíveis até o domínio perceptual pertinente e aloca os sinais em 19 faixas de freqüência, ajustando-as (usando a escala *Bark*) para refletir a sensibilidade humana às freqüências. A escala *Bark* é uma escala de freqüência não uniforme que representa, de maneira mais eficiente que uma escala *Hertz* uniforme, como os humanos percebem as variações de energias auditivas (por exemplo, fala) das diferentes gamas de freqüência, refletindo que em baixas freqüências, o sistema de audição humano tem uma resolução de freqüência maior que em altas freqüências. Nos modelos perceptuais PSQM, PSQM+, PAMS, e PESQ, os sinais de teste comparados são transformados de uma escala de freqüência uniforme para uma escala *Bark*, a fim de melhor representar como os humanos percebem sinais auditivos no domínio da freqüência. Este processo de conversão é feito em amostras com intervalo de duração de 4 milissegundos.

O resultado é uma representação no tempo e freqüência da intensidade sonora percebida, conhecida como Superfície de Sensação, que é análogo à densidade de potência espectral, mas é baseada em como um humano perceberia o sinal em cada célula de tempo-freqüência. O PAMS calcula Superfícies de Sensação para a os sinais referência e degradado.

**Parametrização do erro** - Este método determina as diferenças audíveis nas Superfícies de Sensação dos sinais de entrada (referência e degenerado) subtraindo-as. O resultado é outra representação de tempo-freqüência conhecida como a Superfície de Erro. A Superfície de Erro representa os erros

audíveis, nas células de tempo-frequência, achados quando os sinais de entrada são comparados.

Erros que representam adição da energia no sinal (por exemplo, ruído ou distorção de codec) têm valores positivos em uma célula na Superfície de Erro. Erros que representam perda de energia no sinal (emudecimento, perda de pacote, corte temporal) têm valores negativos. A amplitude de cada célula na Superfície de Erro é relacionada ao nível de percepção ou incômodo humano.

O PAMS analisa a Superfície de Erro de vários modos. Calcula a distorção positiva média (adição de energia no sinal em uma célula) e a distorção negativa média (perda de energia no sinal em uma célula). São calculados vários parâmetros de erro que indicam a quantia de erros audíveis.

**Regressão** - Erros audíveis são avaliados e mapeados em pontuações de percepções da qualidade e esforço de escuta, baseado na correlação com um grande banco de dados de resultados de teste subjetivos. Este banco de dados de resultados de teste subjetivos indica como uma pessoa pontuaria erros audíveis específicos.

### iii. resultado das medições

Como saída desse método, tem-se dois valores distintos:

- Qualidade da escuta - Ylq

Possui relação com a pontuação subjetiva de qualidade da escuta, definida na Recomendação P.800, vide tabela 4. Esta pontuação PAMS é contínua entre 1 e 5, sendo usualmente expressa com duas casas decimais.

- Esforço de escuta - Yle

O PAMS também provê uma pontuação de esforço de escuta, definida pela Recomendação P.800, vide tabela 5. Com ela pode-se testar sistemas onde a inteligibilidade é uma preocupação mais relevante ou onde a qualidade global é geralmente baixa. Esta pontuação PAMS também é contínua entre 1 e 5, sendo usualmente expressa com duas casas decimais.

Como Yle e Ylq são relacionadas a diferentes aspectos de subjetividade, normalmente apresentam valores distintos, caso seja percebida distorção na fala. Normalmente Yle é mais alto que Ylq.

## **Conclusão**

Como não é um padrão formalmente estabelecido e apresenta documentação com divulgação restrita, este método não será adotado neste trabalho.

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

### **Sumário**

O PESQ (*Perceptual Evaluation of Speech Quality*), da mesma forma que o PSQM e o PAMS, é direcionado para medidas de sinais, compreendidos na faixa de 300 a 3.400 Hz, aplicável a sistemas com fala codificada, atraso variável, filtragem, perda de pacotes/células, corte no tempo e erros de canal.

Em 1999, a KPN *Research* promoveu nova melhoria no PSQM+, de forma que refletisse positivamente em testes sob condições reais de rede. Esta melhoria é conhecida como PSQM99.

O PESQ combina a robustez das técnicas de alinhamento no tempo do PAMS com a precisão do modelo perceptual do PSQM99, adicionando novos métodos de equalização e distorção no tempo [54] e [55].

### **Origem**

Entre 1999 e 2000, o Grupo de Trabalho 12 do ITU-T reviu várias propostas de novos métodos para medidas de qualidade de voz, tendo escolhido o trabalho conjunto entre a KPM *Research* e *British Telecommunications* como a nova Recomendação P.862 [55].

### **Escopo**

A medida de qualidade objetiva do PESQ é obtida apenas em testes de escuta ACR. Este método apresenta precisão aceitável em seus resultados, quando a clareza da voz é afetada pelos seguintes processos ou parâmetros:

- Codecs de forma de onda (exemplos: G.711, G.726 e G.727);
- Codecs paramétricos e híbridos (a partir de 4 kbps) incluindo aqueles de múltiplas taxas de transmissão (exemplos: G.728, G.729 e G.723.1);
- Transcodificações (conversão de um formato digital para outro);
- Nível do sinal de entrada no codec;
- Erros no canal de transmissão;
- Efeitos da variação do atraso em testes apenas de escuta;
- Perda de pacotes/células com codecs paramétricos e híbridos
- Ruído ambiente no lado transmissor;
- Taxa de transmissão nos casos de codecs com mais de um modo de operação;
- Deformações temporais do sinal de áudio.

Não é pretendido mensurar o impacto ou não existe precisão nas medidas afetadas pelos seguintes processos ou parâmetros:

- Atraso (cancelado pelo alinhamento de tempo);
- Níveis de escuta e ganho/atenuação total no sistema (é cancelado pelo alinhamento de nível);
- Eco percebido pelo orador;
- Som da própria voz do orador ouvido por retorno no fone.

Por outro lado, não se conhece o impacto dos itens abaixo, nos resultados medidos pelo PESQ:

- Perda de pacotes/células com codecs de forma de onda;
- Cortes (*clipping*) temporal e de nível;
- Dependência de orador;
- Múltiplos oradores simultâneos;
- Diferença entre as taxas de transmissão do codificador e do decodificador;
- Música como sinal de entrada;
- Eco percebido pelo ouvinte;
- Codecs paramétricos ou híbridos com taxa inferior a 4 kbps.

## Processo de medida

### i. material fonte

#### a) escolha do material

Atualmente, todos os testes oficiais de desempenho do PESQ usam voz natural em seus experimentos. As vozes artificial e real concatenada são recomendadas apenas se representarem as estruturas temporal (incluindo intervalos de silêncio) e fonética do sinal de voz real.

Um teste com voz real concatenada deve ser construído agrupando-se pequenos fragmentos de voz real (em torno de um segundo), mantendo-se a estrutura representativa de fala e silêncio.

Caso seja usada voz real, a Seção 7 da Recomendação P.830 deve ser seguida, usando-se no mínimo dois oradores masculinos e dois femininos para cada condição testada. Se a dependência de oradores estiver sendo testada, recomenda-se o uso de oito oradores masculinos, oito femininos e oito crianças. Cabe lembrar que, conforme já comentado anteriormente, a P.862 não foi validada para os testes de dependência de oradores.

#### b) estrutura e duração do material

Os sinais de teste devem incluir rajadas de fala, separadas por períodos de silêncio, a fim de representarem as pausas da fala natural. Um cuidado adicional deve ser tomado pois, certos tipos de VADs (*voice active detectors*) são sensíveis apenas em períodos de silêncio superiores a 200 ms.

Grande parte das experiências usadas para calibragem e validação do PESQ continham pares de sentenças separadas por um período de silêncio, totalizando 8 segundos de duração. Em alguns casos, três ou quatro sentenças foram usadas em gravações que duraram até 12 segundos. É aconselhado o teste com PESQ em experimentos que apresentem estruturas e comprimentos similares aos comentados.

Assim, se uma dada condição deve ser testada por um longo período, é mais adequado que sejam feitas várias gravações de fala distintas (com tempos entre 8 e 20 segundos), e processadas separadamente com PESQ. Esse procedimento traz como benefício o reflexo da variação de tempos, oradores e tipos de fonte, na qualidade mensurada.

É importante alertar que, devido a não linearidade do processo de cálculo do PESQ, a pontuação média obtida sobre um conjunto de arquivos, não é necessariamente a mesma pontuação obtida quando usado um único arquivo concatenado, formado pelo mesmo conjunto.

c) filtragem e calibragem de nível

Os sinais devem passar através de filtros com características de frequência apropriados para emularem um telefone, e equalizadores de nível para o trato do comportamento da voz real.

O PESQ normalmente usa como sinal origem, o material fonte resultante da filtragem do telefone, após também ter sofrido alinhamento de nível.

d) armazenamento

No caso de tratamento de arquivos contendo material de voz para mensuração por PESQ, o formato usado para armazenamento dos sinais de origem e degradado, deve ter amostragem de 8 ou 16 kHz, com codificação PCM linear de 16 bits.

ii. adição de ruído ambiente

É possível usar o PESQ para avaliar a qualidade de sistemas que trafeguem um sinal de voz gerado na presença de ruídos ambiente ou de fundo (como por exemplo: automóveis, rua, público, etc.). O ruído gravado a ser testado deve passar através de um filtro apropriado, com características de transmissão similares ao IRS (*Intermediate Reference System*) modificado, definido no Anexo D da Recomendação P.830. Isso é especialmente importante para sinais de baixa frequência, como o ruído de tráfego de automóveis, que são fortemente atenuados pelo filtro do aparelho telefônico. Após sua filtragem, o sinal deve sofrer alinhamento de nível para um valor desejável para o teste. O processo de adição de ruído é mostrado na figura 7.6.

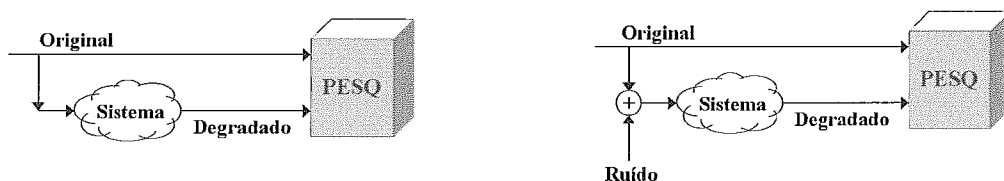


Figura 7.6 – Teste com sinal limpo (esquerda) e com sinal ruidoso (direita).

### iii. descrição do modelo

#### a) visão geral

O PESQ compara um sinal original com um sinal degradado, resultante da passagem através de um sistema de comunicação. A saída PESQ é uma predição da percepção de qualidade que seria obtida por um indivíduo, em teste subjetivo de escuta.

Numa primeira etapa, uma série de atrasos entre os sinais original e degradado são computados, um para cada intervalo de tempo onde os atrasos são significativamente diferentes entre os dois sinais. Para cada intervalo são calculados os pontos de início e fim. O algoritmo de alinhamento tem condições de tratar alterações de atraso tanto nos períodos de silêncio como nos de fala.

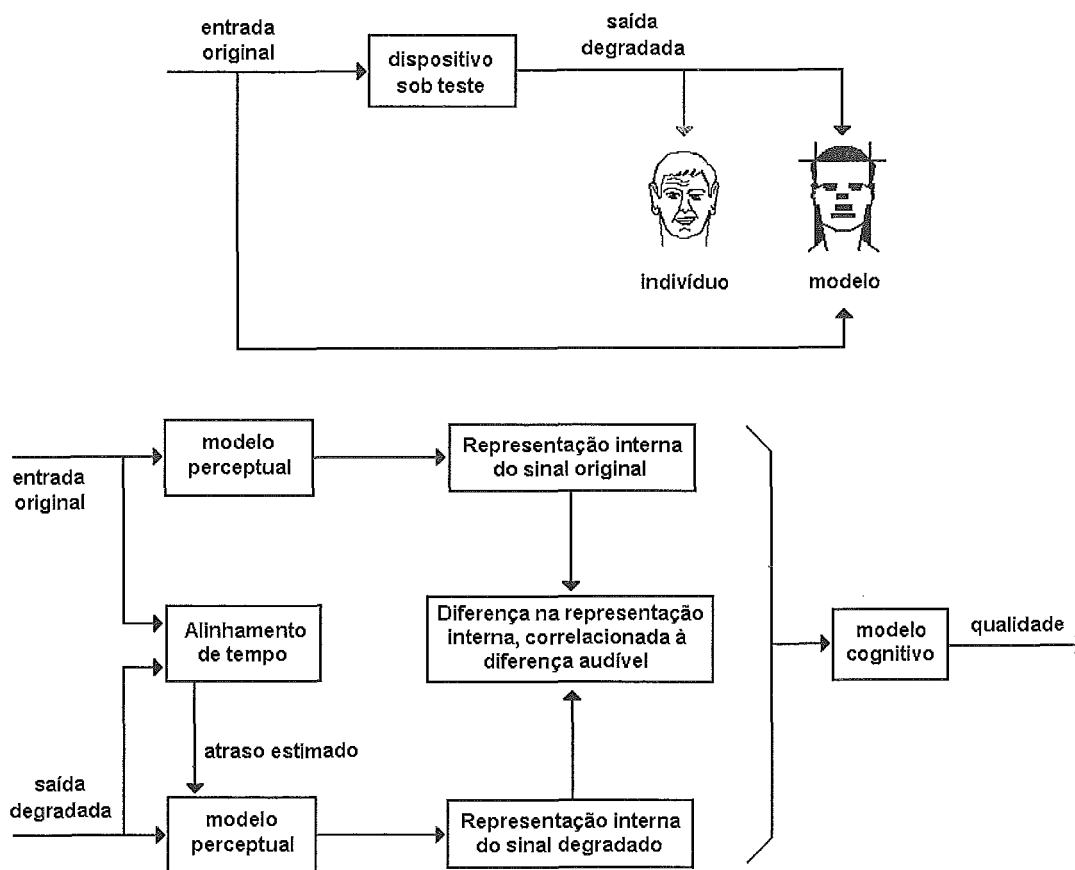


Figura 7.7 – Filosofia básica usada no PESQ. O modelo computacional compara a entrada e a saída do dispositivo em teste.

Baseado no conjunto de atrasos encontrados, os sinais original e degradado já alinhados são comparados utilizando-se um modelo perceptual, conforme ilustrado na figura 7.7. A chave desse processo é transformar ambos os sinais em uma forma de



representação interna que é análoga à representação psicofísica do sinal no sistema auditivo humano, levando em conta a frequência perceptual (*Bark*) e a intensidade (*Sone*). Isto é alcançado em vários estágios: alinhamento de tempo, alinhamento de nível para calibragem do nível de escuta, mapeamento no tempo-frequência, deformação de frequência e compressão da escala de intensidade.

b) descrição do algoritmo

O processo de medição pelo PESQ possui os seguintes estágios:

**Pré-processamento do sinal** – Inicialmente o sinal é analisado pelo modelo perceptual, incluindo as seguintes operações:

- Alinhamento de nível – Os sinais original e degradado são alinhados para um mesmo nível de potência constante, para contabilização nos sistemas de ganho/atenuação. Inicialmente os sinais são filtrados e então computados os valores médios de potência.
- Filtragem – aplicada para formatar os sinais de acordo com as características da banda telefônica. Assim, as características de filtragem de um monofone telefônico não impactam as medidas PESQ.
- Alinhamento de tempo – o sinal degradado é deslocado no tempo para alinhar-se ao sinal original. Isto é executado em segmentos individuais de tempo ou expressões vocais com comprimentos variados. Os atrasos são contabilizados tanto nos períodos de fala como nos de silêncio.

**Modelagem perceptual** – Esta etapa transforma os sinais de origem e degradado em representações internas de acordo com a percepção humana, comparando a diferença entre os sinais. Esta modelagem inclui mapeamento tempo-frequência (similar ao PSQM), deformação no domínio da frequência (usando-se a escala modificada de *Bark*), compressão na escala da intensidade com uma deformação não-linear proporcional à escala em dBm para refletir a sensibilidade humana e trato da distorção que o sinal degradado apresenta em relação ao original.

- Mapeamento tempo-frequência – Semelhante ao PSQM, uma FFT (*Fast Fourier Transform*) é utilizada para transformar os sinais original e degradado do domínio do tempo (potência versos tempo) para o

domínio da frequência. Isto é feito com uma janela de tamanho de 32 milissegundos ou 256 amostras para taxa de amostragem de 8 kHz, segmentando os sinais em células tempo-frequência individuais. Uma célula representa, então, uma banda de frequência específica em um dado quadro de tempo.

- Deformação de frequência – fazendo uso de uma escala de *Bark* modificada, tem a função de refletir a sensibilidade humana a frequências.
- Cálculo da densidade de intensidade – o espectro de potência dos sinais original e degradado são transformados para uma escala *Sone* de intensidade, resultando em grandezas denominadas densidades de intensidade.
- Densidade de perturbação – para cada célula tempo-frequência, a diferença entre as densidades de intensidade degradada e original é calculada. Uma diferença positiva indica ter existido a adição de componentes, como ruído. Já uma diferença negativa indica ter havido a omissão de componentes, tais como distorção de código ou perda de sinal. Estas diferenças calculadas são chamadas de densidade de perturbação.
- Máscara de pequenas distorções – Um limiar é calculado para cada célula tempo-frequência, usando por base o menor valor entre as densidades de intensidade original e degradada, multiplicado por 0,25. Este valor será usado para mascarar o impacto de pequenas distorções que não são percebidas na presença de sinais altos, usando-se a seguinte regra:
  - Se a densidade de perturbação é positiva e maior que o valor de máscara (limiar), o valor de máscara é subtraído da perturbação;
  - Se a densidade de perturbação está entre o valor positivo da máscara e o valor da máscara com sinal negativo, a densidade de perturbação é fixada em zero e
  - Se a densidade de perturbação é menor do que o valor de máscara com sinal negativo, o valor de máscara é acrescentado à densidade de perturbação.

- Processamento da assimetria – A percepção humana de clareza subjetiva de fala é assimétrica. Quando um componente tempo-freqüência não é codificado (perda no sinal de saída), isto afeta menos a qualidade subjetiva que nos casos quando um componente tempo-freqüência inexistente é introduzido (distorção somada). Quer dizer, a adição de uma pequena distorção ao codec é mais notada por um ouvinte humano do que uma pequena perda devido a distorção. Este efeito é modelado calculando-se uma densidade de perturbação assimétrica para cada quadro, multiplicando-se a densidade de perturbação por um fator de assimetria. Este fator de assimetria é obtido pela razão entre as densidades de potência do sinal degradado pelo original, elevada a 1,2. Se o fator de assimetria é menor que 3, é fixado em zero. Caso exceda 12, é cortado neste valor.
- Perturbações de quadro - A densidade de perturbação e densidade de perturbação assimétrica são individualmente integradas (somadas) ao longo do eixo de freqüência, fazendo parte de uma função que dará origem às perturbações de quadro, limitadas ao valor máximo de 45, correspondentes a cada uma das densidades.
- Determinação de variação de atraso – O PESQ pode determinar e identificar quais os quadros envolvidos quando ocorre variação no atraso. No caso de um sinal degradado diminuído por conta de um atraso maior do que 16 ms (metade de uma janela da FFT), entende-se ser melhor ignorar as perturbações de quadro durante este evento na computação da pontuação de qualidade. Como consequência, as perturbações de quadro são zeradas quando isso ocorre, a fim de prevenir falsas pontuações baixas.
- Reavaliação do alinhamento de tempo - Outra avaliação de alinhamento de tempo é executada para perturbações de quadro sucessivas superiores a um limiar. Sendo determinado que o alinhamento de tempo possui um resultando inexato em grandes perturbações de quadro, o alinhamento é repetido e as perturbações de quadro são recalculadas. Os efeitos perceptivos da variação de atraso são preservados nas perturbações de quadro.

- Agregado de valores de perturbação – A seguir, os valores de perturbação de quadro e os valores de perturbação de quadro assimétricos se agregam, independentemente, valendo-se da divisão de subintervalos com 20 quadros sobrepostos, totalizando aproximadamente 320 ms. Estes valores são novamente agregados sobre os intervalos ativos do arquivo de fala.
- Cálculo da pontuação PESQ – Este valor é uma combinação linear do valor médio de perturbação e o valor médio de perturbação assimétrico.

#### iv. pontuação PESQ de qualidade

A pontuação final de qualidade é obtida calculando a distância entre os sinais original e degradado. A faixa de pontuação varia entre -0,5 e 4,5, embora, na maioria dos casos, a faixa de saída deve assumir valores semelhantes aos observados nos testes de escuta do tipo MOS, entre 1,0 e 4,5, sendo esta a faixa normal de valores alcançados em experimentos ACR. Desta forma, quanto maior a semelhança entre os sinais comparados, maior o valor da pontuação obtida, até o máximo de 4,5.

Não existe nenhuma referência na Recomendação P.862 ou em qualquer artigo especializado, mas os profissionais que trabalham com os equipamentos de medição PESQ, atribuem ao valor 3 a menor pontuação para um sinal de boa qualidade. Valores PESQ menores que 3, são tomados como de baixa qualidade, não sendo adequados ao uso corporativo ou comercial, salvo exceções, onde não exista opção com melhor qualidade ou corresponda a uma escolha consciente por solução de baixo custo.

#### v. Comparação entre PESQ e MOS

A proximidade entre o PESQ e a pontuação subjetiva pode ser medida calculando-se o coeficiente de correlação. Normalmente isto é feito com a pontuação média, depois de mapeamento dos valores objetivos para os subjetivos. O coeficiente de correlação ( $r$ ) é calculado pela a fórmula de *Pearson*:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (7.2)$$

Nesta fórmula,  $x_i$  é o valor MOS para condição  $i$ , e  $\bar{x}$  é a média dos valores MOS  $x_i$ .  $y_i$  é a pontuação PESQ para condição  $i$ , e  $\bar{y}$  é a média dos valores PESQ

$y_i$ . Para 22 experimentos de comparação realizados pelo ITU, o coeficiente de correlação médio foi 0,935.

### **Conclusão**

Apesar de ser um método para a identificação da qualidade da fala em testes de escuta simples, é um padrão formalmente estabelecido e apresenta documentação divulgada pelo ITU-T. Possibilita ser notada a interferência causada pela perda de pacotes e variação de atraso por características de propagação na rede, que são os pontos de observação utilizados para identificação das alterações de clareza notadas. Pelo exposto, este método será adotado como principal medida neste trabalho.

## 8. METODOLOGIA DE TRABALHO

### 8.1. Procedimentos de abordagem

Para atender à proposta deste trabalho, foram feitas várias medições de qualidade de voz em ambiente de laboratório, com os três codecs das Recomendações ITU-T de interesse, a saber: G.729 (CS-ACELP) e G.723.1(ACELP e MP-MLQ). Para cada um deles, valores de perdas e *jitter* foram combinados de forma que fosse possível a identificação do comportamento da qualidade medida em função desses valores. Para tanto, foi necessário definir a faixa de valores atribuídos às perdas e ao *jitter*, durante o processo de medidas. Como a bibliografia pesquisada não se mostrou suficiente para auxílio nessa definição, as faixas de valores foram identificadas com testes executados sob redes reais em produção. Assim, o comportamento observado foi reproduzido em laboratório, mas com similaridade a uma rede real. Um alerta importante é que as possíveis imperfeições dos métodos empregados para obtenção dos valores nas redes em produção e suas decorrentes inacurácias, não prejudicaram os resultados alcançados neste trabalho, já que foram utilizados apenas para apontar faixas válidas para as simulações de laboratório e não para identificação de valores exatos.

### 8.2. Medições em redes *em produção*

Foram buscadas medidas em três tipos de enlaces de redes com comportamentos bem distintos e particulares, tentando atender, da melhor forma possível, aos requisitos de cobertura de todos os valores coerentes para perdas e *jitter*. Os enlaces escolhidos foram :

- Trecho da Rede de Comunicação de Dados da Previdência Social entre os Estados do Rio de Janeiro e São Paulo, usando o protocolo *Frame Relay*, com cobertura totalmente terrestre. Este enlace é regido por um contrato com parâmetros rígidos de qualidade, onde os valores de atraso, *jitter* e perda são controlados e mantidos dentro de faixas preestabelecidas;
- Trecho da Rede de Comunicação de Dados da Previdência Social entre os Estados do Rio de Janeiro e Acre, usando o protocolo *Frame Relay*, com cobertura parte terrestre e parte satélite. Este enlace é regido por um contrato com parâmetros rígidos de qualidade, onde os valores de atraso, *jitter* e perda são controlados e mantidos dentro de faixas preestabelecidas;

- Comunicação entre um prédio da DATAPREV no Rio de Janeiro e a UFRJ, através da Internet. Neste caso, os valores de atraso, *jitter* e perda não podem ser controlados.

Deve ficar claro que nos enlaces *Frame Relay*, o protocolo IP também foi usado para transporte da voz. Assim, o protocolo IP foi usado sobre *Frame Relay*.

As medidas de perda e *jitter* nos três enlaces de rede foram feitas através da monitoração do tráfego dos datagramas ICMP gerados por um programa desenvolvido no RAVEL - Laboratório de Redes de Alta Velocidade, que é ligado ao Programa de Engenharia de Sistemas e Computação - PESC, da COPPE/UFRJ. Este programa roda em plataforma Windows e é capaz de gerar tráfego ICMP para um endereço IP especificado pelo usuário, que deve definir também o número de octetos da área de dados do pacote, o intervalo de tempo entre pacotes (em ms), o número de pacotes que será transmitido em uma rajada, o intervalo de tempo entre as rajadas (em ms) e o número de rajadas que deverão ser transmitidas.

A fim de simular o tráfego de voz, para uma maior aproximação com o comportamento real da aplicação VoIP, o número de octetos em cada rajada foi definido seguindo uma distribuição exponencial com: 33 octetos de média, nenhum octeto na área de dados e 30 ms entre a transmissão de cada pacote. Isto porque, segundo as observações do Anexo D, o tempo do período de fala pode ser razoavelmente modelado por uma distribuição exponencial, com média equivalente a 36 pacotes por rajada, para esse caso de uso do codec MP-MLQ (G.723.1 – 6,3 kbps), onde o tamanho de cada datagrama IP foi considerado com 26 octetos (2 octetos para os cabeçalhos IP/UDP/RTP comprimidos e 24 octetos com informação da voz codificada).

No caso ora abordado, como o programa de geração de tráfego está restrito ao uso de pacotes ICMP, mesmo sem nenhum octeto na área de dados, temos 20 octetos do cabeçalho do IP e 8 octetos do cabeçalho do ICMP, totalizando 28 octetos por pacote. Assim, com uma regra de três simples para tentar-se manter o mesmo número de octetos por rajada,

$$26 \text{ octetos} \times 36 \text{ pacotes por rajadas} = 28 \text{ octetos} \times n \text{ pacotes por rajadas}$$

$$\text{onde } n = 26 \times 36 / 28, \quad n = \underline{33 \text{ pacotes por rajadas.}}$$

Para não congestionar a rede desnecessariamente, o tempo entre rajadas foi intencionalmente configurado para valores entre 5 e 15 minutos, dependendo do dia de medição.

As medidas para as conclusões apresentadas no capítulo 9 foram feitas utilizando-se um computador Pentium com sistema operacional Windows 98, rodando o gerador de tráfego e o programa Sniffer da NAI (*Network Associates Technology, Inc.*) para registro dos tempos entre pacotes e das possíveis perdas ocorridas. Este computador estava ligado a uma das portas *Ethernet* do roteador do *backbone* da Rede da Previdência Social, localizado no prédio da DATAPREV do Rio de Janeiro, permitindo acesso às redes *Frame Relay* e Internet.

As medidas de *jitter*, conforme descrito no item 6.3, dependem das aferições dos atrasos de cada pacote. Como não é possível medir o atraso fim-a-fim da rede, entre os pontos origem e destino, com uma acurácia de unidades de milissegundos, o método usado foi de dividir o tempo de ida e volta do datagrama ICMP (*round-trip delay*) por dois, assumindo-se, por simplificação, que os tempos de transmissão em ambos os sentidos são idênticos.

Para o caso da medida das perdas, também foi usado o mesmo critério, assumindo-se que metade das rajadas de perdas observadas ocorreram em cada um dos sentidos da transmissão.

Como já comentado no item 8.1, as aproximações advindas da suposição de equidade de comportamento em ambos os sentidos, não traz qualquer demérito a este trabalho, já que esses dados foram utilizados apenas para apontar a ordem de grandeza a ser adotada nas faixas de variações dos critérios de perdas e *jitter*, que serão simulados em laboratório.

### **8.3. Medições em laboratório**

Para executar as medidas, foi montado em laboratório um ambiente (figura 8.1) composto por: um microcomputador executando um programa emulador de rede de longa distância (WAN), com a finalidade de permitir a alteração do comportamento de perdas de pacotes e variação do atraso de rede, refletindo no *jitter* desejado; um microcomputador executando um programa para obtenção de medidas de qualidade de



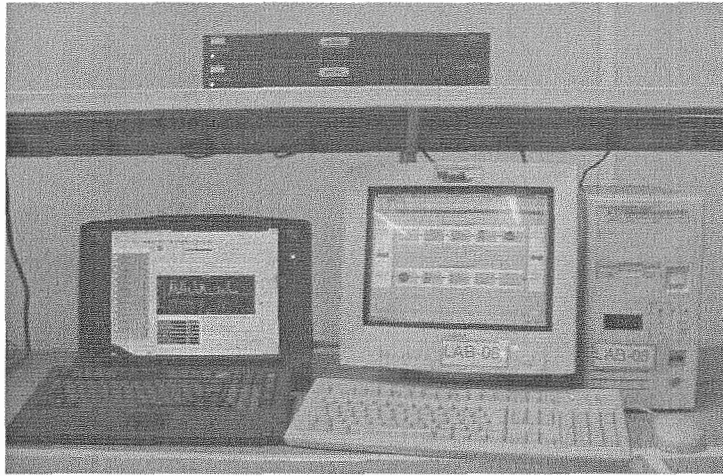


Figura 8.1 – Foto do ambiente de laboratório para as medidas PESQ.

voz usando a Recomendação P.862 (PESQ) e dois roteadores com interfaces FXS, funcionando como *gateways* de voz sobre IP. Os *gateways* foram interligados ao emulador de WAN por *Ethernet* em 10Mbps e ao equipamento de medição PESQ pelas interfaces de voz, conforme figura 8.2.

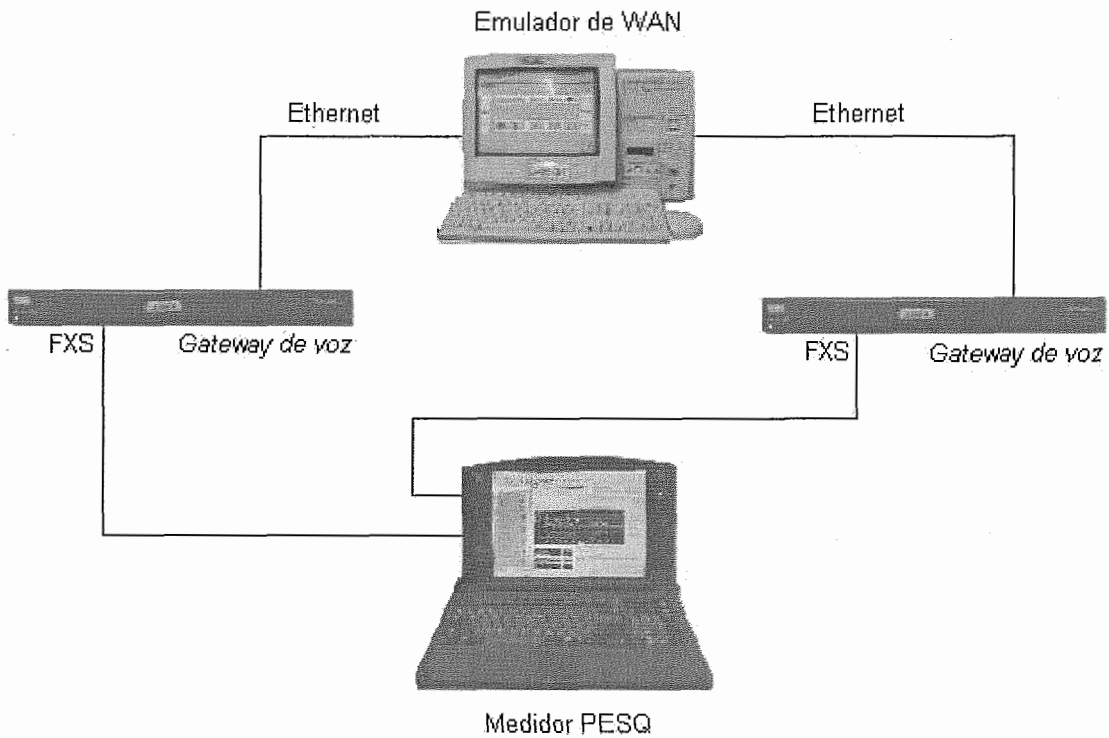


Figura 8.2 – Diagrama esquemático das ligações entre os equipamentos usados para as medidas PESQ.

### 8.3.1. Descrição do ambiente

#### a) Equipamento para emulação de rede de longa distância

*Hardware:* Microcomputador Pentium II 450Mhz com 256Mbytes de memória, com duas placas *Ethernet* 10/100Mbps.

*Software:* Sistema Operacional Windows NT 4.0 com *Service Pack 6* IntSim versão 2.18 da RADCOM Ltda.

Descrição: Este produto é um *software* que emula o comportamento de uma rede de longa distância, através da possibilidade de configuração do atraso, *jitter*, perda de pacotes, velocidade de enlace e inversão de ordem de pacotes. Apenas as opções de *jitter* e perda de pacotes foram utilizadas, para atenderem às situações de cada conjunto de medidas realizadas.

#### b) Equipamento para medição PESQ

*Hardware:* Microcomputador Pentium II 400Mhz com 128Mbytes de memória.

*Software:* Sistema Operacional Windows NT 4.0 com *Service Pack 6* VQT versão 4.100.

Descrição: Este produto é um conjunto de *hardware* e *software*, da Agilent *Technologies* do Brasil, que pode executar medidas de qualidade de voz usando os algoritmos PSQM, PAMS ou PESQ. O exemplar usado continha duas interfaces analógicas de voz para conexão às interfaces FXS do *gateway* de voz e outras duas interfaces para E&M. Sua configuração se dá por várias telas onde são informadas as configurações dos tipos de teste que deverão ser realizados, das interfaces de voz e outras opções disponíveis (figura 8.3).

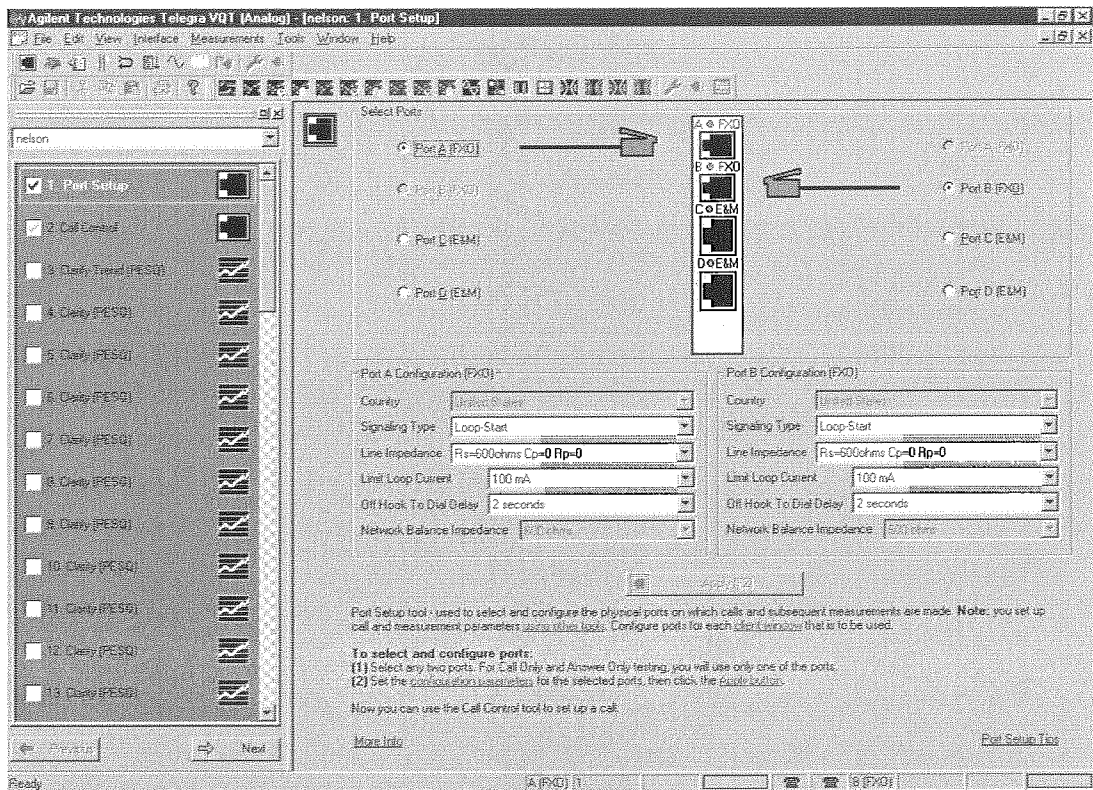


Figura 8.3 – Uma das telas de configuração do VQT. Pode ser observado parte do conjunto de medidas que serão executadas automaticamente (à esquerda) e as interfaces analógicas de voz usadas (centro).

### c) Gateway de voz

*Hardware:* Roteador CISCO 2610 com 48Mbytes de memória DRAM, 16Mbytes de memória *flash*, duas interface FXS (figura 8.4) e uma interface *Ethernet* 10Mbps.

*Software:* IOS CISCO versão 12.1(3a)T7.

Descrição: Roteadores CISCO com interfaces de voz.

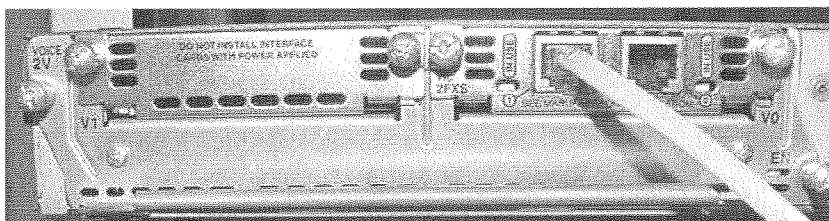


Figura 8.4 – Detalhe da interface FXS no gateway de voz.

### 8.3.2. Material fonte

Como não havia arquivos com falas em português disponíveis no equipamento da Agilent, foram gerados fontes em formato wav, com amostragem de 8kHz em codificação PCM linear de 16 bits, conforme orienta a P.862. Contudo, a pontuação obtida para esse tipo de fonte foi de 3,78, bem abaixo do valor máximo da escala PESQ, mesmo com o uso de G.711 (lei A) e sem considerar-se perdas de pacotes ou *jitter*. Essa pontuação baixa talvez tenha sido em decorrência da forma artesanal com que os arquivos foram criados, sem um equipamento profissional de gravação e filtragem adequada do som.

O material fonte escolhido para as medições foi formado por quatro arquivos wav em espanhol, atendendo ao especificado na Recomendação P.862. Dois arquivos eram de vozes femininas e dois de vozes masculinas, todos com oito segundos de informação cada um, mas com conteúdos diferentes e gerados por quatro pessoas distintas. Esses arquivos já estavam disponíveis no equipamento usado para as medições. Com esse material fonte, foi obtido valor PESQ de 4,01 também com o uso de G.711 (lei A) e sem considerar-se perdas de pacotes ou *jitter*.

Todas as medições foram efetuadas com o mesmo número de repetições de cada um dos quatro arquivos wav, para que nenhuma característica de determinada voz prevaleça nos resultados.

### 8.3.3. Execução das medidas

Os procedimentos de medidas foram feitos configurando-se o produto VQT para executar um conjunto de 48 medidas PESQ de forma automática, eliminando a intervenção de um operador para cada medida unitária. Como já comentado, cada conjunto de medidas continha 12 repetições de um mesmo arquivo wav, abrangendo os 4 arquivos descritos no item anterior.

Os valores de *jitter*, foram obtidos usando-se a opção de *random walk*, onde o atraso gerado tem 50% de chance de aumentar o valor de um passo e 50% de chance de diminuir o valor de um passo, limitado a um valor mínimo e um máximo. Como foi configurado o valor mínimo com zero e tanto o máximo como o passo com o mesmo valor, o atraso variava entre zero e o dobro do valor de *jitter* desejado para a medição.

Para o caso das perdas, foi usada a opção onde configura-se a probabilidade de ocorrer uma rajada de perdas a cada pacote a ser enviado, e o tamanho da rajada desejada.

Os *gateways* de voz foram configurados sem o uso de VAD, para não ocorrer qualquer tipo de interferência proveniente dessa facilidade. Além disso, o *buffer* de *jitter* também foi desabilitado e a entrada do sinal foi configurada com um ganho de 5dB. A seguir foi transcrito parte da configuração feita:

```
interface Ethernet0/0
ip address 10.0.134.2 255.255.255.0

voice-port 1/0/1
input gain 5
playout-delay nominal 0
playout-delay mode fixed

dial-peer voice 1 pots
destination-pattern 100
port 1/0/1

dial-peer voice 2 voip
destination-pattern 200
session target ipv4:10.0.134.3
no vad
```

## 9. RESULTADOS OBTIDOS

### 9.1. Distribuição das Perdas

O objetivo em obter-se o comportamento de distribuição das perdas em uma rede de pacotes, é o de identificar qual a ordem de grandeza e faixa dos valores que devem ser usados na simulação de uma rede em laboratório, durante a medição da qualidade de voz. As medidas para identificação do perfil de comportamento das perdas em uma rede de pacotes foram feitas em três enlaces diferentes, conforme o item 8.2, durante 26 dias diferentes, entre os meses de janeiro a abril de 2002. Os dois primeiros foram trechos da Rede de Telecomunicações da Previdência Social entre os escritórios da DATAPREV nos Estados do Rio de Janeiro, Acre e São Paulo. São enlaces de uma rede *Frame Relay* onde o trecho RJ/SP é totalmente terrestre e o RJ/AC é parcialmente atendido por satélite. Em ambos os casos, o número de perdas foi muito pequeno para que pudesse contribuir com o estudo, no período de realização das medições entre 10:00h e 12:00h e entre 14:00h e 16:00h. No trecho RJ/SP foram observadas 28 perdas em 19.350 pacotes transmitidos, com rajadas de perdas formadas por 1, 2 ou 3 pacotes. Já no trecho RJ/ AC, ocorreu apenas uma perda em 19.245 pacotes transmitidos.

O terceiro enlace utilizado para medição de perdas, foi entre um prédio da DATAPREV no Rio de Janeiro e a UFRJ, através da Internet. Neste caso, as medições se deram de duas formas: considerando apenas o período de maior tráfego (entre 10:00h e 12:00 e entre 14:00h e 16:00h) e no horário comercial completo (entre 09:00h e 18:00h).

No período de maior tráfego, dos 18.481 pacotes transmitidos, 3.142 foram perdidos, agrupados em 2.004 rajadas. Assim, 15.339 pacotes não foram perdidos. Como as medidas foram feitas utilizando-se os pacotes ICMP gerados pelo utilitário PING, as perdas podem ter ocorrido em qualquer um dos sentidos da transmissão. Desta forma, deve-se considerar que das 2.004 rajadas, 1.002 ocorreram em cada sentido. Modelando as perdas de forma que cada pacote tem uma probabilidade  $p$  de iniciar uma rajada, temos  $15.339 + 1.571 + 1.002 = 17.912$  possibilidades de inícios de uma rajada e  $1.002 / 17.912 = 5,59\%$  sendo a probabilidade de um pacote iniciar uma rajada. O tamanho médio das rajadas foi de 1,57 pacotes, com desvio padrão de 1,72. As medidas feitas garantem um grau de confiança de 90%, com intervalo de confiança entre 1,50 e 1,64 para o valor médio (variação de 0,07 em relação à média).

As duas figuras a seguir, mostram a probabilidade de ocorrência de cada tamanho de rajada e a distribuição de probabilidade acumulada, da variável aleatória “tamanho da rajada”, nestes períodos compreendidos entre 10:00h e 12:00h e entre 14:00h e 16:00h.

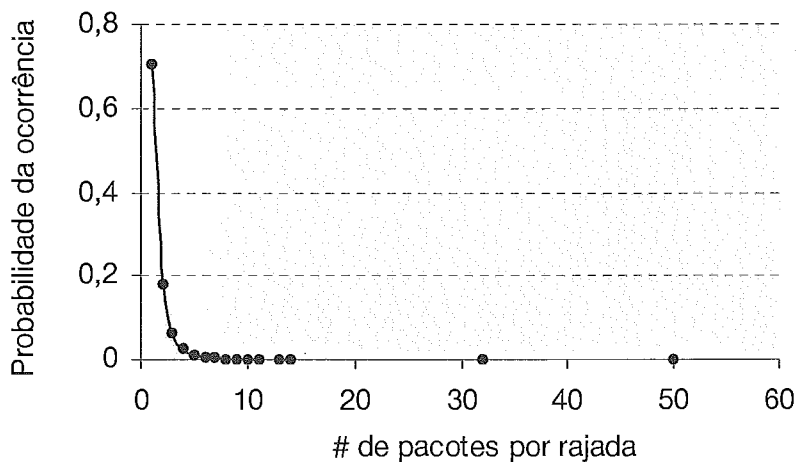


Figura 9.1 – Probabilidade da ocorrência de cada tamanho de rajada de perdas, pela Internet, no horário de maior tráfego.

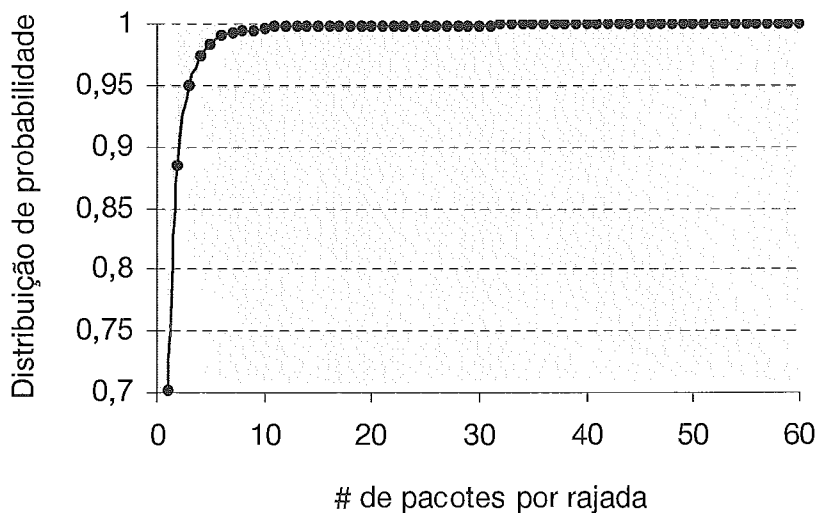


Figura 9.2 – Distribuição de probabilidade acumulada da ocorrência de cada tamanho de rajada de perdas, pela Internet, no horário de maior tráfego.

Já no horário comercial, dos 34.957 pacotes transmitidos, 5.573 foram perdidos, agrupados em 3.700 rajadas. Assim, 29.384 pacotes não foram perdidos. Como da forma anterior, considera-se que 1.850 rajadas de perdas ocorreram em cada sentido. Modelando as perdas de forma que cada pacote tem uma probabilidade  $p$  de iniciar uma rajada, temos  $29.384 + 2.787 + 1.850 = 34.021$  possibilidades de inícios de uma rajada e  $1.850 / 34.021 = 5,44\%$  sendo a probabilidade de um pacote iniciar uma rajada. O tamanho médio das rajadas foi de 1,51 pacotes, com desvio padrão de 1,43. As medidas feitas garantem um grau de confiança de 90%, com intervalo de confiança entre 1,47 e 1,55 para o valor médio (variação de 0,04 em relação à média).

As duas figuras a seguir, mostram a probabilidade de ocorrência de cada tamanho de rajada e a distribuição de probabilidade acumulada, da variável aleatória “tamanho da rajada”, neste período compreendido entre 09:00h e 18:00h.

É sempre bom lembrar que as perdas observadas no horário de maior tráfego também estão contidas nos resultados alcançados dentro do horário comercial.

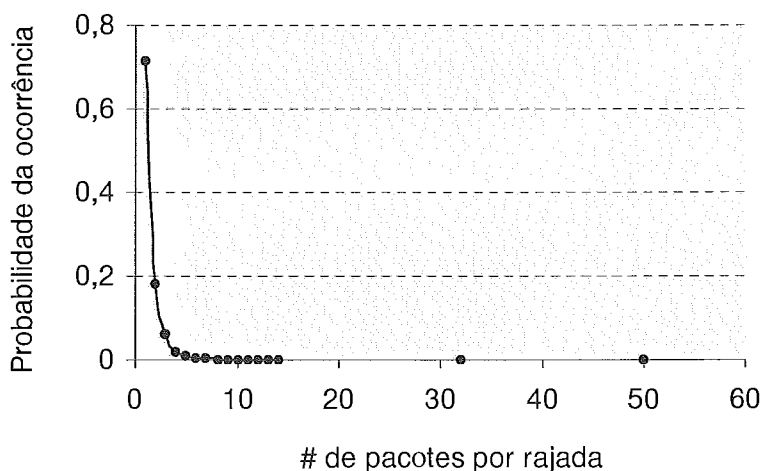


Figura 9.3 – Probabilidade da ocorrência de cada tamanho de rajada de perdas, pela Internet, no horário comercial.



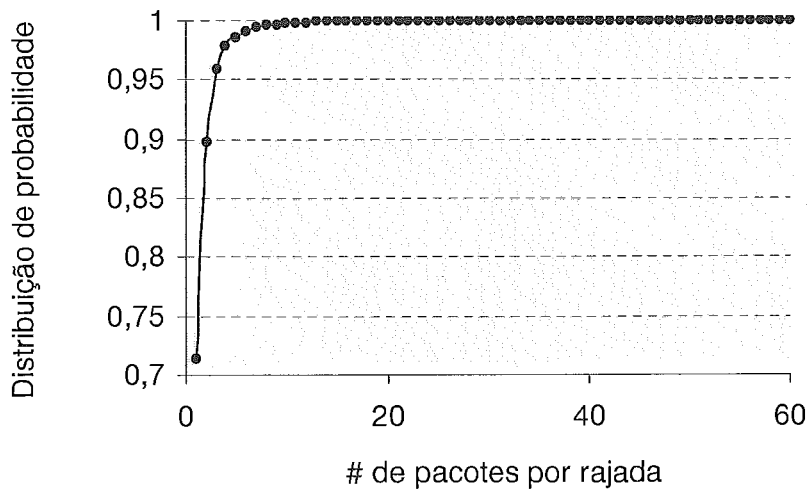


Figura 9.4 – Distribuição de probabilidade acumulada da ocorrência de cada tamanho de rajada de perdas, pela Internet, no horário comercial.

Na figura 9.5 observa-se com mais detalhe o trecho próximo das origens de cada eixo.

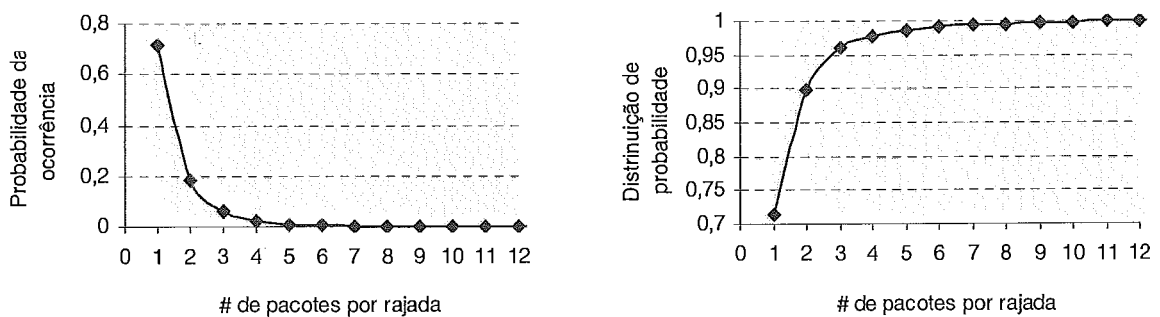


Figura 9.5 – Detalhes das figuras 9.1 e 9.2.

Durante os períodos de maior tráfego, isto é, entre 10:00 e 12:00h e entre 14:00 e 15:00h, 70,26% das ocorrências de rajadas são formadas por apenas uma única perda. Enquanto que, no período entre 09:00 e 18:00h, este percentual aumenta para 71,41%. Isto demonstra um comportamento esperado, com maior número de descartes na rede nos períodos de tráfego mais intenso.

Durante os períodos de maior tráfego, 99,70% das rajadas são formadas por até 10 pacotes perdidos. Enquanto que na observação ao longo do dia, entre 09:00 e 18:00h, este percentual aumenta para 99,76%, considerando o mesmo tamanho de rajada com 10 pacotes.

É certo que em nenhum dos períodos tratados, horário de maior tráfego ou comercial, não se obteve uma média com grau e intervalo de confiança adequados, isto é, grau de confiança maior ou igual a 95% e intervalo de confiança menor ou igual a uma variação de 1% em relação à média. De qualquer forma, a meta de conhecer o comportamento das perdas para servir de base nas simulações de rede em laboratório, foi alcançada. É notório o comportamento geométrico do tamanho das rajadas, com valores entre 1 e 10 sendo os de maior significância. Também foi observado que a probabilidade de um pacote iniciar uma rajada é da ordem de até poucas dezenas percentuais. Com base nesses valores, pode-se configurar o produto de simulação de redes para aferir-se o comportamento das qualidades de voz de cada codec.

## **9.2. Distribuição do *Jitter***

Como já comentado anteriormente, a forma mais prática e viável de identificar-se o comportamento do *jitter* em redes é através da medição do atraso fim-a-fim gerado com pacotes ICMP, para em seguida ser usada a equação (6.1) indicada no item 6.3. Como forma de enriquecer este trabalho, as medidas de atraso obtidas também são apresentadas.

Da mesma forma que a identificação do perfil de comportamento das perdas em uma rede de pacotes, as medições para estudo do *jitter* foram feitas nos mesmos três enlaces: RJ/AC, RJ/SP e DATAPREV-RJ/UFRJ.

Para identificação do jitter, em 26 dias diferentes, entre os meses de janeiro a abril de 2002, nos horários entre 10:00 e 12:00h e entre 14:00 e 16:00h, foram obtidas 19.244 medidas entre RJ e AC, 19.322 medidas entre RJ e SP e 15.339 medidas pela Internet (entre DATAPREV-RJ e UFRJ).

Os valores de jitter aqui identificados servirão de subsídio para a escolha da faixa de valores que será utilizada, nas medições de qualidade de voz em laboratório.

### 9.2.1. Medidas entre RJ e AC, na rede *Frame Relay* da Previdência Social

#### ATRASO

O atraso médio fim-a-fim medido foi de 264,16 ms, com valores mínimo de 253,00 ms, máximo de 729,00 ms e de desvio padrão 13,43 ms. As medidas feitas garantem um grau de confiança de 99%, com intervalo de confiança entre 293,91 ms e 264,41 ms para o valor médio (variação de 0,25 ms em relação à média). As figuras a seguir ilustram o comportamento do atraso neste enlace.

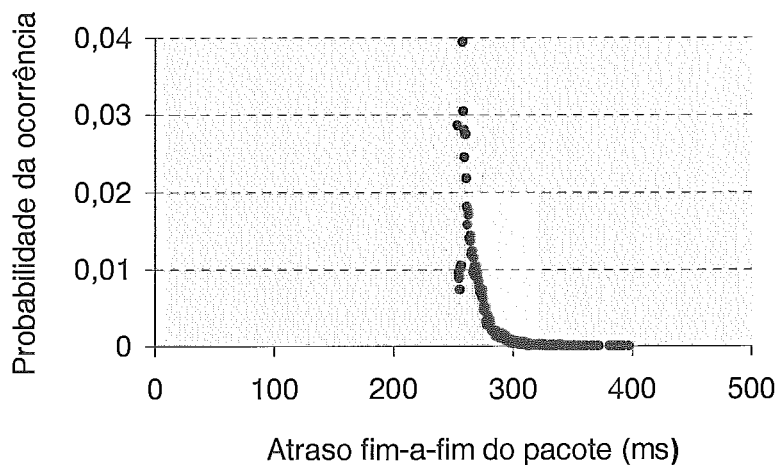


Figura 9.6 – Probabilidade da ocorrência dos valores de atraso entre RJ e AC.

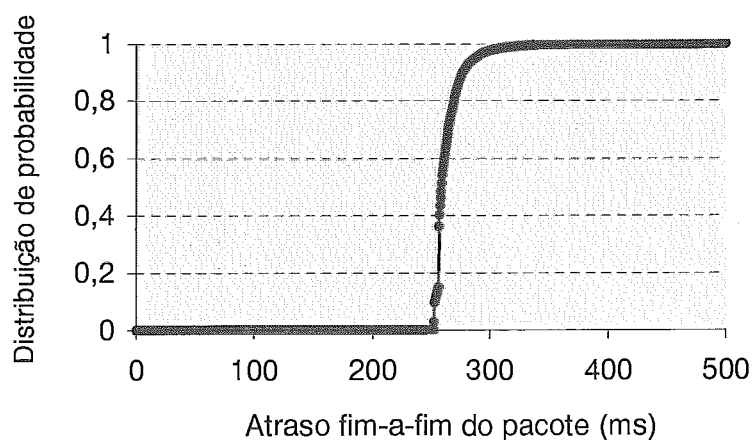


Figura 9.7 – Distribuição de probabilidade acumulada dos valores de atraso entre RJ e AC.

## JITTER

O *jitter* médio calculado foi de 7,00 ms, com valores mínimo de 0,00 ms, máximo de 61,63 ms e de desvio padrão 3,22 ms. As medidas feitas garantem um grau de confiança de 99%, com intervalo de confiança entre 6,94 ms e 7,06 ms para o valor médio (variação de 0,06 ms em relação à média). A figura a seguir ilustra o comportamento do *jitter* neste enlace.

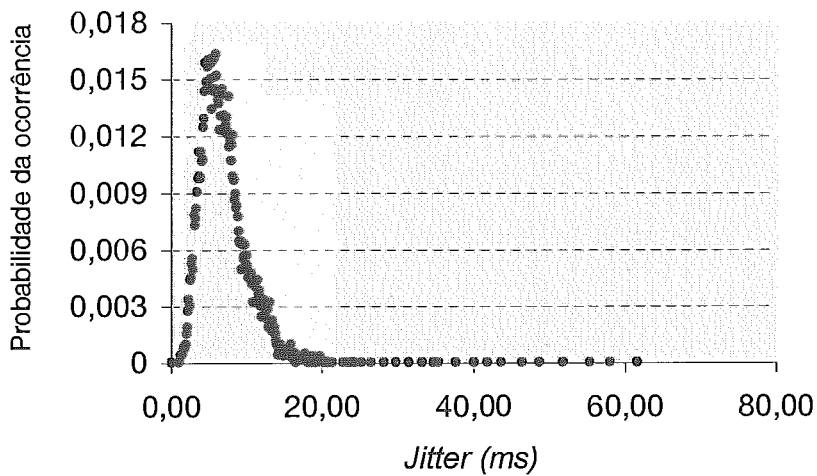


Figura 9.8 – Probabilidade da ocorrência dos valores de *jitter* entre RJ e AC.

### 9.2.2. Medidas entre RJ e SP, na rede *Frame Relay* da Previdência Social

## ATRASSO

O atraso médio fim-a-fim medido foi de 28,19 ms, com valores mínimo de 4,00 ms, máximo de 262,50 ms e de desvio padrão 44,27 ms. As medidas feitas garantem um grau de confiança de 90%, com intervalo de confiança entre 27,66 ms e 28,72 ms para o valor médio (variação de 0,53 ms em relação à média). As figuras a seguir ilustram o comportamento do atraso neste enlace.

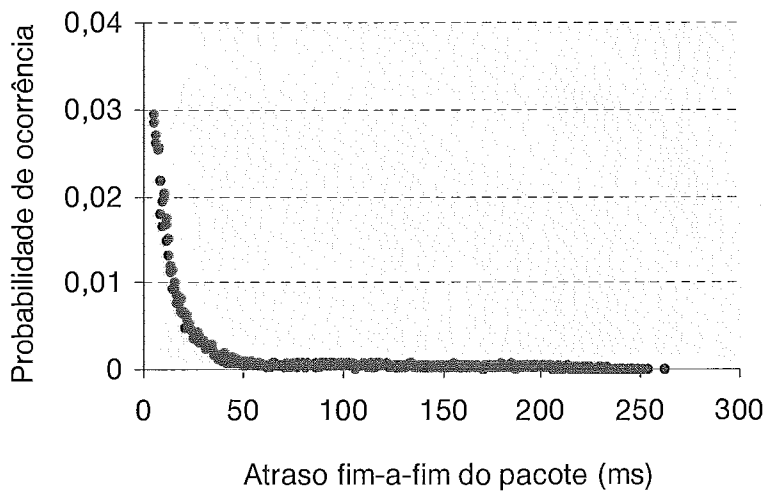


Figura 9.9 – Probabilidade da ocorrência dos valores de atraso entre RJ e SP.

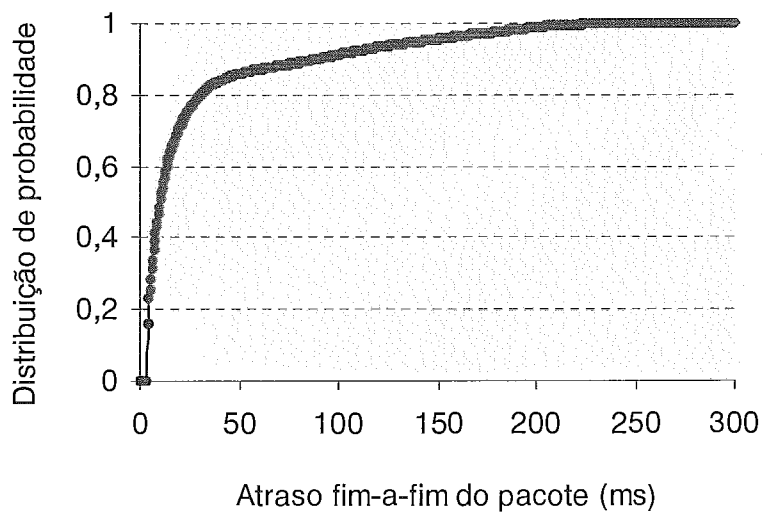


Figura 9.10 – Distribuição de probabilidade acumulada dos valores de atraso entre RJ e SP.

## JITTER

O *jitter* médio calculado foi de 7,59 ms, com valores mínimo de 0,00 ms, máximo de 31,43 ms e de desvio padrão 3,56 ms. As medidas feitas garantem um grau de confiança de 99%, com intervalo de confiança entre 7,52 ms e 7,66 ms para o valor médio (variação de 0,07 ms em relação à média). A figura a seguir ilustra o comportamento do *jitter* neste enlace.

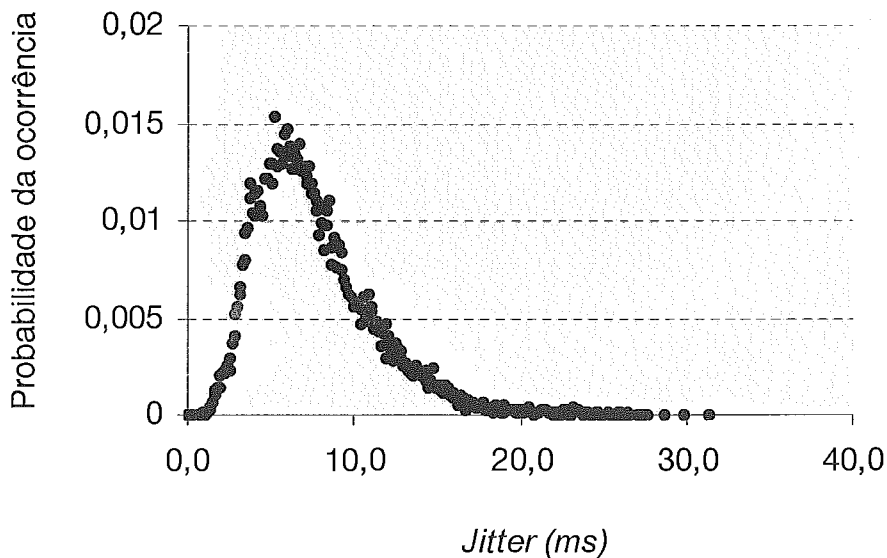


Figura 9.11 – Probabilidade da ocorrência dos valores de *jitter* entre RJ e SP.

### 9.2.3. Medidas entre a DATAPREV (RJ) e a UFRJ, pela Internet

## ATRASSO

O atraso médio fim-a-fim medido foi de 81,70 ms, com valores mínimo de 5,50 ms, máximo de 462,00 ms e de desvio padrão 41,62 ms. As medidas feitas garantem um grau de confiança de 95%, com intervalo de confiança entre 81,04 ms e 82,36 ms para o valor médio (variação de 0,66 ms em relação à média). As figuras a seguir ilustram o comportamento do atraso neste enlace.

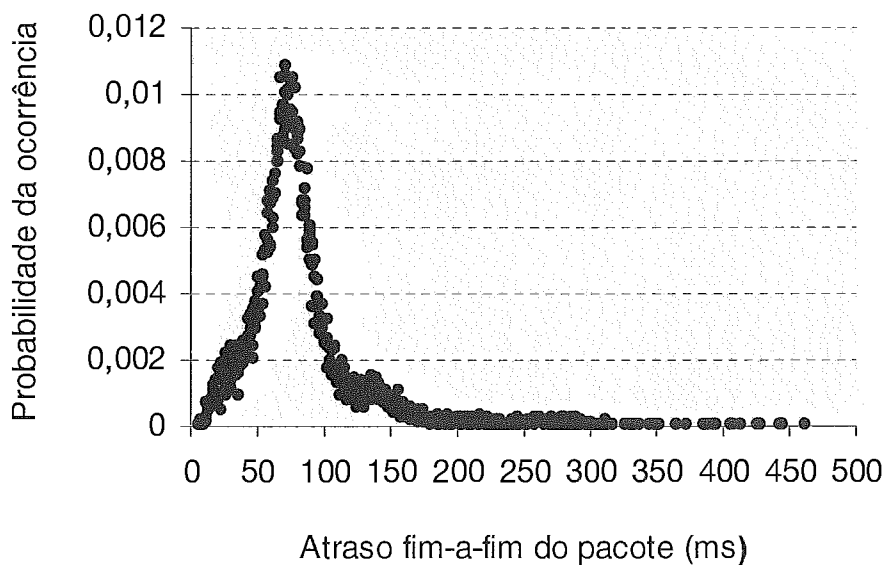


Figura 9.12 – Probabilidade da ocorrência dos valores de atraso entre DATAPREV e UFRJ.

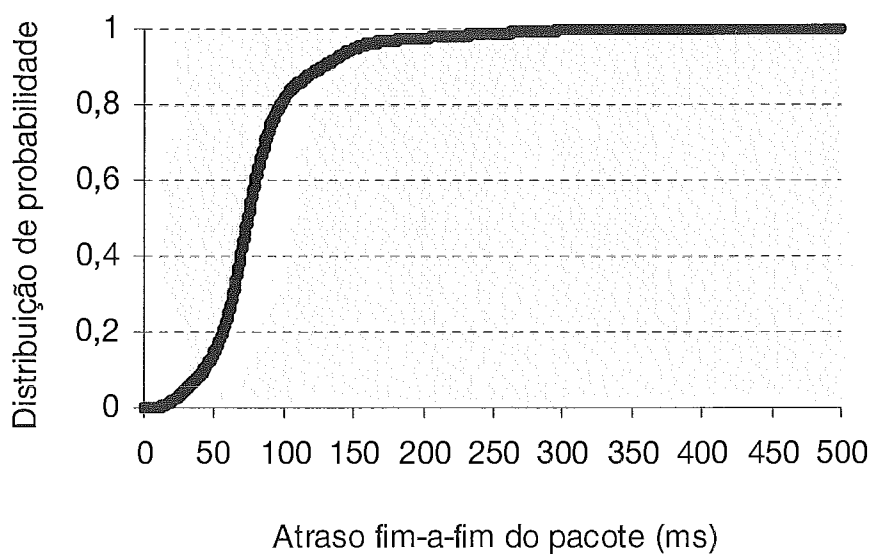


Figura 9.13 – Distribuição de probabilidade acumulada dos valores de atraso entre DATAPREV e UFRJ.

## JITTER

O *jitter* médio calculado foi de 12,15 ms, com valores mínimo de 0,00 ms, máximo de 47,06 ms e de desvio padrão 4,57 ms. As medidas feitas garantem um grau de confiança de 99%, com intervalo de confiança entre 12,05 ms e 12,25 ms para o valor médio (variação de 0,10 ms em relação à média). A figura a seguir ilustra o comportamento do *jitter* neste enlace.

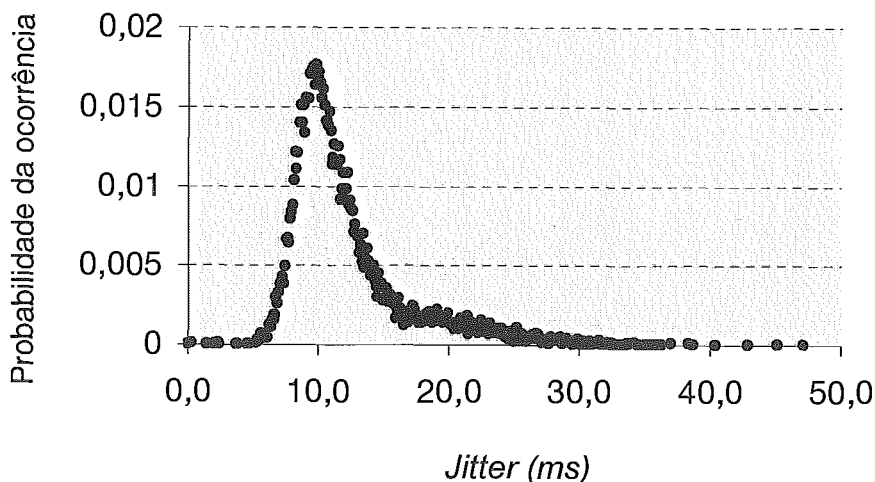


Figura 9.14 – Probabilidade da ocorrência dos valores de *jitter* entre DATAPREV e UFRJ.

### 9.3. Valores PESQ obtidos

Todos os valores PESQ apresentados foram obtidos com um grau de confiança de 95% e buscando-se garantir um intervalo de confiança (IC) máximo de 2,5%. Cada valor PESQ é a média de um múltiplo de 48 medidas realizadas. Conforme já mencionado, o número 48 advém do uso de arquivos wav de quatro vozes diferentes, com doze repetições cada. Como a execução de um conjunto com 48 medidas demanda aproximadamente 15 minutos, a quantidade de medidas para obtenção de um único valor PESQ teve que ser bem controlada, agravado pela necessidade de intervenção manual para início de cada conjunto de medidas. Assim, tentou-se manter como IC o valor máximo de 2,5%, com a menor quantidade de medidas necessárias. Apesar dessa meta, alguns conjuntos de medidas apresentaram IC pouco superior ao perseguido,



como por exemplo os 2,63% da tabela 8. Esses casos, foram limitados pela quantidade de 672 medidas executadas, correspondendo a aproximadamente 3,5 horas de medições.

Fato também a destacar que, quanto menores os valores de *jitter* e perdas considerados, menor o número de medidas necessárias para um IC inferior a 2,5%, já que a diferença entre medidas não é muito grande, isto é, o desvio padrão é pequeno. Em alguns casos, com apenas 48 medidas foi possível observar ICs menores que 1,5% (como exemplo temos os ICs 0,38% na tabela 7 e 1,18% na tabela 9).

Cabe ressaltar que como cada arquivo wav tem duração de 8 segundos, um único conjunto de 48 medidas contém aproximadamente de 12.800 a 38.400 pacotes IP, dependendo do tipo de codec e número de quadros de voz no mesmo pacote. Essas quantidades são significativas para os percentuais de 5 e 10% de perdas utilizadas.

O valor PESQ 4,01, adotado como “valor de referência”, foi obtido com medidas no ambiente de laboratório (figura 8.2), sem perdas e *jitter*, usando-se G711 lei A.

As tabelas e figuras a seguir, apresentam os resultados obtidos nas diversas situações observadas.

### 9.3.1. G.711 lei A

Esse valor será a referência comparativa adotada como a melhor medida passível de ser alcançada, num sistema com voz digitalizada.

Tabela 9.1 – Valor PESQ para G.711 lei A.

Valor de referência	4,01
Desvio padrão	0,05
Intervalo de confiança	0,38%

### 9.3.2. G.729 (10 ms)

Valores obtidos variando-se *jitter* e perdas de pacotes, para G.729, onde cada pacote IP contém um único quadro de voz. Como a formação de cada quadro consome 10 ms, será usada a representação “G729 (10 ms)” para esse caso.

Tabela 9.2 – Valores PESQ para G.729 (10 ms).

jitter		0				10			
pacotes		1	3	6	10	1	3	6	10
prob perda									
0	PESQ	3,53	3,53	3,53	3,53	1,26	1,26	1,26	1,26
	D.P.	0,22	0,22	0,22	0,22	0,21	0,21	0,21	0,21
	Int. conf.	1,76%	1,76%	1,76%	1,76%	2,07%	2,07%	2,07%	2,07%
5	PESQ	3,06	2,12	1,53	1,15	1,21	1,03	0,87	0,75
	D.P.	0,19	0,24	0,22	0,28	0,20	0,22	0,19	0,18
	Int. conf.	1,77%	1,87%	2,35%	2,44%	2,35%	2,45%	2,46%	2,43%
10	PESQ	2,71	1,59	0,94	0,71	1,14	0,83	0,71	0,68
	D.P.	0,24	0,22	0,23	0,22	0,21	0,19	0,21	0,24
	Int. conf.	1,80%	2,26%	2,43%	2,43%	2,34%	2,37%	2,42%	2,63%

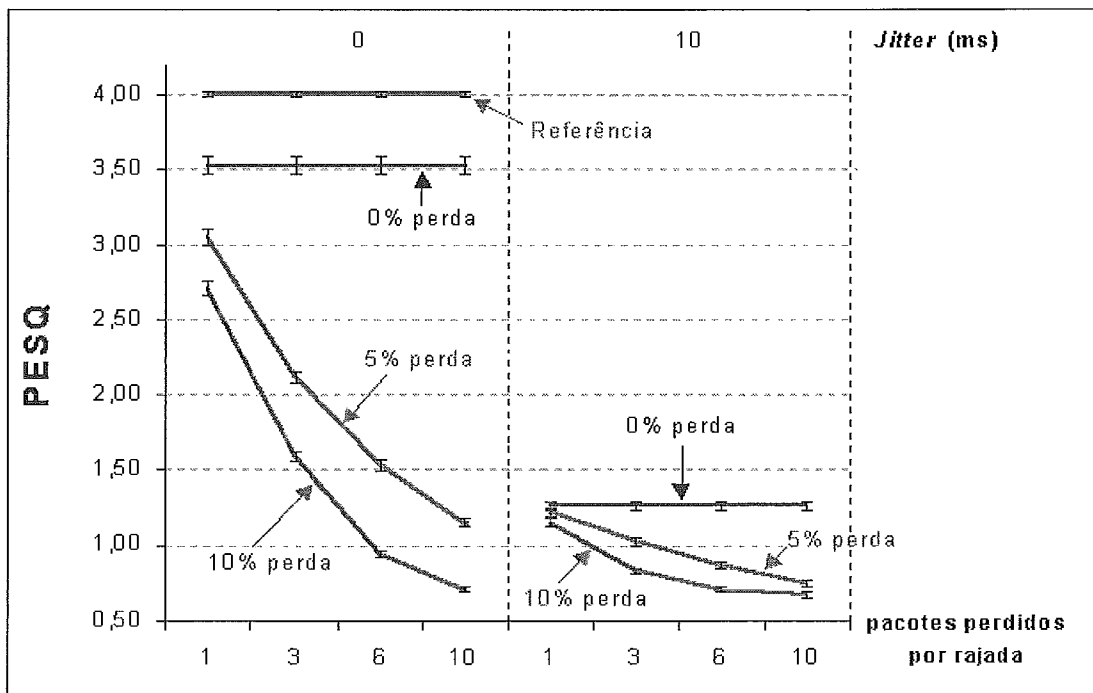


Figura 9.15 – Valores PESQ para medições com G.729 (10 ms).

### 9.3.3. G.729 (20 ms)

Valores obtidos variando-se *jitter* e perdas de pacotes, para G.729, onde cada pacote IP contém dois quadros de voz. Como a formação de cada quadro consome 20 ms, será usada a representação “G729 (20 ms)” para esse caso.

Tabela 9.3 – Valores PESQ para G.729 (20 ms).

jitter		0			
pacotes		1	3	6	10
prob perda					
0	PESQ	3,53	3,53	3,53	3,53
	D.P.	0,15	0,15	0,15	0,15
	Int. conf.	1,18%	1,18%	1,18%	1,18%
5	PESQ	2,89	2,25	1,68	1,28
	D.P.	0,26	0,34	0,30	0,33
	Int. conf.	2,50%	2,48%	2,25%	2,45%
10	PESQ	2,53	1,79	1,24	0,82
	D.P.	0,26	0,35	0,26	0,27
	Int. conf.	2,07%	2,47%	2,41%	2,47%

Legenda: Int. conf. - intervalo de confiança  
D.P. - desvio padrão

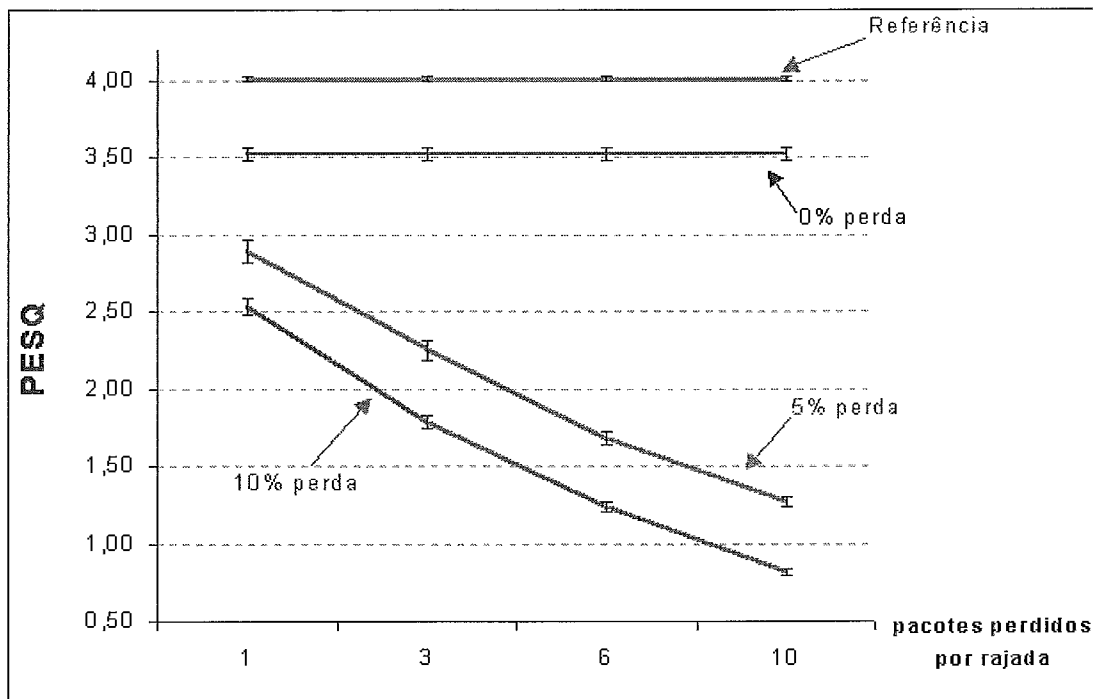


Figura 9.16 – Valores PESQ para medições com G.729 (20 ms).

### 9.3.4. G.729 (30 ms)

Valores obtidos variando-se *jitter* e perdas de pacotes, para G.729, onde cada pacote IP contém três quadros de voz. Como a formação de cada quadro consome 30 ms, será usada a representação “G729 (30 ms)” para esse caso.

Tabela 9.4 – Valores PESQ para G.729 (30 ms).

jitter		0			
pacotes		1	3	6	10
prob perda					
0	PESQ	3,54	3,54	3,54	3,54
	D.P.	0,14	0,14	0,14	0,14
	Int. conf.	1,14%	1,14%	1,14%	1,14%
5	PESQ	2,83	2,09	1,76	1,50
	D.P.	0,27	0,28	0,46	0,50
	Int. conf.	1,90%	2,17%	2,48%	2,52%
10	PESQ	2,40	1,51	1,13	0,87
	D.P.	0,25	0,27	0,33	0,31
	Int. conf.	2,07%	2,29%	2,50%	2,74%

Legenda: Int. conf. - intervalo de confiança  
D.P. - desvio padrão

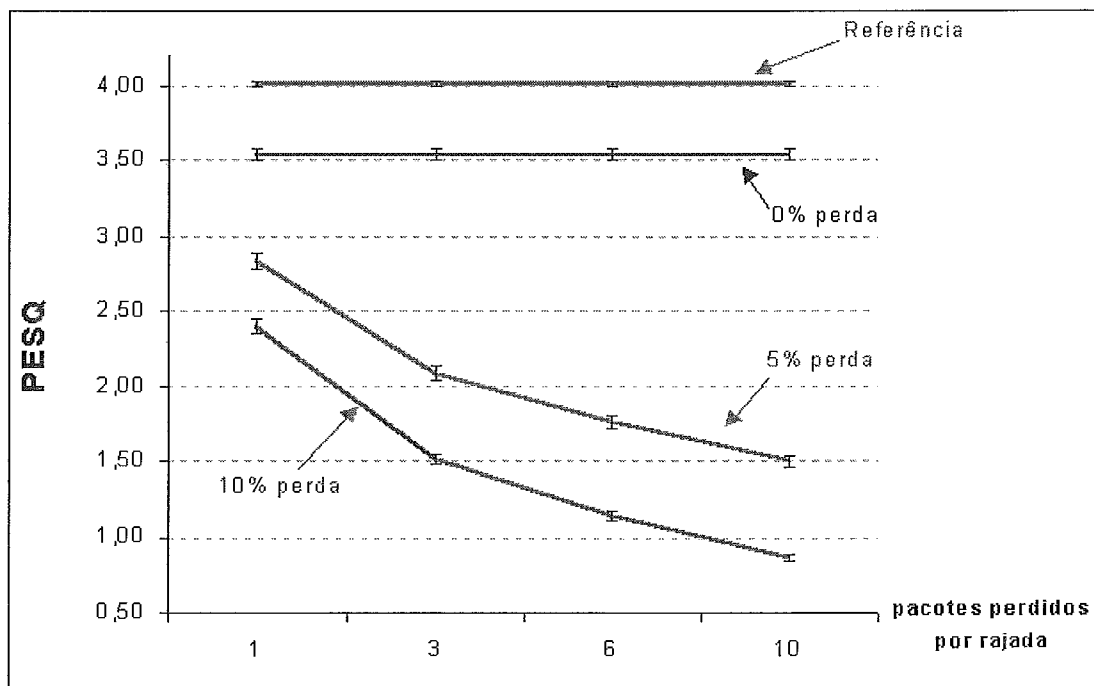


Figura 9.17 – Valores PESQ para medições com G.729 (30 ms).

### 9.3.5. G.723.1 (6,3 Kbps)

Valores obtidos variando-se *jitter* e perdas de pacotes, para G.723.1, com MP-MLQ. Será usada a representação “G723.1 (6,3 Kbps)” para esse caso, devido ao consumo de banda característico desse codec.

Tabela 9.5 – Valores PESQ para G.723.1 (6,3 Kbps).

jitter		0			
pacotes		1	3	6	10
prob perda					
0	PESQ	3,41	3,41	3,41	3,41
	D.P.	0,18	0,18	0,18	0,18
	Int. conf.	1,53%	1,53%	1,53%	1,53%
5	PESQ	2,96	2,09	1,68	1,33
	D.P.	0,24	0,32	0,41	0,43
	Int. conf.	1,59%	2,49%	2,45%	2,42%
10	PESQ	2,52	1,47	1,05	0,84
	D.P.	0,28	0,26	0,28	0,31
	Int. conf.	2,22%	2,50%	2,51%	2,83%

Legenda: Int. conf. - intervalo de confiança  
D.P. - desvio padrão

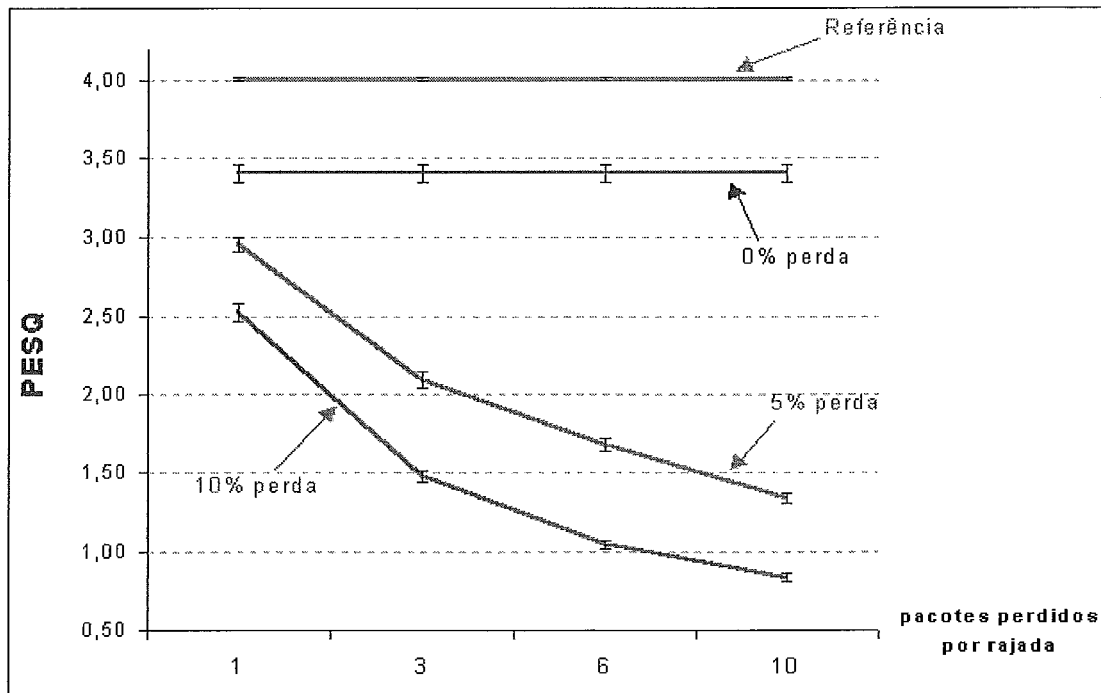


Figura 9.18 – Valores PESQ para medições com G.723.1 (6,3 Kbps).

### 9.3.6. G.723.1 (5,3 Kbps)

Valores obtidos variando-se *jitter* e perdas de pacotes, para G.723.1, com ACELP. Será usada a representação “G723.1 (5,3 Kbps)” para esse caso, devido ao consumo de banda característico desse codec.

Tabela 9.6 – Valores PESQ para G.723.1 (5,3 Kbps).

jitter		0			
pacotes		1	3	6	10
prob perda					
0	PESQ	3,27	3,27	3,27	3,27
	D.P.	0,20	0,20	0,20	0,20
	Int. conf.	1,70%	1,70%	1,70%	1,70%
5	PESQ	2,87	1,95	1,67	1,32
	D.P.	0,22	0,30	0,33	0,35
	Int. conf.	2,15%	2,20%	2,29%	2,50%
10	PESQ	2,56	1,50	1,04	0,78
	D.P.	0,24	0,33	0,29	0,25
	Int. conf.	1,89%	2,35%	2,41%	2,47%

Legenda: Int. conf. - intervalo de confiança  
D.P. - desvio padrão

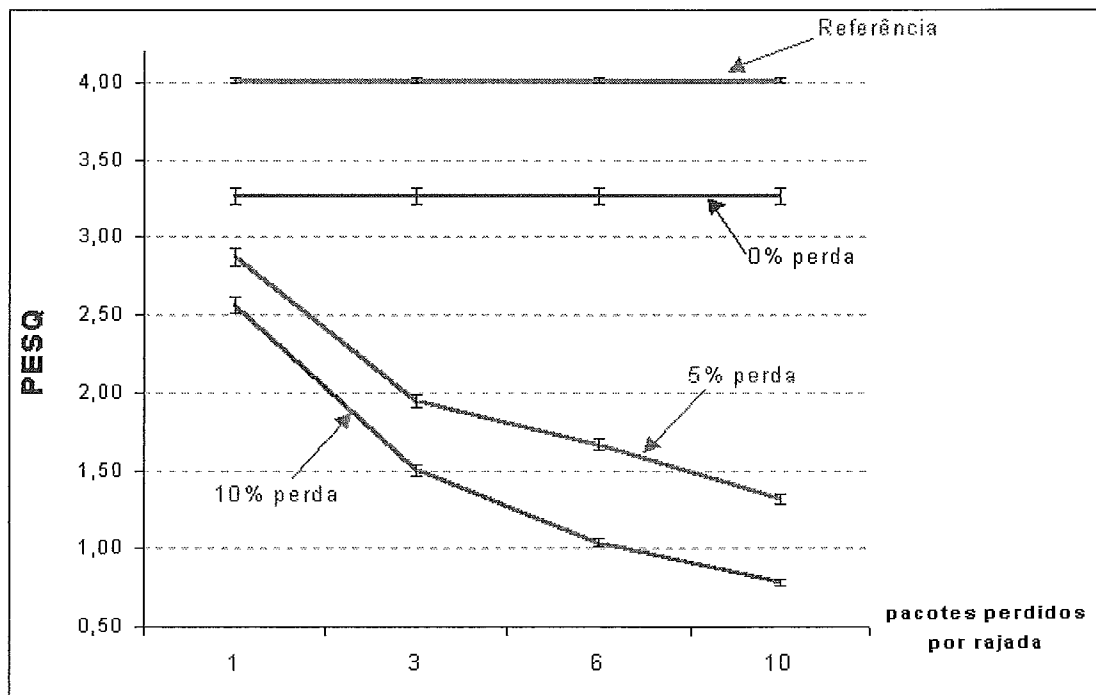


Figura 9.19 – Valores PESQ para medições com G.723.1 (5,3 Kbps).

### 9.3.7. Comparativo entre codecs

#### JITTER VARIÁVEL E SEM PERDA

Tabela 9.7 – Valores PESQ dos codecs, variando-se o *jitter*.

Probabilidade de perda = 0%				
CODEC	Jitter			
	0	10	20	30
G.723.1 (5,3)	3,27	3,28	3,28	0,98
Intervalo de confiança	1,70%	1,44%	1,51%	2,47%
G.723.1 (6,3)	3,41	3,40	3,42	1,02
Intervalo de confiança	1,53%	1,49%	1,39%	2,44%
G.729 (30ms)	3,54	0,96	0,53	
Intervalo de confiança	1,14%	2,49%	2,98%	
G.729 (20ms)	3,53	1,09	0,54	
Intervalo de confiança	1,18%	2,48%	3,23%	
G.729 (10ms)	3,53	1,26	0,53	
Intervalo de confiança	1,76%	2,07%	3,18%	

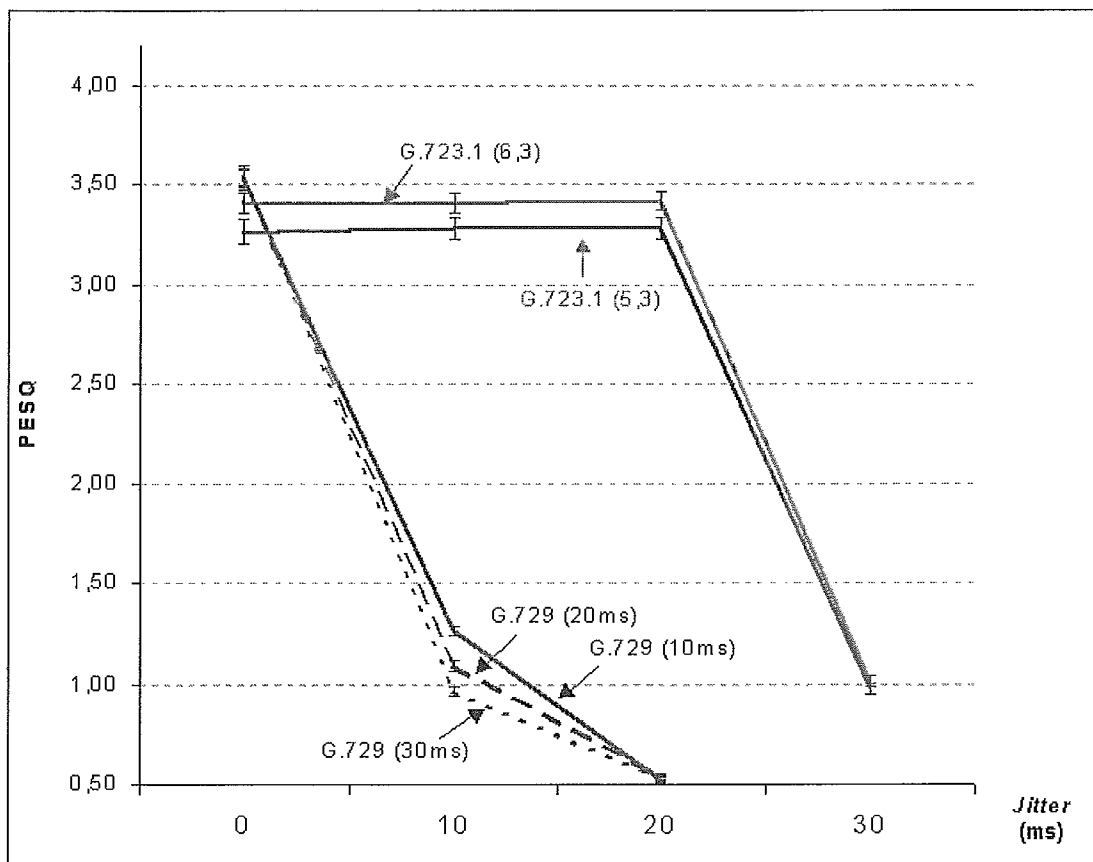


Figura 9.20 – Valores PESQ para comparação entre as opções estudadas, considerando não existir perda de pacote.

PERDA VARIÁVEL E SEM JITTER

Tabela 9.8 – Valores PESQ dos codecs, variando-se a perda.

Probabilidade de perda = 5%					
CODEC	Tamanho da perda (pacotes)				
	0	1	3	6	10
<b>G.723.1 (5,3)</b>	<b>3,27</b>	<b>2,87</b>	<b>1,95</b>	<b>1,67</b>	<b>1,32</b>
Intervalo de confiança	1,70%	2,15%	2,20%	2,29%	2,50%
<b>G.723.1 (6,3)</b>	<b>3,41</b>	<b>2,96</b>	<b>2,09</b>	<b>1,68</b>	<b>1,33</b>
Intervalo de confiança	1,53%	1,59%	2,49%	2,45%	2,42%
<b>G.729 (30ms)</b>	<b>3,54</b>	<b>2,83</b>	<b>2,09</b>	<b>1,76</b>	<b>1,50</b>
Intervalo de confiança	1,14%	1,90%	2,17%	2,48%	2,52%
<b>G.729 (20ms)</b>	<b>3,53</b>	<b>2,89</b>	<b>2,25</b>	<b>1,68</b>	<b>1,28</b>
Intervalo de confiança	1,18%	2,50%	2,48%	2,25%	2,45%
<b>G.729 (10ms)</b>	<b>3,53</b>	<b>3,06</b>	<b>2,12</b>	<b>1,53</b>	<b>1,15</b>
Intervalo de confiança	1,76%	1,77%	1,87%	2,35%	2,44%

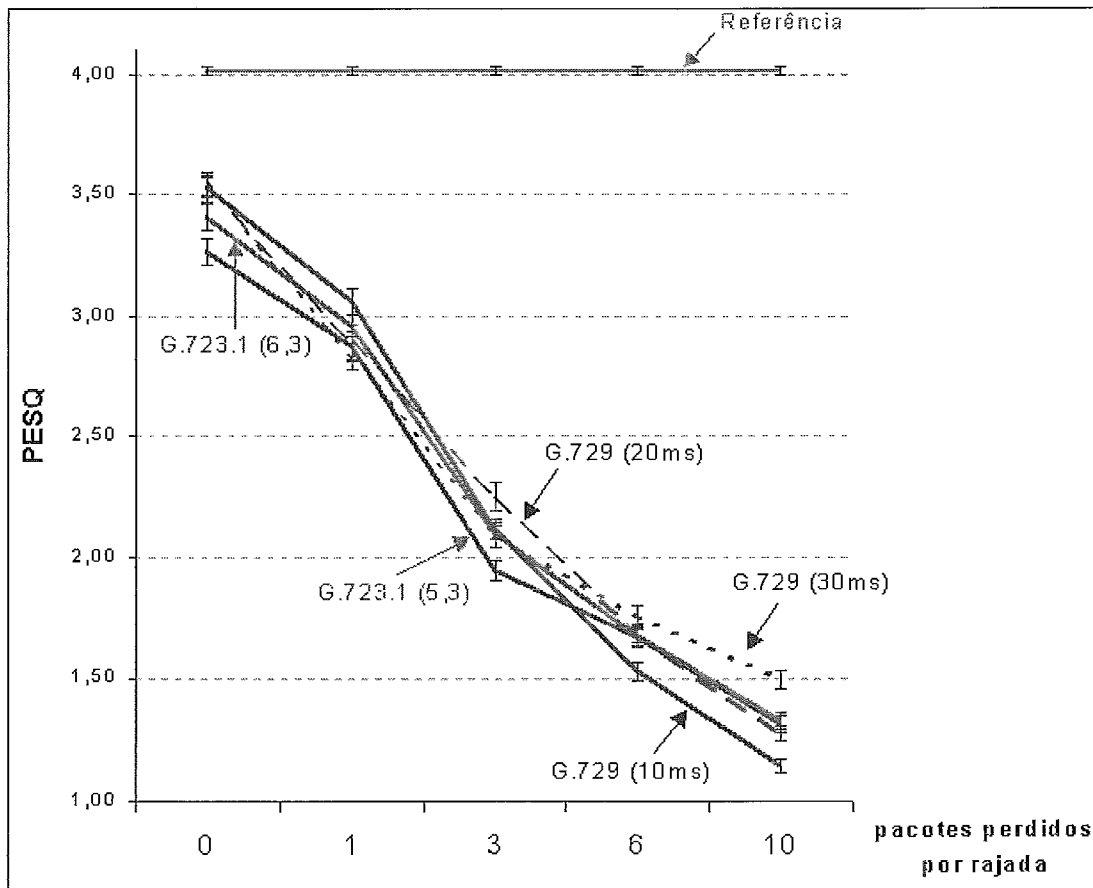


Figura 9.21 – Valores PESQ para comparação entre as opções estudadas, considerando não existir jitter.



Nesta comparação dos comportamentos dos codecs, foi observada uma interessante característica quando variado o número de quadros G.729 em um mesmo pacote IP. Numa avaliação precipitada, era de se esperar que, quanto maior a perda, pior seria a qualidade da fala regenerada para pacotes com mais quadros de voz em seu interior. A figura 9.22 é uma variação da figura 9.21, onde são apresentadas apenas as variações do G.729, para facilitar a visualização do exposto. Neste caso, pode ser constatado que a afirmativa anterior não se concretiza. Na verdade, quanto maior o número de quadros de voz no mesmo pacote, melhor a qualidade quando se aumenta a quantidade de perdas, como é claramente mostrado no caso de rajadas de perdas com 10 pacotes IP.

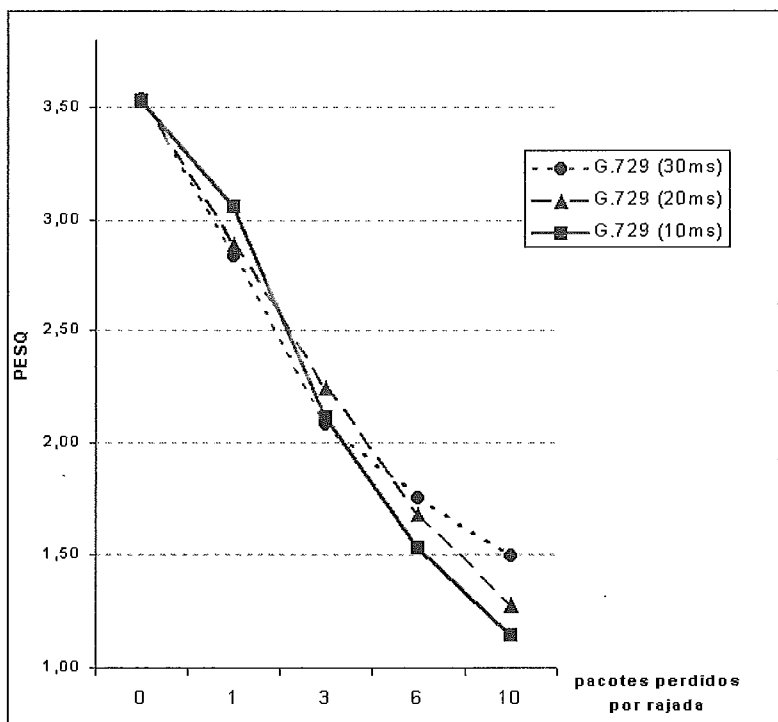


Figura 9.22 – Observação do comportamento da qualidade de voz, segundo o número de quadros G.729 no mesmo pacote IP.

A explicação para o fato é simples. Considerando-se como exemplo 9.000 pacotes IP, com um quadro de voz cada, para uma perda de 5% com rajadas de 10 pacotes, existirão 450 rajadas em média, com 10 pacotes cada. Distribuindo-se

homogeneamente o restante dos pacotes não perdidos entre cada rajada, tem-se 10 pacotes, em média, e por conseguinte 10 quadros de voz entre perdas. Já para o caso de ter-se 3 quadros de voz em cada pacote IP, não mais teremos 9.000 pacotes, mas sim apenas 3.000, para o mesmo trecho de voz considerado anteriormente. Para os mesmos 5% de perdas e 10 pacotes por rajada, existirão 150 rajadas em média. Distribuindo-se homogeneamente o restante dos pacotes não perdidos entre cada rajada, tem-se os mesmos 10 pacotes, em média, mas agora com 30 quadros de voz entre perdas. É certo que a quantidade de quadros perdidos por rajada também aumenta no segundo exemplo, mas é mais danoso ter-se um sinal com mais interrupções de curta duração do que um com menos interrupções um pouco mais longas. Isto ocorre porque poucos décimos de segundos guardam apenas um fonema, onde pacotes vizinhos carregam informações sonoras muito próximas. Assim, o número de fonemas alterados no primeiro caso é muito maior do que no segundo.

Contudo, para o caso de rajadas muito pequenas, como é o caso de perdas com apenas um pacote, a alteração do fonema no caso do G.729 (10 ms) é menor que a dos outros dois. Pode-se dizer que, para o caso de perdas equivalentes a algumas dezenas até centenas de milissegundos, o fonema é muito alterado, influenciando bastante na regeneração da voz. Mas para poucas unidades, até uma dezena de milissegundos, essa alteração não “deforma” tanto o fonema, refletindo em melhor qualidade PESQ.

Existe então um limiar onde o tamanho da “alteração” fonética irá interferir menos na qualidade PESQ do que o número de suas ocorrências.

### 9.3.8. Comparativo entre RJ/AC e RJ/SP

Aqui, buscou-se estudar em laboratório as duas situações práticas apontadas nos itens 9.2.1 e 9.2.2. Como o valor PESQ não é influenciado com o atraso fim-a-fim, o caso RJ/AC foi simulado com a variação do atraso modelada por uma exponencial de média 11,2 ms (valor médio de 264,16 ms, menos o menor valor observado de 253,00 ms). Cabe registrar que foi necessário arredondar o valor do atraso com uma casa decimal, por restrição do programa de emulação de rede usado. O caso RJ/SP foi simulado com a variação do atraso modelada por uma exponencial de média 28,2 ms (valor médio de 28,19 ms, menos o menor valor observado de 4,00 ms).

Tabela 9.9 – Valores PESQ dos codecs, para RJ/AC e RJ/SP.

Probabilidade de perda = 0%		
CODEC	<i>média da exp</i>	
	11,2	24,2
<b>G.723.1 (5,3)</b>	<b>3,21</b>	<b>2,01</b>
Intervalo de confiança	1,48%	2,34%
<b>G.723.1 (6,3)</b>	<b>3,23</b>	<b>2,01</b>
Intervalo de confiança	2,08%	2,17%
<b>G.729 (30ms)</b>	<b>0,92</b>	<b>0,56</b>
Intervalo de confiança	2,40%	2,85%
<b>G.729 (20ms)</b>	<b>0,91</b>	<b>0,55</b>
Intervalo de confiança	2,24%	3,12%
<b>G.729 (10ms)</b>	<b>0,87</b>	<b>0,53</b>
Intervalo de confiança	2,29%	3,47%

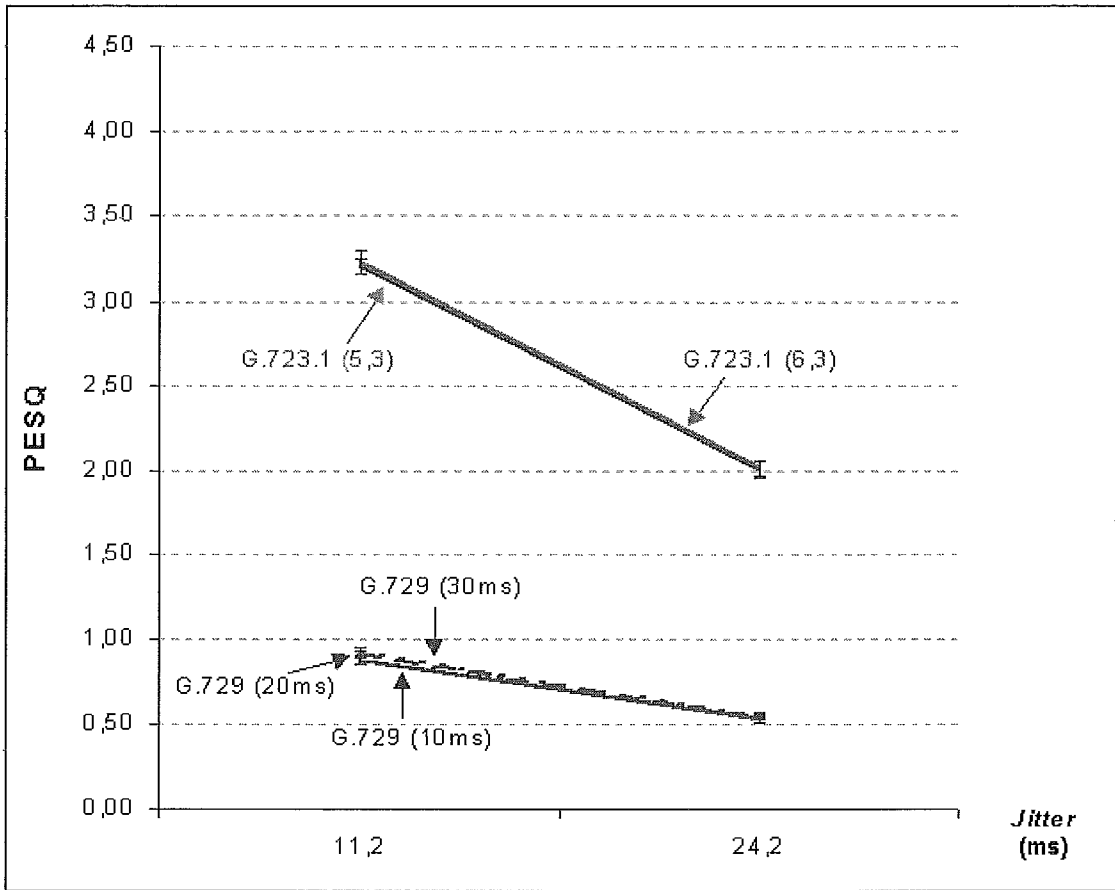


Figura 9.23 – Observação do comportamento da qualidade de voz, nos enlaces *Frame Relay* entre RJ/AC e RJ/SP.

### 9.3.9. Comparativo entre G.729 e G.729A

Com a intenção de comprovar a similaridade da qualidade entre G.729 e G.729A (G.729 Anexo A), foram comparados valores PESQ com *jitter* 0 e 10 ms, sem perdas de pacotes. Também foram comparados valores referentes a 5% de probabilidade de perda de pacotes, sem *jitter*, variando o tamanho das rajadas entre 1, 3, 6 e 10 pacotes IP.

Tabela 9.10 – Valores PESQ comparativos entre G.729 e G.729A.

COEDC	jitter		0				10			
	pacotes	prob perda	1	3	6	10	1	3	6	10
G.729A	0	PESQ	3,54	3,54	3,54	3,54	1,23	1,23	1,23	1,23
		D.P.	0,13	0,13	0,13	0,13	0,18	0,18	0,18	0,18
		Int. conf.	1,07%	1,07%	1,07%	1,07%	1,95%	1,95%	1,95%	1,95%
	5	PESQ	3,01	2,09	1,49	1,13				
		D.P.	0,22	0,23	0,26	0,28				
		Int. conf.	2,07%	1,80%	2,49%	2,48%				
G.729	0	PESQ	3,53	3,53	3,53	3,53	1,26	1,26	1,26	1,26
		D.P.	0,22	0,22	0,22	0,22	0,21	0,21	0,21	0,21
		Int. conf.	1,76%	1,76%	1,76%	1,76%	2,07%	2,07%	2,07%	2,07%
	5	PESQ	3,06	2,12	1,53	1,15				
		D.P.	0,19	0,24	0,22	0,28				
		Int. conf.	1,77%	1,87%	2,35%	2,44%				

Legenda: Int. conf. - intervalo de confiança  
D.P. - desvio padrão

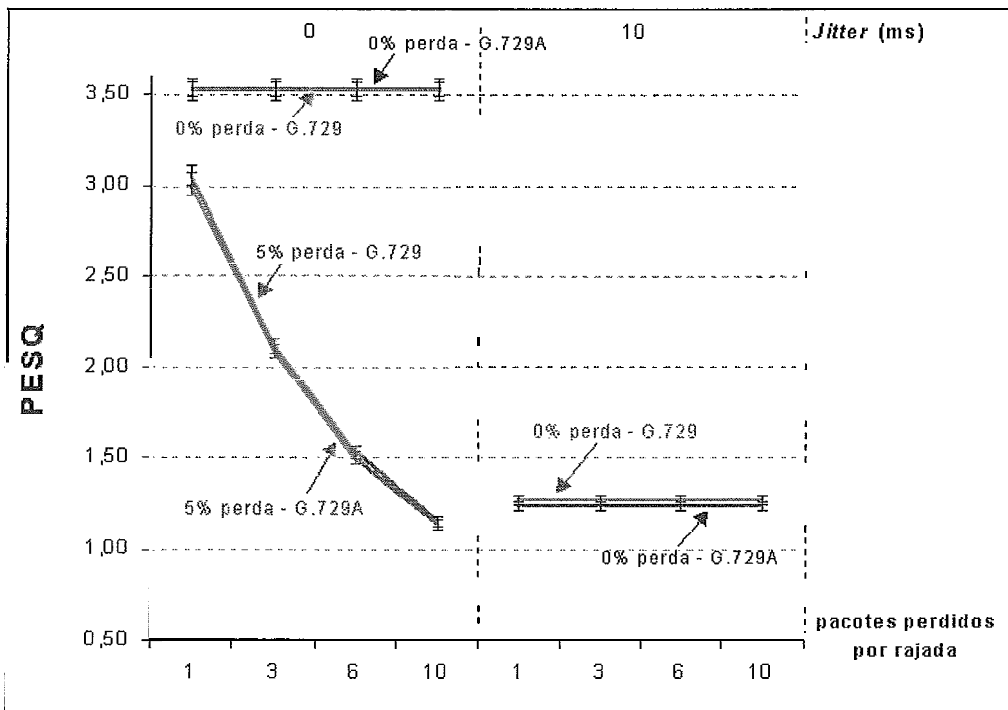


Figura 9.24 – Valores PESQ para comparação entre G.729 e G.729A.

## 10. CONCLUSÕES

As informações apresentadas no item 9.3, serviram como base para as diversas conclusões aqui enumeradas. Em aderência ao objetivo inicial, este trabalho contribuiu com várias informações úteis a projetos práticos, além de trazer consigo importantes apontamentos de cunho acadêmico, caracterizando o correto funcionamento dos codificadores estudados, bem como o emprego ideal de cada um deles.

A seguir, são listadas as conclusões resultantes deste estudo. Em prosseguimento, é feito um resumo com as principais idéias, destacando-se os pontos de maior relevância. Por fim, importantes fatos observados durante o processo de medição foram registrados, agregando conhecimento ao tema.

### Conclusões:

- a) Segundo o gráfico da figura 9.20, os codecs da G.723.1 são menos sensíveis ao *jitter* que o G.729. Isto é, para situações onde não seja possível controlar o *jitter* de forma eficaz, o G.723.1 apresentará uma qualidade melhor que o G.729.
- b) Segundo o mesmo gráfico da figura 9.20, para situações onde a rede apresente variações no atraso que reflita em valores de *jitter* de até 20ms, esta característica não será observada pelos ouvintes, caso seja usado o G.723.1.
- c) Observando os gráficos das figuras 9.15, 9.16, 9.17 e 9.20, no caso de 10% de perdas com tamanho de 3 pacotes, o G.729 apresenta valores PESQ superiores aos do *jitter* de 10 ms. No caso do G.729 com 20 e 30 ms, a perda pode apresentar tamanho de até 6 pacotes, que ainda assim a qualidade é superior ao *jitter* de 10ms, demonstrando ser muito sensível ao *jitter*.
- d) Segundo o gráfico da figura 9.21, para o caso de uma rede com comportamento muito próximo a inexistência de perdas ou *jitter*, o G.729 apresenta qualidade superior ao G.723.1 de 6,3 Kbps, que por sua vez é melhor que o G.723.1 de 5,3 Kbps.
- e) Ainda pelo mesmo gráfico, apesar de na prática não representar uma observação relevante, já que apresentam valores PESQ inferiores a 2, para 5% de perdas com tamanho de 10 pacotes, a qualidade do G.729(30 ms) é superior ao do G.729(20 ms) que é superior ao do G.729(10 ms). Este fato é detalhadamente

explicado no item 9.3.7 e representa um interessante comportamento da qualidade de voz.

- f) O gráfico da figura 9.23 apresenta as situações de *jitter* observadas na rede *Frame Relay* da DATAPREV entre RJ-SP e RJ-AC, conforme itens 9.1 e 9.2 desta tese. Em ambos os casos, considerando que a quantidade de pacotes perdidos seja desprezível, a qualidade do G.729 é muito inferior ao do G.723.1. Vale alertar que não está sendo considerada a opção de uso do *buffer* de *jitter* para tratamento do mesmo.
- g) O gráfico da figura 9.24 mostra que as observações feitas para o G.729, também são válidas para o G.729A.

As conclusões acima enumeradas podem ser resumidas nos tópicos a seguir:

- O *jitter* é muito mais agressivo à qualidade de voz, do que a ocorrência de perdas de pacotes;
- No caso de enlaces com *jitter* apresentando valores até 20 ms, o G723.1 não sofre influência e tem uma qualidade melhor do que o G.729;
- No caso de enlaces onde o *jitter* não seja notado pela aplicação de voz, o G729 apresenta qualidade superior ao G.723.1;
- G.729 e o G.729A apresentam comportamentos idênticos, com relação à resposta de qualidade, em função da presença de *jitter* e perdas de pacotes;
- Para casos onde o *jitter* e as perdas de pacotes sejam desprezíveis, o G.729 tem qualidade superior ao G.723.1 com 6,3 Kbps, que por sua vez é superior ao G.723.1 com 5,3 Kbps.

Apesar de não serem classificados como conclusões do trabalho, os dois itens a seguir estão aqui registrados por significarem observações importantes do processo de medição.

- Durante o período de montagem do ambiente que seria usado para as medições, foi observado que o valor PESQ obtido variou em função do ganho de sinal configurado na porta analógica (FXS) do *gateway* de voz, o que não era o esperado, segundo a Recomendação P.862. Por conta disso, os *gateways* foram configurados com ganho de 5dB, conforme registrado no item 8.3.3, valor esse que apresentou maior pontuação PESQ das medidas usadas nesses procedimentos iniciais.
- Por mais críticos que fossem os valores configurados de perda de pacotes e *jitter*, o menor valor PESQ obtido em todas as medições executadas foi de -0,1.
- Para situações sem perdas e com *jitter* de 5 ms, os valores PESQ medidos para as três opções do G.729 (10, 20 e 30 ms) foram extremamente similares aos obtidos para os casos sem *jitter*, isto é, valor de *jitter* igual a 0ms. Esta informação não foi registrada no item 9.3, mas foi garantido o grau de confiança de 95% com um intervalo de confiança menor que 2%.



## 11. SUGESTÕES PARA TRABALHOS FUTUROS

Adotando uma abordagem mais ampla, sem o direcionamento único de qualidade do serviço de voz sobre IP, são enumeradas propostas para futuros trabalhos acadêmicos pertinentes ao assunto.

1. Identificação da acurácia do método para medição objetiva de qualidade de voz, descrito na Recomendação P.862 do ITU-T.

Neste trabalho foram comentados vários métodos para medição de qualidade de voz, informando suas cronologias e características funcionais. O mais atual, e teoricamente, mais indicado método é o descrito na Recomendação P.862, já que é dito trazer consigo a independência de fatores subjetivos e os aspectos positivos de outros métodos. Contudo, não foram encontradas nos documentos difundidos pelo ITU-T ou em outras fontes, informações claras e objetivas sobre o quão acurado é o valor resultante das medidas que se valem deste método e como variam segundo as diferenças entre os sinais original e degradado. Com a expectativa de uso desse método em larga escala, em substituição a todos os até então empregados, esse estudo será de grande valia para sua adoção baseada em critérios claros e objetivos, em detrimento dos vários outros métodos já usados.

2. Comparativo entre as pontuações obtidas pelos métodos MOS, PSQM, PAMS e PESQ.

Até hoje, muitos trabalhos que necessitam mensurar a qualidade de voz para identificação da degradação causada por um codec ou sistema, ainda fazem uso do método MOS, devido ao grande número de documentos já publicados, facilitando o comparativo com outras experiências. O PSQM e o PAMS também tiveram uso em alguns estudos, mesmo que em número bem menor que o MOS. O PESQ, por se tratar do método objetivo mais atual, deverá ser a próxima referência como procedimento de medida de qualidade de voz. Assim, uma comparação detalhada dos quatro métodos e comparativos práticos de suas pontuações, será de grande valia para os profissionais envolvidos com este assunto.

### 3. Uso do PESQ em enlaces reais.

Para este trabalho, observações em laboratório foram suficientes para seu desfecho com todas as conclusões bem fundamentadas. Entretanto, um conjunto mais amplo de medidas em enlaces reais, usando-se PESQ, em conjunto com a identificação do comportamento das perdas de pacotes e *jitter* de cada enlace, traz uma visão complementar ao que foi apresentado neste estudo.

### 4. Garantia de comunicação segura com VoIP.

A grande maioria das implementações de redes de comunicação de dados não possui mecanismos de criptografia. Com isso, o uso de voz sobre IP sem uso de qualquer outra funcionalidade adicional, permite que se armazene e reproduza conversas inteiras, utilizando-se equipamentos compatíveis para captura e análise de dados, disponíveis no mercado. O estudo de alternativas para uma conversação telefônica IP segura é bastante útil, como exercício acadêmico e como subsídio para implementações dirigidas a ambientes com maior rigor de sigilo.

## 12. BIBLIOGRAFIA

- [1] SCHULZRINNE, H., "Internet Telephony", *IEEE Network*, v.13, n. 3, pp. 6-7, May/June 1999.
- [2] COX, R. V., "Three New Speech Coders from the ITU Cover a Range of Application", *IEEE Communication Magazine*, pp. 40, Sept. 1997.
- [3] ALENCAR, M.S., *Telefonia Digital*. 1 ed., São Paulo, Brasil, Érica, 1998.
- [4] DOVROLIS C. e RAMANATHAN P., "A Case for Relative Differentiated Services and the Proportional Differentiation Model", *IEEE Network*, pp.26-34, Sept./Oct. 1999.
- [5] COMER, D. E., "Internet Protocol: Connectionless Datagram Delivery". In: Apt, A., Zucker, I. (eds), *Internetworking with TCP/IP Vol I: Principles, Protocols, and Architecture*, 3 ed., chapter 7, New Jersey, USA, Prentice Hall Inc., 1995.
- [6] NICHOLS, K., BLAKE, S., BAKER, F., et al., "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Header", IETF RFC 2474, Dec. 1998.
- [7] ZHANG, L., DEERING, S., ESTRIN, D., et al., "RSVP: A New Resource ReSerVation Protocol", *IEEE Network*, pp. 8-18, Sept. 1993.
- [8] BRADEN, R., ZHANG, L., BERSON, S., et al., "Resource ReSerVation Protocol (RSVP)", IETF RFC 2205, Sept. 1997.
- [9] METZ, C., "RSVP: General-Purpose Signaling for IP", *IEEE Internet Computing*, pp.95-99, May/June 1999.
- [10] ITU-T Rec. H.323, "Packed-based multimedia communications systems", Nov. 2000.
- [11] THOM, G. A., "H.323: The Multimedia Communications Standard for Local Area Networks", *IEEE Communication Magazine*, pp. 52-56, Dec. 1996.
- [12] <http://www.packetizer.com/iptel/h323>.
- [13] SCHULZRINNE, H., ROSENBERG, J., "The IETF Internet Telephony Architecture and Protocols", *IEEE Network*, pp. 18-23, May/June 1999.

- [14] HANDLEY, M., SCHULZRINNE, H., SCHOOLER, E., et al., "SIP: Session Initiation Protocol", IETF RFC 2543, Jan. 1996.
- [15] SCHULZRINNE, H., ROSENBERG, J., "A Comparison of SIP and H.323 for Internet Telephony", <http://www.nossdav.org/1998/papers/nossdav98-043.ps.gz>, *Proc. NOSSDAV*, Cambridge, U.K., July 1998.
- [16] DALGIC, I., FANG, H., "Comparison of H.323 and SIP for IP Telephony Signaling", [http://www.cs.columbia.edu/~hgs/papers/others/Dalg9909\\_Comparison.pdf](http://www.cs.columbia.edu/~hgs/papers/others/Dalg9909_Comparison.pdf), *Proc. of Photonics East, Boston, Massachusetts*, Sept. 1999.
- [17] SCHULZRINNE, H., CASNER, S., FREDERICK, R., et al., "RTP: A Transport Protocol for Real-Time Applications", IETF RFC 1889, Jan. 1996.
- [18] ITU-T Rec. G.711, "General Aspects of Digital Transmission Systems Terminal Equipments - Pulse Code Modulation (PCM) of Voice Frequencies", 1972.
- [19] ITU-T Rec. G.728, "General Aspects of Digital Transmission Systems; Terminal Equipments - Coding of Speech at 16 kbit/s Using Low-Delay Code Excited Linear Prediction", Sept. 1992.
- [20] 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)", Mar. 1996.
- [21] SCHRÖDER, G. and SHERIF, M. H., "The Road to G.729: ITU 8-kb/s Speech Coding Algorithm with Wireline Quality", *IEEE Communication Magazine*, Sept. 1997.
- [22] GIBSON, J. D., BERGER, T., LOOKABAUGH, T., et al., "Linear Predictive Speech Coding Standards". In: Fox, E., *Digital Compression for Multimedia: Principles and Standards*, 1 ed., chapter 6, San Francisco, USA, Morgan Kaufmann Publishers, 1998.
- [23] SALAMI, R., LAFLAMME, C., ADOUL, J., et al., "A Toll Quality 8 Kb/s Speech Codec for the Personal Communications System (PCS)", *IEEE Transactions on Vehicular Technology*, v. 43, n. 3, pp. 808-816, Aug. 1994.
- [24] 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*, Sept. 1997.

- [25] 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 kbit/s", Mar. 1996.
- [26] Georgia Tech, "Graphics, Visualization & Usability Center's, Eight www User Survey", [http://www.gvu.gatech.edu/user\\_surveys/survey-1997-10/graphs/technology/Connection\\_Speed.html](http://www.gvu.gatech.edu/user_surveys/survey-1997-10/graphs/technology/Connection_Speed.html), Oct. 1997.
- [27] Georgia Tech, "Graphics, Visualization & Usability Center's, Eight www User Survey", [http://www.cc.gatech.edu/gvu/user\\_surveys/survey-1998-10/graphs/technology/q01.htm](http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/technology/q01.htm), Oct. 1998.
- [28] DOUSKALIS, B., "IP Telephony - The Integration of Robust VoIP Services", 1 ed., pp. 251-252, New Jersey, Prentice Hall, 2000.
- [29] 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.
- [30] CASNER, S. e JACOBSON, V., "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", IETF RFC 2508, Feb. 1999.
- [31] KOSTAS, T. J., BORELLA M. S., SIDHU, I., et al., "Real-Time Voice Over Packet-Switched Networks", *IEEE Network*, Jan./Feb. 1998.
- [32] ITU-T Rec. G.114, "One-way transmission time", May 2000.
- [33] BOGER, Y., "Fine-tuning Voice over Packet Services", RADCOM Inc., [http://www.radcom-inc.com/radcom/technlgy/pdf/finetuning\\_voip.pdf](http://www.radcom-inc.com/radcom/technlgy/pdf/finetuning_voip.pdf).
- [34] BOLOT, J. C., "End-to-End Packet Delay and Loss Behavior in the Internet", Proc. SIGCOMM '93, N.Y., USA, Sept. 1993.
- [35] CUEVAS, F. A., BERTRAN, M., OLLER, F., et al., "Voice Synchronization in Packet Switching Networks", *IEEE Networks*, Sept. 1993.
- [36] STONE, D. L., JEFFAY, K., "Queue Monitoring - A Delay Jitter Management Policy", *Proc. NOSSDAV*, Lancaster, U.K., Nov. 1993.
- [37] HASSAN, M., NAYANDORO, A., "Internet Telephony: Services, Technical Challenges, and Products.", *IEEE Communications Magazine*, April 2000.

- [38]“Diagnosing Cisco Call Manager Problems” – Cisco Systems Inc., [http://www.cisco.com/univercd/cc/td/doc/product/voice/c\\_callmg/3\\_1/trouble/trblch3.htm](http://www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/3_1/trouble/trblch3.htm).
- [39]HAGSAND, O., HANSON, K., MARSH, I., “Measuring Internet Telephony Quality: Where Are We Today ?”, *Global Telecommunications Conference - Globecom '99*, Rio de Janeiro, Brazil, Dec.1999.
- [40]“Playout Delay Enhancements for Voice over IP”, Cisco Systems Inc., [http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t5/dt\\_pod.htm](http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t5/dt_pod.htm).
- [41]BOLOT, J., GARCÍA, A., “The Case for FEC-Based Error Control for Packet Audio in the Internet”, *Proc. NOSSDAV*, Cambridge U.K., Apr. 1996.
- [42]FIGUEIREDO, D. R., SILVA, E. S., “Efficient Mechanisms for recovering Voice Packets in the Internet”, *Global Telecommunications Conference - Globecom '99*, Rio de Janeiro, Brazil, Dec.1999.
- [43]PERKINS, C., HODSON, O., HARDMAN, V., “A survey of Packets-Loss Recovery Techniques for Streaming Audio”, London, Aug. 1998, <http://www-mice.cs.ucl.ac.uk/multimedia/publications/ErrorCorrection.ps>.
- [44]ITU-T Rec. G.131, “Control of talker echo”, Aug. 1996.
- [45]ITU-T Rec. G.164, “Echo Suppressors”, 1988.
- [46]JAIN, R., “Comparing Systems Using Sample Data”, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, 1 ed., New York, John Wiley and Sons Inc., 1991.
- [47]ITU-T Rec. P.800, “Methods for Subjective Determination of Transmission Quality”, Aug. 1996.
- [48]ITU-T Rec. P.861, “Objective Quality Measurement of Telephone-band (300-3400 Hz) Speech Codecs”, Feb. 1998.
- [49]WILLIS, D., “Measuring Voice Quality: Listening by Numbers”, *Network Computing*, <http://www.nwc.com/1011/1011colwillis.html>, May 1999.
- [50]DAI, R., “A Technical White Paper on Sage’s PSQM Test”, [http://www.sageinst.com/pdf/psqmw8\\_00.pdf](http://www.sageinst.com/pdf/psqmw8_00.pdf), Aug. 2000.

- [51] ITU-T Rec. P.830, "Subjective Performance Assessment of Telephone-band and Wideband digital Codecs", Feb. 1996.
- [52] AGILENT TECHNOLOGIES, "Methods for Measuring Perceptual Speech Quality, White Paper", <http://literature.agilent.com/litweb/pdf/5988-2352EN.pdf>, May 2001.
- [53] AGILENT TECHNOLOGIES, "Agilent Telegra VQT – PAMS Overview", Apr, 2000, [http://onenetworks.comms.agilent.com/VQT/literature/PAMS\\_overview.pdf](http://onenetworks.comms.agilent.com/VQT/literature/PAMS_overview.pdf).
- [54] "News about PESQ", [www.pesq.orq](http://www.pesq.orq).
- [55] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", Feb. 2001.
- [56] SCHWARTZ, M., *Broadband Integrated Networks*. New York, Prentice Hall, 1996.
- [57] BRADY, P. T., "A Statistical Analysis of On-Off Patterns in 16 Conversations", *The Bell System Technical Journal*, v. 47, n. 1, pp. 73-91, Jan. 1968.
- [58] BRADY, P. T., "A model for Generating On-Off Speech Patterns in Two-Way Conversation", *The Bell System Technical Journal*, v. 48, n. 9, pp. 2445-2472, Sep. 1969.
- [59] SRIRAM, K., WHITT, W., "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data", *IEEE Journal on Selected Areas in Communications*, v. SAC-4, n. 6, Sep. 1986.
- [60] YATES, D., KUROSE, J., TOWSLEY, D., et al, "On per-session end-to-end delay distributions and the call admission problem for real-time applications with QOS requirements", SIGCOMM'93, Ithaca, N.Y., Sep. 1993.
- [61] ROUSSEEUW, P., "Least Median of Squares Regression", *Journal of the American Statistical Association*, v. 79, n. 388, Dec. 1984.
- [62] KLEINROCK, L., *Queueing Systems*, v.1, 1 ed., pp. 370-372, New York, John Wiley&Sons, 1975.

## ANEXO A – MEDIDAS DE CONVERSÇÕES TELEFÔNICAS REAIS

Tabela A.1 – Medidas de conversções telefônicas.

Pacotes por rajada de fala	Ocorrências	Probabilidade das rajadas	Logaritmo neperiano das prob. das rajadas	Pacotes por rajada de fala	Ocorrências	Probabilidade das rajadas	Logaritmo neperiano das prob. das rajadas
7	89	0.02900913	-3.54014	55	12	0.00391134	-5.54387
8	210	0.06844850	-2.68167	56	5	0.00162973	-6.41934
9	83	0.02705346	-3.60994	57	13	0.00423729	-5.46383
10	68	0.02216428	-3.80927	58	7	0.00228162	-6.08287
11	81	0.02640156	-3.63433	59	14	0.00456323	-5.38972
12	47	0.01531943	-4.17863	60	21	0.00684485	-4.98426
13	105	0.03422425	-3.37482	61	10	0.00325945	-5.72620
14	64	0.02086050	-3.86990	62	9	0.00293351	-5.83156
15	101	0.03292047	-3.41366	63	10	0.00325945	-5.72620
16	88	0.02868318	-3.55144	64	20	0.00651890	-5.03305
17	57	0.01857888	-3.98573	65	15	0.00488918	-5.32073
18	69	0.02249022	-3.79467	66	11	0.00358540	-5.63089
19	65	0.02118644	-3.85439	67	10	0.00325945	-5.72620
20	63	0.02053455	-3.88565	68	12	0.00391134	-5.54387
21	52	0.01694915	-4.07754	69	7	0.00228162	-6.08287
22	57	0.01857888	-3.98573	70	14	0.00456323	-5.38972
23	43	0.01401565	-4.26758	71	9	0.00293351	-5.83156
24	46	0.01499348	-4.20014	72	8	0.00260756	-5.94934
25	32	0.01043025	-4.56305	73	11	0.00358540	-5.63089
26	38	0.01238592	-4.39120	74	8	0.00260756	-5.94934
27	48	0.01564537	-4.15758	75	16	0.00521512	-5.25619
28	33	0.01075619	-4.53227	76	14	0.00456323	-5.38972
29	29	0.00945241	-4.66149	77	5	0.00162973	-6.41934
30	32	0.01043025	-4.56305	78	7	0.00228162	-6.08287
31	34	0.01108214	-4.50242	79	10	0.00325945	-5.72620
32	61	0.01988266	-3.91791	80	13	0.00423729	-5.46383
33	39	0.01271186	-4.36522	81	9	0.00293351	-5.83156
34	25	0.00814863	-4.80991	82	8	0.00260756	-5.94934
35	34	0.01108214	-4.50242	83	6	0.00195567	-6.23702
36	18	0.00586701	-5.13841	84	11	0.00358540	-5.63089
37	26	0.00847458	-4.77068	85	3	0.00097784	-6.93017
38	23	0.00749674	-4.89329	86	11	0.00358540	-5.63089
39	21	0.00684485	-4.98426	87	5	0.00162973	-6.41934
40	21	0.00684485	-4.98426	88	8	0.00260756	-5.94934
41	60	0.01955671	-3.93444	89	7	0.00228162	-6.08287
42	27	0.00880052	-4.73294	90	6	0.00195567	-6.23702
43	26	0.00847458	-4.77068	91	6	0.00195567	-6.23702
44	21	0.00684485	-4.98426	92	6	0.00195567	-6.23702
45	18	0.00586701	-5.13841	93	7	0.00228162	-6.08287
46	24	0.00782269	-4.85073	94	6	0.00195567	-6.23702
47	27	0.00880052	-4.73294	95	11	0.00358540	-5.63089
48	21	0.00684485	-4.98426	96	6	0.00195567	-6.23702
49	60	0.01955671	-3.93444	97	4	0.00130378	-6.64249
50	17	0.00554107	-5.19557	98	6	0.00195567	-6.23702
51	9	0.00293351	-5.83156	99	9	0.00293351	-5.83156
52	21	0.00684485	-4.98426	100	6	0.00195567	-6.23702
53	18	0.00586701	-5.13841	101	6	0.00195567	-6.23702
54	11	0.00358540	-5.63089	102	7	0.00228162	-6.08287



Tabela A.1 – Medidas de conversações telefônicas (continuação).

Pacotes por rajada de fala	Ocorrências	Probabilidade das rajadas	Logaritmo neperiano das prob. das rajadas	Pacotes por rajada de fala	Ocorrências	Probabilidade das rajadas	Logaritmo neperiano das prob. das rajadas
103	9	0.00293351	-5.83156	156	1	0.00032595	-
104	6	0.00195567	-6.23702	157	3	0.00097784	-
105	4	0.00130378	-6.64249	158	4	0.00130378	-
106	7	0.00228162	-6.08287	159	1	0.00032595	-
107	3	0.00097784	-6.93017	160	1	0.00032595	-
108	6	0.00195567	-6.23702	161	1	0.00032595	-
109	8	0.00260756	-5.94934	162	1	0.00032595	-
110	8	0.00260756	-5.94934	163	3	0.00097784	-
111	4	0.00130378	-6.64249	164	3	0.00097784	-
112	7	0.00228162	-6.08287	165	2	0.00065189	-
113	6	0.00195567	-6.23702	166	4	0.00130378	-
114	6	0.00195567	-6.23702	167	1	0.00032595	-
115	4	0.00130378	-6.64249	168	1	0.00032595	-
116	5	0.00162973	-6.41934	170	1	0.00032595	-
117	4	0.00130378	-6.64249	171	1	0.00032595	-
118	5	0.00162973	-6.41934	172	2	0.00065189	-
119	6	0.00195567	-6.23702	173	2	0.00065189	-
120	3	0.00097784	-6.93017	174	2	0.00065189	-
121	6	0.00195567	-6.23702	176	1	0.00032595	-
122	6	0.00195567	-6.23702	177	4	0.00130378	-
123	4	0.00130378	-6.64249	178	1	0.00032595	-
124	3	0.00097784	-6.93017	181	4	0.00130378	-
125	3	0.00097784	-6.93017	182	3	0.00097784	-
126	11	0.00358540	-5.63089	183	3	0.00097784	-
127	5	0.00162973	-6.41934	184	1	0.00032595	-
128	3	0.00097784	-6.93017	185	2	0.00065189	-
129	4	0.00130378	-6.64249	188	2	0.00065189	-
130	5	0.00162973	-6.41934	189	1	0.00032595	-
131	3	0.00097784	-6.93017	190	1	0.00032595	-
132	4	0.00130378	-6.64249	192	1	0.00032595	-
133	6	0.00195567	-6.23702	195	1	0.00032595	-
134	9	0.00293351	-5.83156	196	1	0.00032595	-
135	7	0.00228162	-6.08287	197	2	0.00065189	-
136	2	0.00065189	-7.33563	198	3	0.00097784	-
137	6	0.00195567	-6.23702	199	3	0.00097784	-
138	1	0.00032595	-8.02878	200	1	0.00032595	-
139	2	0.00065189	-7.33563	201	4	0.00130378	-
140	2	0.00065189	-7.33563	208	2	0.00065189	-
141	4	0.00130378	-6.64249	210	1	0.00032595	-
142	4	0.00130378	-6.64249	211	1	0.00032595	-
143	4	0.00130378	-6.64249	216	1	0.00032595	-
144	3	0.00097784	-6.93017	218	1	0.00032595	-
145	3	0.00097784	-	219	1	0.00032595	-
146	2	0.00065189	-	220	1	0.00032595	-
147	1	0.00032595	-	221	2	0.00065189	-
148	2	0.00065189	-	226	1	0.00032595	-
149	3	0.00097784	-	228	1	0.00032595	-
150	1	0.00032595	-	232	2	0.00065189	-
151	5	0.00162973	-	239	1	0.00032595	-
152	3	0.00097784	-	240	2	0.00065189	-
153	2	0.00065189	-	241	1	0.00032595	-
154	1	0.00032595	-	242	1	0.00032595	-

Tabela A.1 – Medidas de conversações telefônicas (continuação).

Pacotes por rajada de fala	Ocorrências	Probabilidade das rajadas	Logaritmo neperiano das prob. das rajadas
243	1	0.00032595	-
249	1	0.00032595	-
259	1	0.00032595	-
260	2	0.00065189	-
263	1	0.00032595	-
264	1	0.00032595	-
268	1	0.00032595	-
269	2	0.00065189	-
272	1	0.00032595	-
276	2	0.00065189	-
279	1	0.00032595	-
281	1	0.00032595	-
286	1	0.00032595	-
289	1	0.00032595	-
306	1	0.00032595	-
308	2	0.00065189	-
339	1	0.00032595	-
341	1	0.00032595	-
357	2	0.00065189	-
363	1	0.00032595	-
370	1	0.00032595	-
380	1	0.00032595	-
442	1	0.00032595	-
443	1	0.00032595	-
449	1	0.00032595	-
464	1	0.00032595	-
465	1	0.00032595	-
516	1	0.00032595	-
616	1	0.00032595	-
632	1	0.00032595	-
658	1	0.00032595	-
707	1	0.00032595	-
921	1	0.00032595	-

## ANEXO B – MEDIDAS DE PERDAS

As medidas apresentadas na tabela B.1 foram obtidas entre os meses de janeiro e abril de 2002, nos horários entre 10:00 e 12:00h e entre 14:00 e 16:00h.

Tabela B.1 – Medidas de perdas no horário de maior tráfego.

pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#
1	1408	11	2	21	0	31	0	41	0
2	365	12	0	22	0	32	1	42	0
3	132	13	1	23	0	33	0	43	0
4	48	14	1	24	0	34	0	44	0
5	17	15	0	25	0	35	0	45	0
6	14	16	0	26	0	36	0	46	0
7	6	17	0	27	0	37	0	47	0
8	2	18	0	28	0	38	0	48	0
9	2	19	0	29	0	39	0	49	0
10	4	20	0	30	0	40	0	50	1

Legenda : # – número de ocorrências de cada evento.

As medidas apresentadas na tabela B.2 foram obtidas entre os meses de janeiro e abril de 2002, nos horários entre 09:00 e 18:00h.

Tabela B.2 – Medidas de perdas entre 09:00 e 18:00h.

pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#	pacotes/ rajada	#
1	2643	11	4	21	0	31	0	41	0
2	678	12	1	22	0	32	1	42	0
3	229	13	1	23	0	33	0	43	0
4	72	14	1	24	0	34	0	44	0
5	30	15	0	25	0	35	0	45	0
6	19	16	0	26	0	36	0	46	0
7	10	17	0	27	0	37	0	47	0
8	4	18	0	28	0	38	0	48	0
9	2	19	0	29	0	39	0	49	0
10	5	20	0	30	0	40	0	50	1

Legenda : # – número de ocorrências de cada evento.

## ANEXO C – MEDIDAS DE ATRASOS E JITTER

### C.1 - Medidas entre RJ e AC, na rede *Frame Relay* da DATAPREV

As medidas apresentadas nas tabelas C.1 e C.2 foram obtidas entre os meses de janeiro e abril de 2002, nos horários entre 10:00 e 12:00h e entre 14:00 e 16:00h.

Tabela C.1 – Medidas de atraso entre RJ e AC.

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
253	552	273	138	293	24	313	1	333	3	353	0
253,5	1281	273,5	124	293,5	18	313,5	3	333,5	0	353,5	0
254	184	274	142	294	16	314	8	334	4	354	0
254,5	176	274,5	128	294,5	27	314,5	6	334,5	3	354,5	0
255	171	275	119	295	19	315	3	335	2	355	0
255,5	143	275,5	92	295,5	15	315,5	3	335,5	0	355,5	1
256	195	276	102	296	20	316	5	336	2	356	1
256,5	203	276,5	89	296,5	16	316,5	5	336,5	3	356,5	2
257	4080	277	98	297	16	317	3	337	3	357	0
257,5	759	277,5	64	297,5	18	317,5	3	337,5	0	357,5	0
258	587	278	54	298	11	318	4	338	5	358	0
258,5	540	278,5	64	298,5	7	318,5	4	338,5	3	358,5	1
259	473	279	77	299	17	319	4	339	0	359	0
259,5	530	279,5	67	299,5	9	319,5	3	339,5	6	359,5	0
260	529	280	60	300	15	320	6	340	0	360	0
260,5	420	280,5	68	300,5	16	320,5	4	340,5	0	360,5	0
261	348	281	58	301	9	321	6	341	2	361	1
261,5	304	281,5	45	301,5	11	321,5	5	341,5	1	361,5	0
262	340	282	40	302	8	322	3	342	2	362	0
262,5	328	282,5	42	302,5	7	322,5	4	342,5	2	362,5	0
263	271	283	37	303	10	323	4	343	2	363	0
263,5	271	283,5	41	303,5	14	323,5	5	343,5	2	363,5	0
264	264	284	41	304	10	324	5	344	0	364	0
264,5	278	284,5	39	304,5	8	324,5	6	344,5	3	364,5	0
265	229	285	33	305	6	325	3	345	3	365	0
265,5	237	285,5	41	305,5	5	325,5	4	345,5	2	365,5	0
266	238	286	27	306	5	326	5	346	0	366	1
266,5	186	286,5	26	306,5	9	326,5	3	346,5	2	366,5	2
267	222	287	37	307	9	327	4	347	1	367	1
267,5	199	287,5	26	307,5	12	327,5	3	347,5	1	367,5	0
268	176	288	32	308	11	328	0	348	0	368	0
268,5	193	288,5	36	308,5	8	328,5	0	348,5	2	368,5	0
269	201	289	21	309	13	329	2	349	0	369	0
269,5	184	289,5	29	309,5	4	329,5	2	349,5	2	369,5	0
270	186	290	23	310	11	330	3	350	0	370	0
270,5	172	290,5	27	310,5	6	330,5	1	350,5	3	370,5	0
271	162	291	33	311	3	331	3	351	0	371	0
271,5	147	291,5	26	311,5	3	331,5	5	351,5	0	371,5	0
272	137	292	27	312	7	332	1	352	3	372	1
272,5	145	292,5	30	312,5	6	332,5	1	352,5	2	372,5	0

Legenda : Atraso – atraso fim-a-fim na rede;

# – número de ocorrências de cada evento.

Tabela C.1 – Medidas de atraso entre RJ e AC (continuação).

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
373	0	398	0	423	0	448	0	473	0	498	0
373,5	0	398,5	0	423,5	0	448,5	0	473,5	0	498,5	0
374	0	399	0	424	0	449	0	474	0	499	0
374,5	0	399,5	0	424,5	0	449,5	0	474,5	0	499,5	0
375	0	400	0	425	0	450	0	475	0	500	0
375,5	0	400,5	0	425,5	0	450,5	0	475,5	0	500,5	0
376	0	401	1	426	0	451	0	476	0	501	0
376,5	0	401,5	2	426,5	0	451,5	0	476,5	0	501,5	0
377	0	402	0	427	0	452	0	477	0	502	0
377,5	0	402,5	0	427,5	0	452,5	0	477,5	0	502,5	0
378	0	403	0	428	0	453	0	478	0	503	0
378,5	0	403,5	0	428,5	0	453,5	0	478,5	0	503,5	0
379	0	404	0	429	0	454	0	479	0	504	0
379,5	0	404,5	0	429,5	0	454,5	0	479,5	0	504,5	0
380	0	405	0	430	0	455	0	480	0	505	0
380,5	0	405,5	0	430,5	0	455,5	0	480,5	0	505,5	0
381	1	406	0	431	0	456	0	481	0	506	0
381,5	0	406,5	1	431,5	1	456,5	0	481,5	0	506,5	0
382	0	407	0	432	0	457	0	482	0	507	0
382,5	0	407,5	0	432,5	0	457,5	0	482,5	0	507,5	0
383	0	408	0	433	0	458	0	483	0	508	0
383,5	2	408,5	0	433,5	0	458,5	0	483,5	0	508,5	0
384	1	409	1	434	0	459	0	484	0	509	0
384,5	0	409,5	0	434,5	0	459,5	0	484,5	0	509,5	0
385	0	410	0	435	0	460	0	485	0	510	0
385,5	1	410,5	0	435,5	0	460,5	0	485,5	0	510,5	0
386	0	411	0	436	0	461	0	486	0	511	0
386,5	0	411,5	0	436,5	0	461,5	0	486,5	0	511,5	0
387	0	412	0	437	0	462	0	487	0	512	0
387,5	0	412,5	0	437,5	0	462,5	0	487,5	0	512,5	0
388	0	413	0	438	0	463	0	488	0	513	0
388,5	1	413,5	2	438,5	0	463,5	0	488,5	0	513,5	0
389	0	414	0	439	0	464	0	489	0	514	0
389,5	0	414,5	0	439,5	0	464,5	0	489,5	0	514,5	0
390	0	415	0	440	0	465	0	490	0	515	0
390,5	0	415,5	0	440,5	0	465,5	0	490,5	0	515,5	0
391	0	416	0	441	0	466	0	491	0	516	0
391,5	0	416,5	1	441,5	0	466,5	0	491,5	0	516,5	0
392	1	417	0	442	0	467	0	492	0	517	0
392,5	0	417,5	0	442,5	0	467,5	0	492,5	0	517,5	0
393	0	418	0	443	0	468	0	493	0	.	0
393,5	0	418,5	0	443,5	0	468,5	0	493,5	0	.	0
394	0	419	0	444	0	469	0	494	0	.	0
394,5	0	419,5	0	444,5	0	469,5	0	494,5	0	.	0
395	0	420	0	445	0	470	0	495	0	.	0
395,5	0	420,5	0	445,5	0	470,5	0	495,5	0	.	0
396	0	421	0	446	0	471	0	496	0	.	0
396,5	0	421,5	0	446,5	0	471,5	0	496,5	0	728	0
397	1	422	0	447	0	472	0	497	0	728,5	0
397,5	0	422,5	0	447,5	0	472,5	0	497,5	0	729	1

Tabela C.2 – Medidas de *jitter* entre RJ e AC.

<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#
0	1	4,7	279	9,4	123	14,1	20	18,8	8	23,5	0
0,1	0	4,8	302	9,5	121	14,2	9	18,9	2	23,6	0
0,2	0	4,9	309	9,6	107	14,3	16	19	2	23,7	1
0,3	0	5	282	9,7	110	14,4	13	19,1	1	23,8	0
0,4	0	5,1	310	9,8	121	14,5	20	19,2	1	23,9	0
0,5	0	5,2	290	9,9	101	14,6	12	19,3	1	24	0
0,6	0	5,3	306	10	110	14,7	15	19,4	2	24,1	0
0,7	0	5,4	259	10,1	100	14,8	14	19,5	0	24,2	0
0,8	0	5,5	278	10,2	88	14,9	8	19,6	3	24,3	0
0,9	1	5,6	310	10,3	95	15	12	19,7	2	24,4	1
1	1	5,7	280	10,4	91	15,1	12	19,8	8	24,5	0
1,1	2	5,8	292	10,5	86	15,2	15	19,9	2	24,6	0
1,2	9	5,9	315	10,6	92	15,3	15	20	2	24,7	0
1,3	8	6	293	10,7	81	15,4	15	20,1	2	24,8	0
1,4	9	6,1	276	10,8	84	15,5	14	20,2	2	24,9	0
1,5	11	6,2	268	10,9	63	15,6	13	20,3	2	25	0
1,6	16	6,3	265	11	86	15,7	15	20,4	1	25,1	0
1,7	17	6,4	238	11,1	62	15,8	8	20,5	0	25,2	0
1,8	16	6,5	274	11,2	79	15,9	20	20,6	2	25,3	1
1,9	20	6,6	252	11,3	86	16	8	20,7	1	25,4	0
2	31	6,7	280	11,4	64	16,1	6	20,8	2	25,5	0
2,1	37	6,8	239	11,5	85	16,2	5	20,9	1	25,6	0
2,2	54	6,9	237	11,6	71	16,3	6	21	0	25,7	0
2,3	65	7	253	11,7	75	16,4	11	21,1	0	25,8	0
2,4	59	7,1	251	11,8	69	16,5	4	21,2	0	25,9	0
2,5	85	7,2	243	11,9	59	16,6	1	21,3	0	26	0
2,6	88	7,3	251	12	48	16,7	6	21,4	1	26,1	0
2,7	102	7,4	237	12,1	62	16,8	8	21,5	0	26,2	0
2,8	107	7,5	234	12,2	64	16,9	4	21,6	0	26,3	0
2,9	96	7,6	272	12,3	65	17	6	21,7	0	26,4	0
3	151	7,7	221	12,4	62	17,1	5	21,8	0	26,5	1
3,1	141	7,8	227	12,5	59	17,2	4	21,9	0	26,6	0
3,2	148	7,9	207	12,6	55	17,3	7	22	0	26,7	0
3,3	158	8	223	12,7	46	17,4	6	22,1	0	26,8	0
3,4	175	8,1	233	12,8	56	17,5	7	22,2	0	26,9	0
3,5	189	8,2	191	12,9	56	17,6	4	22,3	0	27	0
3,6	216	8,3	187	13	37	17,7	4	22,4	0	27,1	0
3,7	191	8,4	167	13,1	63	17,8	10	22,5	0	27,2	0
3,8	189	8,5	173	13,2	44	17,9	5	22,6	0	27,3	0
3,9	216	8,6	159	13,3	41	18	2	22,7	0	27,4	0
4	206	8,7	161	13,4	33	18,1	5	22,8	1	27,5	0
4,1	208	8,8	150	13,5	44	18,2	7	22,9	0	27,6	0
4,2	241	8,9	135	13,6	33	18,3	1	23	0	27,7	0
4,3	250	9	125	13,7	37	18,4	4	23,1	0	27,8	0
4,4	277	9,1	123	13,8	31	18,5	5	23,2	0	27,9	0
4,5	306	9,2	125	13,9	31	18,6	4	23,3	0	28	0
4,6	287	9,3	96	14	24	18,7	1	23,4	1	28,1	0

Legenda : *Jitter* – *jitter* medido conforme item 6.3;  
# – número de ocorrências de cada evento.

Tabela C.2 – Medidas de *jitter* entre RJ e AC (continuação).

<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#
28,2	1	33,2	0	38,2	0	43,2	0	48,2	0	53,2	0
28,3	0	33,3	1	38,3	0	43,3	0	48,3	0	53,3	0
28,4	0	33,4	0	38,4	0	43,4	0	48,4	0	.	0
28,5	0	33,5	0	38,5	0	43,5	0	48,5	0	.	0
28,6	0	33,6	0	38,6	0	43,6	1	48,6	0	.	0
28,7	0	33,7	0	38,7	0	43,7	0	48,7	1	.	0
28,8	0	33,8	0	38,8	0	43,8	0	48,8	0	.	0
28,9	0	33,9	0	38,9	0	43,9	0	48,9	0	.	0
29	0	34	0	39	0	44	0	49	0	.	0
29,1	0	34,1	0	39,1	0	44,1	0	49,1	0	.	0
29,2	0	34,2	0	39,2	0	44,2	0	49,2	0	.	0
29,3	0	34,3	0	39,3	0	44,3	0	49,3	0	55	0
29,4	0	34,4	0	39,4	0	44,4	0	49,4	0	55,1	0
29,5	0	34,5	0	39,5	0	44,5	0	49,5	0	55,2	0
29,6	0	34,6	0	39,6	0	44,6	0	49,6	0	55,3	1
29,7	0	34,7	1	39,7	0	44,7	0	49,7	0	55,4	0
29,8	1	34,8	0	39,8	0	44,8	0	49,8	0	55,5	0
29,9	0	34,9	0	39,9	0	44,9	0	49,9	0	55,6	0
30	0	35	0	40	1	45	0	50	0	.	0
30,1	0	35,1	0	40,1	0	45,1	0	50,1	0	.	0
30,2	0	35,2	0	40,2	0	45,2	0	50,2	0	.	0
30,3	0	35,3	0	40,3	0	45,3	0	50,3	0	.	0
30,4	0	35,4	1	40,4	0	45,4	0	50,4	0	.	0
30,5	0	35,5	0	40,5	0	45,5	0	50,5	0	.	0
30,6	0	35,6	0	40,6	0	45,6	0	50,6	0	.	0
30,7	0	35,7	0	40,7	0	45,7	0	50,7	0	.	0
30,8	0	35,8	0	40,8	0	45,8	0	50,8	0	.	0
30,9	0	35,9	0	40,9	0	45,9	0	50,9	0	.	0
31	0	36	0	41	0	46	0	51	0	57,7	0
31,1	0	36,1	0	41,1	0	46,1	0	51,1	0	57,8	0
31,2	0	36,2	0	41,2	0	46,2	0	51,2	0	57,9	0
31,3	0	36,3	0	41,3	0	46,3	0	51,3	0	58	1
31,4	0	36,4	0	41,4	0	46,4	1	51,4	0	55,7	0
31,5	1	36,5	0	41,5	0	46,5	0	51,5	0	55,8	0
31,6	0	36,6	0	41,6	0	46,6	0	51,6	0	55,9	0
31,7	0	36,7	0	41,7	0	46,7	0	51,7	0	.	0
31,8	0	36,8	0	41,8	1	46,8	0	51,8	1	.	0
31,9	0	36,9	0	41,9	0	46,9	0	51,9	0	.	0
32	0	37	0	42	0	47	0	52	0	.	0
32,1	0	37,1	0	42,1	0	47,1	0	52,1	0	.	0
32,2	0	37,2	0	42,2	0	47,2	0	52,2	0	.	0
32,3	0	37,3	0	42,3	0	47,3	0	52,3	0	.	0
32,4	0	37,4	0	42,4	0	47,4	0	52,4	0	.	0
32,5	0	37,5	0	42,5	0	47,5	0	52,5	0	.	0
32,6	0	37,6	1	42,6	0	47,6	0	52,6	0	.	0
32,7	0	37,7	0	42,7	0	47,7	0	52,7	0	.	0
32,8	0	37,8	0	42,8	0	47,8	0	52,8	0	61,3	0
32,9	0	37,9	0	42,9	0	47,9	0	52,9	0	61,4	0
33	0	38	0	43	0	48	0	53	0	61,5	0
33,1	0	38,1	0	43,1	0	48,1	0	53,1	0	61,6	1

## C.2 - Medidas entre RJ e SP, na rede *Frame Relay* da DATAPREV

As medidas apresentadas nas tabelas C.3 e C.4 foram feitas entre os meses de janeiro e abril de 2002, nos horários entre 10:00 e 12:00h e entre 14:00 e 16:00h.

Tabela C.3 – Medidas de atraso entre RJ e SP.

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
0,5	0	20,5	123	40,5	39	60,5	11	80,5	11	100,5	13
1	0	21	94	41	34	61	11	81	13	101	9
1,5	0	21,5	110	41,5	20	61,5	6	81,5	5	101,5	11
2	0	22	120	42	16	62	11	82	11	102	12
2,5	0	22,5	103	42,5	30	62,5	8	82,5	8	102,5	12
3	0	23	93	43	18	63	7	83	16	103	11
3,5	0	23,5	100	43,5	16	63,5	6	83,5	11	103,5	10
4	3097	24	87	44	23	64	10	84	7	104	12
4,5	1304	24,5	88	44,5	21	64,5	7	84,5	11	104,5	7
5	551	25	81	45	27	65	16	85	7	105	16
5,5	570	25,5	69	45,5	20	65,5	13	85,5	11	105,5	16
6	523	26	82	46	12	66	8	86	10	106	15
6,5	505	26,5	75	46,5	25	66,5	8	86,5	13	106,5	2
7	496	27	82	47	21	67	10	87	6	107	9
7,5	490	27,5	66	47,5	22	67,5	8	87,5	11	107,5	8
8	420	28	58	48	14	68	8	88	9	108	11
8,5	349	28,5	74	48,5	12	68,5	9	88,5	5	108,5	7
9	321	29	60	49	12	69	9	89	13	109	11
9,5	377	29,5	58	49,5	16	69,5	11	89,5	12	109,5	8
10	392	30	59	50	10	70	8	90	16	110	11
10,5	385	30,5	63	50,5	17	70,5	11	90,5	11	110,5	14
11	338	31	52	51	8	71	9	91	15	111	9
11,5	326	31,5	44	51,5	20	71,5	4	91,5	14	111,5	9
12	286	32	56	52	14	72	12	92	15	112	11
12,5	255	32,5	51	52,5	13	72,5	10	92,5	10	112,5	6
13	291	33	52	53	13	73	5	93	7	113	9
13,5	230	33,5	54	53,5	12	73,5	11	93,5	8	113,5	11
14	219	34	50	54	18	74	3	94	15	114	15
14,5	222	34,5	57	54,5	13	74,5	4	94,5	7	114,5	8
15	181	35	47	55	10	75	9	95	14	115	5
15,5	196	35,5	36	55,5	14	75,5	13	95,5	14	115,5	14
16	179	36	32	56	12	76	10	96	10	116	13
16,5	165	36,5	33	56,5	13	76,5	8	96,5	12	116,5	6
17	148	37	34	57	13	77	12	97	10	117	12
17,5	150	37,5	24	57,5	9	77,5	5	97,5	7	117,5	7
18	164	38	24	58	7	78	12	98	16	118	10
18,5	157	38,5	21	58,5	10	78,5	12	98,5	12	118,5	10
19	131	39	33	59	14	79	8	99	14	119	10
19,5	130	39,5	28	59,5	10	79,5	9	99,5	7	119,5	13
20	123	40	20	60	7	80	11	100	11	120	7

Legenda : Atraso – atraso fim-a-fim na rede;

# – número de ocorrências de cada evento.



Tabela C.3 – Medidas de atraso entre RJ e SP (continuação).

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
120,5	8	145,5	9	170,5	5	195,5	4	220,5	5	245,5	1
121	8	146	5	171	9	196	10	221	3	246	2
121,5	11	146,5	3	171,5	6	196,5	5	221,5	3	246,5	1
122	7	147	11	172	7	197	4	222	3	247	0
122,5	14	147,5	8	172,5	4	197,5	4	222,5	2	247,5	0
123	5	148	9	173	4	198	10	223	0	248	0
123,5	4	148,5	5	173,5	3	198,5	4	223,5	5	248,5	1
124	8	149	4	174	7	199	6	224	3	249	0
124,5	5	149,5	7	174,5	6	199,5	6	224,5	2	249,5	0
125	10	150	10	175	4	200	5	225	3	250	0
125,5	7	150,5	9	175,5	4	200,5	6	225,5	2	250,5	1
126	5	151	10	176	3	201	7	226	2	251	0
126,5	7	151,5	6	176,5	8	201,5	7	226,5	2	251,5	0
127	6	152	7	177	3	202	6	227	0	252	0
127,5	6	152,5	4	177,5	0	202,5	5	227,5	2	252,5	0
128	11	153	10	178	7	203	6	228	2	253	0
128,5	5	153,5	7	178,5	3	203,5	5	228,5	1	253,5	0
129	5	154	8	179	9	204	9	229	4	254	0
129,5	4	154,5	10	179,5	14	204,5	4	229,5	0	254,5	2
130	10	155	5	180	10	205	7	230	2	255	0
130,5	8	155,5	7	180,5	5	205,5	1	230,5	1	255,5	0
131	9	156	13	181	6	206	3	231	2	256	0
131,5	4	156,5	4	181,5	4	206,5	4	231,5	3	256,5	0
132	7	157	3	182	6	207	0	232	1	257	0
132,5	8	157,5	7	182,5	7	207,5	3	232,5	0	257,5	0
133	9	158	7	183	3	208	4	233	1	258	0
133,5	13	158,5	5	183,5	6	208,5	2	233,5	3	258,5	0
134	9	159	8	184	10	209	7	234	0	259	0
134,5	12	159,5	4	184,5	12	209,5	2	234,5	0	259,5	0
135	7	160	11	185	6	210	0	235	0	260	0
135,5	9	160,5	7	185,5	5	210,5	3	235,5	1	260,5	0
136	6	161	10	186	6	211	1	236	1	261	0
136,5	7	161,5	3	186,5	10	211,5	4	236,5	0	261,5	0
137	3	162	7	187	4	212	5	237	2	262	0
137,5	6	162,5	11	187,5	2	212,5	3	237,5	0	262,5	1
138	6	163	9	188	8	213	7	238	1	263	0
138,5	4	163,5	7	188,5	9	213,5	6	238,5	1	263,5	0
139	6	164	4	189	6	214	4	239	1	264	0
139,5	7	164,5	11	189,5	9	214,5	2	239,5	1	264,5	0
140	10	165	9	190	4	215	2	240	1	265	0
140,5	6	165,5	8	190,5	4	215,5	5	240,5	1	265,5	0
141	5	166	3	191	8	216	3	241	1	266	0
141,5	7	166,5	4	191,5	6	216,5	2	241,5	0	266,5	0
142	4	167	8	192	6	217	2	242	0	267	0
142,5	5	167,5	7	192,5	4	217,5	1	242,5	0	267,5	0
143	8	168	3	193	9	218	3	243	0	268	0
143,5	5	168,5	5	193,5	5	218,5	2	243,5	2	268,5	0
144	11	169	5	194	6	219	3	244	1	269	0
144,5	6	169,5	5	194,5	4	219,5	1	244,5	1	269,5	0
145	4	170	2	195	6	220	3	245	0	270	0

Tabela C.4 – Medidas de *jitter* entre RJ e SP.

<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#
0	1	4,7	236	9,4	134	14,1	40	18,8	3	23,5	6
0,1	0	4,8	235	9,5	126	14,2	41	18,9	9	23,6	2
0,2	0	4,9	249	9,6	127	14,3	45	19	3	23,7	1
0,3	1	5	229	9,7	119	14,4	35	19,1	5	23,8	2
0,4	0	5,1	250	9,8	118	14,5	27	19,2	0	23,9	2
0,5	0	5,2	295	9,9	118	14,6	35	19,3	4	24	3
0,6	1	5,3	265	10	108	14,7	33	19,4	5	24,1	3
0,7	1	5,4	246	10,1	111	14,8	47	19,5	4	24,2	1
0,8	1	5,5	262	10,2	112	14,9	28	19,6	6	24,3	0
0,9	2	5,6	262	10,3	110	15	30	19,7	2	24,4	1
1	1	5,7	250	10,4	104	15,1	26	19,8	4	24,5	3
1,1	0	5,8	279	10,5	90	15,2	24	19,9	3	24,6	0
1,2	2	5,9	284	10,6	118	15,3	30	20	5	24,7	2
1,3	4	6	258	10,7	92	15,4	22	20,1	4	24,8	1
1,4	8	6,1	266	10,8	99	15,5	30	20,2	2	24,9	1
1,5	13	6,2	245	10,9	120	15,6	25	20,3	4	25	0
1,6	19	6,3	258	11	108	15,7	26	20,4	7	25,1	0
1,7	24	6,4	259	11,1	102	15,8	20	20,5	5	25,2	2
1,8	28	6,5	255	11,2	86	15,9	18	20,6	3	25,3	0
1,9	40	6,6	245	11,3	93	16	21	20,7	1	25,4	1
2	27	6,7	268	11,4	90	16,1	17	20,8	3	25,5	1
2,1	41	6,8	242	11,5	84	16,2	11	20,9	2	25,6	0
2,2	44	6,9	248	11,6	82	16,3	18	21	2	25,7	1
2,3	41	7	248	11,7	69	16,4	19	21,1	3	25,8	2
2,4	48	7,1	229	11,8	81	16,5	17	21,2	6	25,9	1
2,5	45	7,2	238	11,9	90	16,6	15	21,3	4	26	2
2,6	57	7,3	247	12	57	16,7	5	21,4	5	26,1	0
2,7	70	7,4	231	12,1	79	16,8	15	21,5	5	26,2	1
2,8	79	7,5	219	12,2	70	16,9	14	21,6	2	26,3	0
2,9	101	7,6	221	12,3	63	17	12	21,7	3	26,4	2
3	108	7,7	212	12,4	53	17,1	12	21,8	2	26,5	1
3,1	121	7,8	202	12,5	60	17,2	12	21,9	1	26,6	0
3,2	128	7,9	178	12,6	70	17,3	8	22	3	26,7	0
3,3	150	8	206	12,7	66	17,4	8	22,1	1	26,8	0
3,4	155	8,1	190	12,8	63	17,5	10	22,2	3	26,9	0
3,5	182	8,2	163	12,9	56	17,6	9	22,3	2	27	1
3,6	187	8,3	203	13	50	17,7	13	22,4	1	27,1	0
3,7	229	8,4	188	13,1	51	17,8	7	22,5	0	27,2	1
3,8	215	8,5	212	13,2	48	17,9	5	22,6	4	27,3	1
3,9	201	8,6	150	13,3	45	18	6	22,7	4	27,4	0
4	220	8,7	167	13,4	43	18,1	4	22,8	4	27,5	0
4,1	198	8,8	176	13,5	42	18,2	7	22,9	1	27,6	1
4,2	222	8,9	146	13,6	39	18,3	7	23	1	27,7	0
4,3	205	9	170	13,7	49	18,4	3	23,1	0	27,8	0
4,4	208	9,1	169	13,8	42	18,5	9	23,2	8	27,9	0
4,5	197	9,2	145	13,9	44	18,6	7	23,3	0	28	0
4,6	236	9,3	162	14	41	18,7	7	23,4	4	28,1	0

Legenda : *Jitter* – *jitter* medido conforme item 6.3;  
# – número de ocorrências de cada evento.

Tabela C.4 – Medidas de *jitter* entre RJ e SP (continuação).

<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#	<i>Jitter</i> (ms)	#
28,2	0	33,2	0	38,2	0	43,2	0	48,2	0	53,2	0
28,3	0	33,3	0	38,3	0	43,3	0	48,3	0	53,3	0
28,4	0	33,4	0	38,4	0	43,4	0	48,4	0	53,4	0
28,5	0	33,5	0	38,5	0	43,5	0	48,5	0	53,5	0
28,6	0	33,6	0	38,6	0	43,6	0	48,6	0	53,6	0
28,7	1	33,7	0	38,7	0	43,7	0	48,7	0	53,7	0
28,8	0	33,8	0	38,8	0	43,8	0	48,8	0	53,8	0
28,9	0	33,9	0	38,9	0	43,9	0	48,9	0	53,9	0
29	0	34	0	39	0	44	0	49	0	54	0
29,1	0	34,1	0	39,1	0	44,1	0	49,1	0	54,1	0
29,2	0	34,2	0	39,2	0	44,2	0	49,2	0	54,2	0
29,3	0	34,3	0	39,3	0	44,3	0	49,3	0	54,3	0
29,4	0	34,4	0	39,4	0	44,4	0	49,4	0	54,4	0
29,5	0	34,5	0	39,5	0	44,5	0	49,5	0	54,5	0
29,6	0	34,6	0	39,6	0	44,6	0	49,6	0	54,6	0
29,7	0	34,7	0	39,7	0	44,7	0	49,7	0	54,7	0
29,8	1	34,8	0	39,8	0	44,8	0	49,8	0	54,8	0
29,9	0	34,9	0	39,9	0	44,9	0	49,9	0	54,9	0
30	0	35	0	40	0	45	0	50	0	55	0
30,1	0	35,1	0	40,1	0	45,1	0	50,1	0	55,1	0
30,2	0	35,2	0	40,2	0	45,2	0	50,2	0	55,2	0
30,3	0	35,3	0	40,3	0	45,3	0	50,3	0	55,3	0
30,4	0	35,4	0	40,4	0	45,4	0	50,4	0	55,4	0
30,5	0	35,5	0	40,5	0	45,5	0	50,5	0	55,5	0
30,6	0	35,6	0	40,6	0	45,6	0	50,6	0	55,6	0
30,7	0	35,7	0	40,7	0	45,7	0	50,7	0	55,7	0
30,8	0	35,8	0	40,8	0	45,8	0	50,8	0	55,8	0
30,9	0	35,9	0	40,9	0	45,9	0	50,9	0	55,9	0
31	0	36	0	41	0	46	0	51	0	56	0
31,1	0	36,1	0	41,1	0	46,1	0	51,1	0	56,1	0
31,2	0	36,2	0	41,2	0	46,2	0	51,2	0	56,2	0
31,3	0	36,3	0	41,3	0	46,3	0	51,3	0	56,3	0
31,4	1	36,4	0	41,4	0	46,4	0	51,4	0	56,4	0
31,5	0	36,5	0	41,5	0	46,5	0	51,5	0	56,5	0
31,6	0	36,6	0	41,6	0	46,6	0	51,6	0	56,6	0
31,7	0	36,7	0	41,7	0	46,7	0	51,7	0	56,7	0
31,8	0	36,8	0	41,8	0	46,8	0	51,8	0	56,8	0
31,9	0	36,9	0	41,9	0	46,9	0	51,9	0	56,9	0
32	0	37	0	42	0	47	0	52	0	57	0
32,1	0	37,1	0	42,1	0	47,1	0	52,1	0	57,1	0
32,2	0	37,2	0	42,2	0	47,2	0	52,2	0	57,2	0
32,3	0	37,3	0	42,3	0	47,3	0	52,3	0	57,3	0
32,4	0	37,4	0	42,4	0	47,4	0	52,4	0	57,4	0
32,5	0	37,5	0	42,5	0	47,5	0	52,5	0	57,5	0
32,6	0	37,6	0	42,6	0	47,6	0	52,6	0	57,6	0
32,7	0	37,7	0	42,7	0	47,7	0	52,7	0	57,7	0
32,8	0	37,8	0	42,8	0	47,8	0	52,8	0	57,8	0
32,9	0	37,9	0	42,9	0	47,9	0	52,9	0	57,9	0
33	0	38	0	43	0	48	0	53	0	58	0
33,1	0	38,1	0	43,1	0	48,1	0	53,1	0	58,1	0

### C.3 - Medidas entre a DATAPREV e a UFRJ, através da Internet

As medidas apresentadas nas tabelas C.5 e C.6 foram feitas entre os meses de janeiro e abril de 2002, nos horários entre 10:00 e 12:00h e entre 14:00 e 16:00h.

Tabela C.5 – Medidas de atraso através da Internet.

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
0,5	0	20,5	13	40,5	35	60,5	97	80,5	133	100,5	39
1	0	21	23	41	38	61	92	81	120	101	41
1,5	0	21,5	8	41,5	34	61,5	114	81,5	127	101,5	50
2	0	22	27	42	36	62	106	82	138	102	27
2,5	0	22,5	30	42,5	42	62,5	108	82,5	104	102,5	38
3	0	23	16	43	40	63	108	83	103	103	37
3,5	0	23,5	16	43,5	50	63,5	117	83,5	98	103,5	31
4	0	24	26	44	45	64	123	84	104	104	36
4,5	0	24,5	22	44,5	52	64,5	129	84,5	104	104,5	30
5	0	25	34	45	42	65	133	85	98	105	28
5,5	1	25,5	26	45,5	38	65,5	127	85,5	101	105,5	35
6	0	26	21	46	32	66	142	86	110	106	24
6,5	0	26,5	15	46,5	55	66,5	144	86,5	119	106,5	38
7	1	27	23	47	53	67	161	87	83	107	28
7,5	0	27,5	26	47,5	45	67,5	146	87,5	86	107,5	36
8	3	28	27	48	47	68	149	88	91	108	26
8,5	2	28,5	23	48,5	49	68,5	134	88,5	89	108,5	23
9	1	29	37	49	58	69	142	89	93	109	21
9,5	3	29,5	22	49,5	63	69,5	167	89,5	77	109,5	26
10	0	30	18	50	54	70	163	90	85	110	21
10,5	2	30,5	29	50,5	69	70,5	138	90,5	75	110,5	23
11	3	31	33	51	51	71	155	91	84	111	15
11,5	5	31,5	26	51,5	65	71,5	131	91,5	56	111,5	27
12	11	32	27	52	64	72	156	92	56	112	22
12,5	9	32,5	36	52,5	56	72,5	153	92,5	77	112,5	26
13	9	33	24	53	57	73	155	93	48	113	31
13,5	8	33,5	26	53,5	69	73,5	141	93,5	69	113,5	26
14	9	34	38	54	71	74	160	94	56	114	27
14,5	9	34,5	29	54,5	80	74,5	147	94,5	68	114,5	21
15	13	35	25	55	89	75	158	95	60	115	24
15,5	13	35,5	30	55,5	65	75,5	161	95,5	51	115,5	19
16	17	36	15	56	82	76	146	96	47	116	18
16,5	15	36,5	35	56,5	86	76,5	142	96,5	51	116,5	13
17	17	37	28	57	81	77	141	97	43	117	17
17,5	19	37,5	28	57,5	99	77,5	157	97,5	50	117,5	22
18	11	38	30	58	105	78	128	98	45	118	19
18,5	16	38,5	37	58,5	83	78,5	137	98,5	47	118,5	19
19	22	39	40	59	84	79	126	99	57	119	21
19,5	27	39,5	29	59,5	109	79,5	141	99,5	40	119,5	20
20	23	40	33	60	103	80	135	100	39	120	12

Legenda : Atraso – atraso fim-a-fim na rede;  
# – número de ocorrências de cada evento.

Tabela C.5 – Medidas de atraso através da Internet (continuação).

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
120,5	18	145,5	14	170,5	6	195,5	0	220,5	1	245,5	2
121	20	146	16	171	2	196	0	221	0	246	1
121,5	15	146,5	12	171,5	6	196,5	1	221,5	2	246,5	3
122	23	147	17	172	7	197	2	222	1	247	0
122,5	15	147,5	9	172,5	3	197,5	0	222,5	4	247,5	2
123	21	148	10	173	4	198	3	223	1	248	2
123,5	10	148,5	13	173,5	2	198,5	3	223,5	1	248,5	2
124	20	149	13	174	8	199	2	224	0	249	2
124,5	9	149,5	13	174,5	0	199,5	2	224,5	1	249,5	2
125	18	150	11	175	2	200	0	225	1	250	1
125,5	18	150,5	10	175,5	4	200,5	4	225,5	2	250,5	1
126	13	151	12	176	3	201	1	226	3	251	1
126,5	16	151,5	7	176,5	2	201,5	5	226,5	2	251,5	1
127	22	152	12	177	3	202	0	227	1	252	3
127,5	17	152,5	13	177,5	4	202,5	2	227,5	3	252,5	1
128	15	153	8	178	4	203	0	228	2	253	3
128,5	14	153,5	15	178,5	0	203,5	1	228,5	2	253,5	2
129	17	154	7	179	1	204	2	229	3	254	2
129,5	19	154,5	9	179,5	1	204,5	2	229,5	1	254,5	2
130	9	155	9	180	1	205	2	230	0	255	2
130,5	19	155,5	9	180,5	3	205,5	2	230,5	2	255,5	0
131	19	156	17	181	5	206	5	231	1	256	2
131,5	17	156,5	7	181,5	2	206,5	2	231,5	3	256,5	1
132	19	157	8	182	3	207	1	232	3	257	2
132,5	22	157,5	5	182,5	2	207,5	0	232,5	1	257,5	4
133	20	158	11	183	3	208	4	233	2	258	1
133,5	12	158,5	7	183,5	0	208,5	0	233,5	2	258,5	2
134	20	159	5	184	2	209	3	234	1	259	2
134,5	13	159,5	6	184,5	1	209,5	1	234,5	1	259,5	0
135	14	160	11	185	3	210	2	235	1	260	4
135,5	24	160,5	5	185,5	6	210,5	1	235,5	1	260,5	2
136	23	161	6	186	1	211	2	236	1	261	2
136,5	21	161,5	8	186,5	1	211,5	4	236,5	1	261,5	0
137	18	162	6	187	3	212	5	237	1	262	4
137,5	19	162,5	7	187,5	4	212,5	2	237,5	1	262,5	2
138	14	163	5	188	2	213	0	238	1	263	0
138,5	23	163,5	5	188,5	2	213,5	3	238,5	1	263,5	0
139	18	164	9	189	0	214	0	239	1	264	2
139,5	20	164,5	4	189,5	4	214,5	1	239,5	0	264,5	0
140	12	165	3	190	0	215	1	240	2	265	1
140,5	18	165,5	5	190,5	3	215,5	2	240,5	1	265,5	0
141	21	166	4	191	1	216	0	241	3	266	0
141,5	17	166,5	5	191,5	1	216,5	0	241,5	1	266,5	3
142	16	167	5	192	3	217	1	242	0	267	2
142,5	18	167,5	6	192,5	3	217,5	6	242,5	2	267,5	2
143	17	168	7	193	6	218	2	243	2	268	2
143,5	13	168,5	4	193,5	0	218,5	5	243,5	3	268,5	2
144	12	169	6	194	4	219	2	244	1	269	1
144,5	11	169,5	3	194,5	0	219,5	0	244,5	5	269,5	1
145	16	170	8	195	1	220	1	245	3	270	1

Tabela C.5 – Medidas de atraso através da Internet (continuação).

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
270,5	5	295,5	0	320,5	0	345,5	0	370,5	0	395,5	0
271	1	296	2	321	0	346	0	371	0	396	0
271,5	1	296,5	0	321,5	0	346,5	0	371,5	1	396,5	0
272	3	297	1	322	0	347	0	372	0	397	0
272,5	0	297,5	1	322,5	0	347,5	1	372,5	0	397,5	0
273	2	298	1	323	0	348	0	373	0	398	0
273,5	1	298,5	1	323,5	0	348,5	0	373,5	0	398,5	0
274	3	299	0	324	0	349	0	374	0	399	0
274,5	3	299,5	1	324,5	0	349,5	1	374,5	0	399,5	0
275	2	300	2	325	1	350	0	375	0	400	0
275,5	1	300,5	0	325,5	0	350,5	1	375,5	0	400,5	0
276	1	301	0	326	0	351	1	376	0	401	0
276,5	1	301,5	1	326,5	1	351,5	0	376,5	0	401,5	0
277	4	302	0	327	0	352	0	377	0	402	0
277,5	1	302,5	0	327,5	0	352,5	0	377,5	0	402,5	0
278	0	303	0	328	0	353	0	378	0	403	0
278,5	3	303,5	0	328,5	0	353,5	0	378,5	0	403,5	0
279	1	304	0	329	0	354	0	379	0	404	0
279,5	1	304,5	1	329,5	0	354,5	0	379,5	0	404,5	0
280	4	305	1	330	0	355	0	380	0	405	1
280,5	1	305,5	0	330,5	1	355,5	0	380,5	0	405,5	0
281	1	306	0	331	0	356	0	381	0	406	1
281,5	2	306,5	0	331,5	0	356,5	0	381,5	0	406,5	0
282	0	307	0	332	1	357	0	382	0	407	0
282,5	2	307,5	0	332,5	1	357,5	0	382,5	0	407,5	1
283	1	308	1	333	0	358	0	383	0	408	0
283,5	0	308,5	0	333,5	0	358,5	0	383,5	0	408,5	0
284	0	309	1	334	0	359	0	384	0	409	0
284,5	1	309,5	0	334,5	0	359,5	0	384,5	1	409,5	0
285	1	310	1	335	0	360	0	385	0	410	0
285,5	4	310,5	0	335,5	0	360,5	0	385,5	0	410,5	0
286	1	311	3	336	0	361	0	386	0	411	0
286,5	1	311,5	1	336,5	0	361,5	0	386,5	0	411,5	0
287	2	312	0	337	1	362	0	387	0	412	0
287,5	2	312,5	0	337,5	0	362,5	0	387,5	1	412,5	1
288	4	313	1	338	0	363	0	388	0	413	0
288,5	1	313,5	0	338,5	0	363,5	0	388,5	1	413,5	0
289	2	314	0	339	0	364	0	389	0	414	0
289,5	2	314,5	0	339,5	0	364,5	1	389,5	0	414,5	0
290	1	315	0	340	0	365	0	390	0	415	0
290,5	1	315,5	0	340,5	0	365,5	0	390,5	0	415,5	0
291	2	316	0	341	0	366	0	391	0	416	0
291,5	2	316,5	1	341,5	0	366,5	0	391,5	0	416,5	0
292	2	317	1	342	0	367	0	392	0	417	0
292,5	1	317,5	0	342,5	0	367,5	0	392,5	0	417,5	0
293	3	318	0	343	0	368	0	393	0	418	0
293,5	1	318,5	0	343,5	0	368,5	0	393,5	1	418,5	0
294	2	319	0	344	1	369	0	394	0	419	0
294,5	0	319,5	0	344,5	0	369,5	0	394,5	0	419,5	0
295	0	320	0	345	0	370	0	395	1	420	0

Tabela C.5 – Medidas de atraso através da Internet (continuação).

Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#	Atraso (ms)	#
420,5	0	445,5	0	470,5	0	495,5	0	520,5	0	545,5	0
421	0	446	0	471	0	496	0	521	0	546	0
421,5	0	446,5	0	471,5	0	496,5	0	521,5	0	546,5	0
422	0	447	0	472	0	497	0	522	0	547	0
422,5	0	447,5	0	472,5	0	497,5	0	522,5	0	547,5	0
423	0	448	0	473	0	498	0	523	0	548	0
423,5	0	448,5	0	473,5	0	498,5	0	523,5	0	548,5	0
424	0	449	0	474	0	499	0	524	0	549	0
424,5	0	449,5	0	474,5	0	499,5	0	524,5	0	549,5	0
425	0	450	0	475	0	500	0	525	0	550	0
425,5	0	450,5	0	475,5	0	500,5	0	525,5	0	550,5	0
426	1	451	0	476	0	501	0	526	0	551	0
426,5	1	451,5	0	476,5	0	501,5	0	526,5	0	551,5	0
427	1	452	0	477	0	502	0	527	0	552	0
427,5	0	452,5	0	477,5	0	502,5	0	527,5	0	552,5	0
428	0	453	0	478	0	503	0	528	0	553	0
428,5	0	453,5	0	478,5	0	503,5	0	528,5	0	553,5	0
429	0	454	0	479	0	504	0	529	0	554	0
429,5	0	454,5	0	479,5	0	504,5	0	529,5	0	554,5	0
430	0	455	0	480	0	505	0	530	0	555	0
430,5	0	455,5	0	480,5	0	505,5	0	530,5	0	555,5	0
431	0	456	0	481	0	506	0	531	0	556	0
431,5	0	456,5	0	481,5	0	506,5	0	531,5	0	556,5	0
432	0	457	0	482	0	507	0	532	0	557	0
432,5	0	457,5	0	482,5	0	507,5	0	532,5	0	557,5	0
433	0	458	0	483	0	508	0	533	0	558	0
433,5	0	458,5	0	483,5	0	508,5	0	533,5	0	558,5	0
434	0	459	0	484	0	509	0	534	0	559	0
434,5	0	459,5	0	484,5	0	509,5	0	534,5	0	559,5	0
435	0	460	0	485	0	510	0	535	0	560	0
435,5	0	460,5	0	485,5	0	510,5	0	535,5	0	560,5	0
436	0	461	0	486	0	511	0	536	0	561	0
436,5	0	461,5	0	486,5	0	511,5	0	536,5	0	561,5	0
437	0	462	1	487	0	512	0	537	0	562	0
437,5	0	462,5	0	487,5	0	512,5	0	537,5	0	562,5	0
438	0	463	0	488	0	513	0	538	0	563	0
438,5	0	463,5	0	488,5	0	513,5	0	538,5	0	563,5	0
439	0	464	0	489	0	514	0	539	0	564	0
439,5	0	464,5	0	489,5	0	514,5	0	539,5	0	564,5	0
440	0	465	0	490	0	515	0	540	0	565	0
440,5	0	465,5	0	490,5	0	515,5	0	540,5	0	565,5	0
441	0	466	0	491	0	516	0	541	0	566	0
441,5	0	466,5	0	491,5	0	516,5	0	541,5	0	566,5	0
442	0	467	0	492	0	517	0	542	0	567	0
442,5	0	467,5	0	492,5	0	517,5	0	542,5	0	567,5	0
443	1	468	0	493	0	518	0	543	0	568	0
443,5	0	468,5	0	493,5	0	518,5	0	543,5	0	568,5	0
444	0	469	0	494	0	519	0	544	0	569	0
444,5	1	469,5	0	494,5	0	519,5	0	544,5	0	569,5	0
445	0	470	0	495	0	520	0	545	0	570	0

Tabela C.6 – Medidas de *jitter* através da Internet.

Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#
0	1	4,7	1	9,4	263	14,1	74	18,8	25	23,5	12
0,1	0	4,8	0	9,5	266	14,2	77	18,9	23	23,6	14
0,2	0	4,9	0	9,6	269	14,3	62	19	24	23,7	16
0,3	2	5	0	9,7	251	14,4	56	19,1	34	23,8	16
0,4	0	5,1	0	9,8	271	14,5	47	19,2	25	23,9	11
0,5	0	5,2	2	9,9	259	14,6	69	19,3	25	24	16
0,6	0	5,3	5	10	264	14,7	56	19,4	24	24,1	13
0,7	0	5,4	3	10,1	254	14,8	60	19,5	22	24,2	13
0,8	0	5,5	4	10,2	239	14,9	43	19,6	27	24,3	8
0,9	0	5,6	11	10,3	248	15	44	19,7	25	24,4	8
1	0	5,7	6	10,4	232	15,1	45	19,8	28	24,5	6
1,1	0	5,8	8	10,5	217	15,2	55	19,9	31	24,6	13
1,2	0	5,9	10	10,6	217	15,3	42	20	22	24,7	11
1,3	0	6	8	10,7	212	15,4	49	20,1	20	24,8	11
1,4	1	6,1	7	10,8	226	15,5	48	20,2	21	24,9	4
1,5	1	6,2	19	10,9	207	15,6	43	20,3	21	25	7
1,6	0	6,3	17	11	175	15,7	40	20,4	16	25,1	17
1,7	0	6,4	24	11,1	181	15,8	40	20,5	22	25,2	8
1,8	0	6,5	24	11,2	194	15,9	26	20,6	17	25,3	12
1,9	0	6,6	29	11,3	179	16	35	20,7	22	25,4	5
2	1	6,7	46	11,4	174	16,1	45	20,8	16	25,5	12
2,1	1	6,8	40	11,5	192	16,2	36	20,9	26	25,6	10
2,2	0	6,9	51	11,6	179	16,3	35	21	18	25,7	5
2,3	2	7	59	11,7	141	16,4	29	21,1	22	25,8	9
2,4	1	7,1	61	11,8	151	16,5	19	21,2	20	25,9	4
2,5	0	7,2	66	11,9	167	16,6	31	21,3	21	26	8
2,6	0	7,3	59	12	141	16,7	27	21,4	28	26,1	6
2,7	0	7,4	76	12,1	151	16,8	32	21,5	11	26,2	9
2,8	0	7,5	102	12,2	167	16,9	22	21,6	19	26,3	10
2,9	0	7,6	103	12,3	136	17	24	21,7	22	26,4	8
3	0	7,7	99	12,4	140	17,1	32	21,8	15	26,5	7
3,1	0	7,8	122	12,5	132	17,2	38	21,9	19	26,6	12
3,2	0	7,9	129	12,6	130	17,3	34	22	18	26,7	7
3,3	0	8	135	12,7	109	17,4	31	22,1	19	26,8	10
3,4	0	8,1	159	12,8	117	17,5	28	22,2	12	26,9	8
3,5	0	8,2	187	12,9	108	17,6	22	22,3	20	27	7
3,6	0	8,3	170	13	105	17,7	28	22,4	24	27,1	3
3,7	1	8,4	186	13,1	89	17,8	25	22,5	19	27,2	5
3,8	1	8,5	215	13,2	97	17,9	22	22,6	13	27,3	5
3,9	0	8,6	232	13,3	80	18	26	22,7	20	27,4	5
4	0	8,7	215	13,4	108	18,1	23	22,8	13	27,5	6
4,1	0	8,8	232	13,5	75	18,2	30	22,9	17	27,6	2
4,2	0	8,9	205	13,6	84	18,3	28	23	11	27,7	2
4,3	0	9	239	13,7	93	18,4	24	23,1	21	27,8	8
4,4	0	9,1	238	13,8	81	18,5	32	23,2	13	27,9	6
4,5	1	9,2	238	13,9	74	18,6	27	23,3	14	28	6
4,6	1	9,3	262	14	67	18,7	22	23,4	13	28,1	7

Legenda : *Jitter* – *jitter* medido conforme item 6.3;  
 # – número de ocorrências de cada evento.



Tabela C.6 – Medidas de *jitter* através da Internet (continuação).

Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#	Jitter (ms)	#
28,2	5	33,2	1	38,2	0	43,2	0	48,2	0	53,2	0
28,3	6	33,3	0	38,3	0	43,3	0	48,3	0	53,3	0
28,4	7	33,4	0	38,4	2	43,4	0	48,4	0	53,4	0
28,5	8	33,5	1	38,5	0	43,5	0	48,5	0	53,5	0
28,6	5	33,6	2	38,6	0	43,6	0	48,6	0	53,6	0
28,7	2	33,7	2	38,7	1	43,7	0	48,7	0	53,7	0
28,8	7	33,8	0	38,8	0	43,8	0	48,8	0	53,8	0
28,9	4	33,9	2	38,9	0	43,9	0	48,9	0	53,9	0
29	7	34	0	39	0	44	0	49	0	54	0
29,1	5	34,1	0	39,1	0	44,1	0	49,1	0	54,1	0
29,2	3	34,2	3	39,2	0	44,2	0	49,2	0	54,2	0
29,3	3	34,3	0	39,3	0	44,3	0	49,3	0	54,3	0
29,4	3	34,4	0	39,4	0	44,4	0	49,4	0	54,4	0
29,5	1	34,5	1	39,5	0	44,5	0	49,5	0	54,5	0
29,6	4	34,6	1	39,6	0	44,6	0	49,6	0	54,6	0
29,7	3	34,7	0	39,7	0	44,7	0	49,7	0	54,7	0
29,8	3	34,8	0	39,8	0	44,8	0	49,8	0	54,8	0
29,9	6	34,9	0	39,9	0	44,9	0	49,9	0	54,9	0
30	4	35	1	40	0	45	0	50	0	55	0
30,1	1	35,1	0	40,1	0	45,1	1	50,1	0	55,1	0
30,2	2	35,2	0	40,2	0	45,2	0	50,2	0	55,2	0
30,3	4	35,3	1	40,3	0	45,3	0	50,3	0	55,3	0
30,4	2	35,4	0	40,4	1	45,4	0	50,4	0	55,4	0
30,5	1	35,5	1	40,5	0	45,5	0	50,5	0	55,5	0
30,6	0	35,6	0	40,6	0	45,6	0	50,6	0	55,6	0
30,7	1	35,7	1	40,7	0	45,7	0	50,7	0	55,7	0
30,8	4	35,8	1	40,8	0	45,8	0	50,8	0	55,8	0
30,9	1	35,9	0	40,9	0	45,9	0	50,9	0	55,9	0
31	1	36	0	41	0	46	0	51	0	56	0
31,1	0	36,1	0	41,1	0	46,1	0	51,1	0	56,1	0
31,2	2	36,2	1	41,2	0	46,2	0	51,2	0	56,2	0
31,3	2	36,3	1	41,3	0	46,3	0	51,3	0	56,3	0
31,4	0	36,4	0	41,4	0	46,4	0	51,4	0	56,4	0
31,5	5	36,5	0	41,5	0	46,5	0	51,5	0	56,5	0
31,6	0	36,6	0	41,6	0	46,6	0	51,6	0	56,6	0
31,7	1	36,7	0	41,7	0	46,7	0	51,7	0	56,7	0
31,8	3	36,8	0	41,8	0	46,8	0	51,8	0	56,8	0
31,9	1	36,9	0	41,9	0	46,9	0	51,9	0	56,9	0
32	1	37	2	42	0	47	0	52	0	57	0
32,1	0	37,1	0	42,1	0	47,1	1	52,1	0	57,1	0
32,2	3	37,2	0	42,2	0	47,2	0	52,2	0	57,2	0
32,3	2	37,3	0	42,3	0	47,3	0	52,3	0	57,3	0
32,4	1	37,4	0	42,4	0	47,4	0	52,4	0	57,4	0
32,5	3	37,5	0	42,5	0	47,5	0	52,5	0	57,5	0
32,6	2	37,6	0	42,6	0	47,6	0	52,6	0	57,6	0
32,7	0	37,7	0	42,7	0	47,7	0	52,7	0	57,7	0
32,8	1	37,8	0	42,8	0	47,8	0	52,8	0	57,8	0
32,9	0	37,9	0	42,9	1	47,9	0	52,9	0	57,9	0
33	1	38	0	43	0	48	0	53	0	58	0
33,1	0	38,1	0	43,1	0	48,1	0	53,1	0	58,1	0

## ANEXO D – MODELO DE UMA FONTE DE PACOTES DE VOZ

Em primeiro lugar deve ficar claro a real necessidade da definição de um modelo para fontes de voz em redes de pacotes. A variação dos parâmetros utilizados para definição dos perfis funcionais de redes, conforme o Anexo D, resultarão em diferentes medidas de qualidades da voz recebida. Essas medidas serão feitas em laboratório e em redes reais.

Para obter os valores de atraso, *jitter* e perdas nas redes reais, será gerado um tráfego de pacotes que simule, de forma mais aproximada possível, o tráfego real gerado por um *gateway* de voz. Isto porque dependendo da quantidade e do tempo entre pacotes enviados à rede e a relação entre este tráfego específico para medida e o tráfego global da rede, temos como resposta diferentes comportamentos desses parâmetros.

As perdas sofrem influência direta com o tipo de tráfego gerado para a medição. Pois, sendo geradas longas rajadas de pacotes formados por muitos octetos e com um pequeno intervalo de tempo entre eles, por exemplo: 10.000 pacotes de 1.500 octetos cada e com intervalo de tempo de 1 ms entre eles, a probabilidade da rede descartar alguns pacotes por falta de espaço em filas internas aos nós ou desta rajada ter algum octeto alterado por influências físicas externas (interferência eletromagnética, por exemplo), é muito maior do que no caso de uma rajada de 5 pacotes com 30 octetos cada e com 1 s de tempo entre pacotes.

No caso dos valores de atraso e *jitter*, a influência maior se dá com o tipo de topologia e tráfego global da rede. O uso de meios de comunicações distintos, como fibra óptica e satélite, é o principal fator para grandes diferenças em medições do atraso da rede. Já a variação do tráfego global e alterações dinâmicas de rotas, são os principais fatores responsáveis pelo *jitter*.

### D.1. Modelo de voz tradicional

É importante ter-se em mente que o escopo deste trabalho resume-se a fontes de voz humana, geradas em conversações telefônicas. Este tipo de tráfego é diferente de uma fonte multimídia ou mesmo de uma fonte de som genérica, que não seja a voz humana. Em fontes de som, como por exemplo uma música, o tráfego de dados gerado é contínuo, devido às componentes de som responsáveis pelos instrumentos ou vozes de fundo.

Já na conversa telefônica convencional, existem períodos de silêncio entre e intra orações e entre os períodos de fala de cada participante, quando cada um está na situação de ouvinte do outro participante.

Considerando a telefonia analógica, há muitos anos trabalha-se com a representação da fonte de voz como um processo de dois estados alternados, onde um estado representa o período de fala, com tempos médios de permanência entre 0,4 e 1,2 segundos, e o outro o período de silêncio da voz humana, com tempos médios de permanência entre 0,6 e 1,8 segundos [56]. Com razoável aproximação, estes estados podem ser considerados com tamanhos seguindo uma distribuição exponencial. O período de silêncio não é tão fidedigno à distribuição exponencial como o período de fala, contudo, uma representação mais próxima à real, refletiria no aumento do grau de complexidade do modelo que não seria justificável em comparação com a acurácia dos resultados passíveis de serem alcançados [57],[58] e [59].

Na figura D.1 temos um modelo de nascimento e morte de dois estados que representa a fala descrita no parágrafo anterior, onde  $\alpha$  é a taxa de transição do estado de fala para o de silêncio e  $\lambda$  é a taxa de transição do estado de silêncio para o de fala. Assim, podemos ter como exemplo:  $1/\alpha=0,4$  segundos (tempo médio de duração do estado de fala) e  $1/\lambda=0,6$  segundos (tempo médio de duração do estado de silêncio).

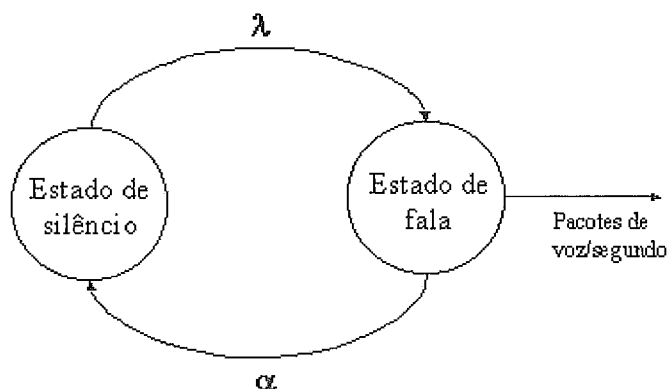


Figura D.1 - Modelo de fonte de voz com dois estados.

## D.2. Modelo de voz digitalizada

O modelo representativo da fala digitalizada, é consequência do modelo da fala analógica, anteriormente discutida. Como abordado em [59],[60], neste caso o modelo exponencial da fala é substituído pelo seu equivalente discreto, que é a distribuição geométrica. A demonstração de tal relação entre as duas distribuições pode ser visto em detalhes no Anexo E.

Para enriquecer este trabalho, foram observadas 33 conversações reais, gerando um total de 146.309 pacotes de voz, o que representa um total de 73,2 minutos de conversação com tráfego efetivo. Este tempo total de observação é maior que o utilizado em [58], que foi de 20 minutos. Cada conversação foi feita com oradores diferentes e aleatoriamente escolhidos entre os sexos masculino e feminino em seis dias de observação. O ambiente montado para a coleta de dados é ilustrado na figura D.2, onde todo tráfego de voz forçosamente passa por elementos de codificação e decodificação da voz em MP-MLQ, conforme a Recomendação G.723.1 do ITU-T. No trecho digital, um equipamento responsável pela captura dos pacotes os armazenas para análise posterior, contabilizando o número de pacotes e o intervalo de tempo entre cada um deles.

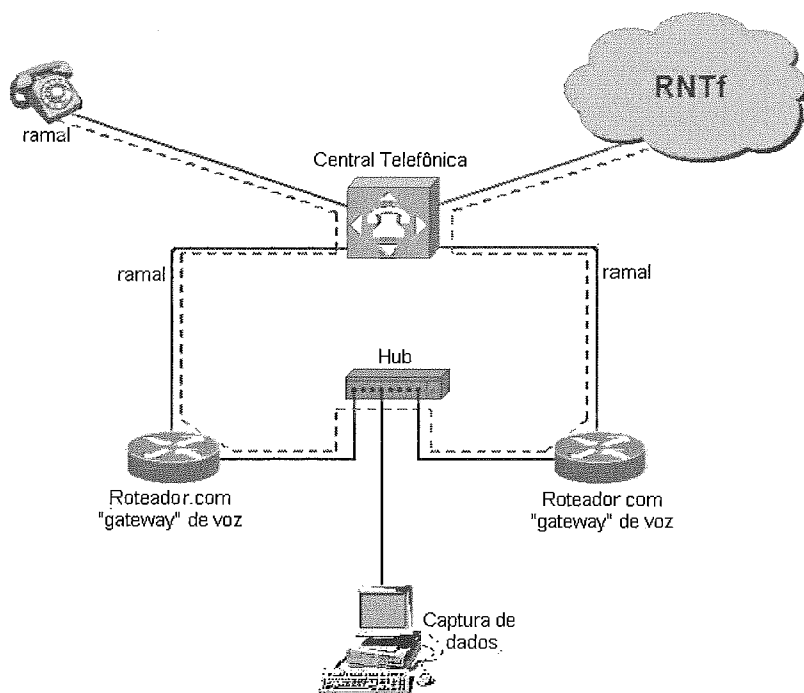


Figura D.2 – Ambiente para captura dos pacotes de voz em conversa telefônica real.

Com estas informações (apresentadas no Anexo A) é possível obter-se o modelo que caracteriza o comportamento do tamanho dos períodos de fala em uma conversa telefônica. Para tentar minimizar o efeito do ruído de fundo, unindo duas ou mais rajadas reais de fala em uma única grande rajada, trabalhou-se apenas com as ocorrências que representaram 95% de todo o conjunto, equivalente a 55,7 minutos de conversação. Dessa forma, a maior rajada considerada representa 4,3 segundos de fala, enquanto a maior medida registrada foi de 27,6 segundos, o que sem dúvida não condiz com um período de fala real, já que as pausas de orações e intervalos entre palavras, não permitiriam uma medida dessa magnitude numa conversa telefônica normal.

Apresentando as medidas do Anexo A em um gráfico, tem-se a probabilidade de ocorrência dos vários tamanhos das rajadas de fala, expressos na quantidade de pacotes de voz da G.723.1. Para ter-se o valor equivalente ao tempo de cada período de fala, deve-se multiplicar o número de pacotes da rajada por 30 ms.

É de se esperar que o gráfico da probabilidade dos períodos de fala apresente pontos discretos medidos que remetam a uma curva bem próxima de uma função exponencial. Através de ferramentas matemáticas pode-se obter a função  $f(x) = k e^{-kx}$ , que melhor se acomode às medidas feitas. Para isso foi usado o programa MAPLE V *Release 5*. Como este programa tem a restrição para ajuste de curvas apenas com funções lineares, foi usado um artifício para resolver o problema em questão.

Aplicando logaritmo neperiano ( $\ln$ ) em ambos os lados, tem-se :

$$f(x) = y = k e^{-kx} , \quad (D.1)$$

$$\ln(y) = \ln(k e^{-kx}) \text{ e} \quad (D.2)$$

$$\ln(y) = \ln(k) - kx. \quad (D.3)$$

Substituindo  $\ln(k)$  por  $b$ ,  $-k$  por  $a$  e  $\ln(y)$  por  $z$ , tem-se:

$$z = ax + b. \quad (D.4)$$

Desta forma, tem-se uma função linear, podendo-se obter as constantes  $a$  e  $b$ , através dos métodos matemáticos para regressão denominados mínimos quadrados (*Least Squares – LS*) e mínima média dos quadrados (*Least Median of Squares – LMS*), disponíveis na ferramenta.

O primeiro e clássico método, usado maciçamente, busca encontrar os coeficientes  $a$  e  $b$  que minimizem o valor da soma dos quadrados dos erros ( $\sum_{i=1, \dots, n} e_i^2$ ) na função  $Y_i = aX_i + b + e_i$ , onde  $X_i$  e  $Y_i$  são variáveis aleatórias que assumem os valores coletados experimentalmente e  $e_i$ , o erro aleatório inserido em cada par de medidas. De todos os possíveis valores para  $a$  e  $b$ , apenas um será responsável pelo menor somatório dos quadrados de  $e_i$ , encontrando-se assim a reta que melhor se acomoda aos pontos experimentais observados. O LS exige pouco consumo computacional, entretanto, um ponto com comportamento distante dos demais, interfere bastante no resultado estimado global. O segundo modelo (LMS) é mais robusto que o primeiro, exigindo um esforço computacional muito superior, mas resultando em valores mais precisos. Este método substitui a soma dos quadrados dos erros pela média destes, buscando minimizar o valor para a média dos quadrados de  $e_i$  [61].

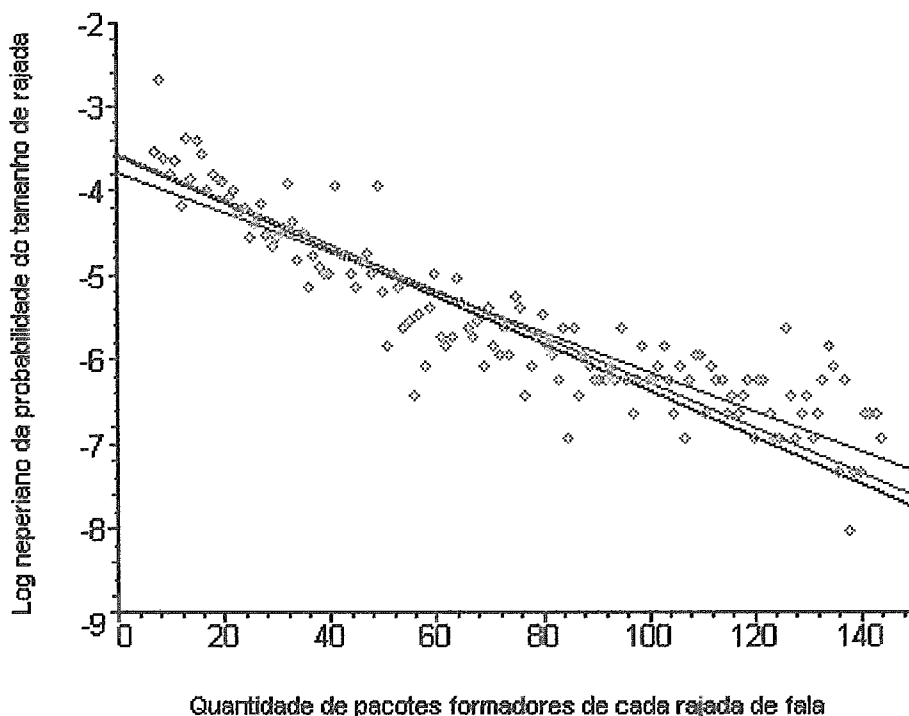


Figura D.3 – Retas obtidas pelos métodos LS (azul), LMS (vermelho) e cálculo adequando os coeficientes obtidos do método LMS (preto).

A figura D.3 mostra o gráfico do logaritmo neperiano das probabilidades de ocorrência dos períodos de fala, formados pelos pacotes de voz. Através do método LS, os valores dos coeficientes  $a$  e  $b$  foram  $-0.02353$  e  $-3.78139$ , respectivamente. Já pelo

método LMS foram obtidos os valores  $-0,02698$  e  $-3,55830$ . Como no nosso caso específico,  $-a$  e  $e^b$  devem assumir o mesmo valor, foi tirada a média aritmética entre  $0,02698$  e  $e^{-3,55830}$ , dando como resultado  $0,0277$ , que representa o valor da constante  $k$ . Assim, a reta que melhor se ajusta aos pontos medidos é  $z = -0,0277x + \ln(0,0277)$ .

Adotando os mesmos valores de coeficientes das três retas da figura D.3, e os pontos originalmente coletados, a figura D.4 mostra as probabilidades de ocorrência dos períodos de fala formados pelos pacotes de voz, por meio das curvas formadas pelas funções  $y = e^{-3,78139} e^{-0,02353x}$  (LS),  $y = e^{-3,55830} e^{-0,02698x}$  (LMS) e  $y = 0,0277 e^{-0,0277x}$  (LMS modificado).

Para o caso onde o valor de  $k$  é  $0,0277$ , em  $y = k e^{-kx}$ , a média da função é  $1/k$ , igual a  $36,10$ . Temos então, que as rajadas na média são formadas por aproximadamente  $36$  pacotes, o que representa um tempo de fala entre silêncios de  $1,08$  segundos, estando de acordo com os tempos comentados no item D.1.

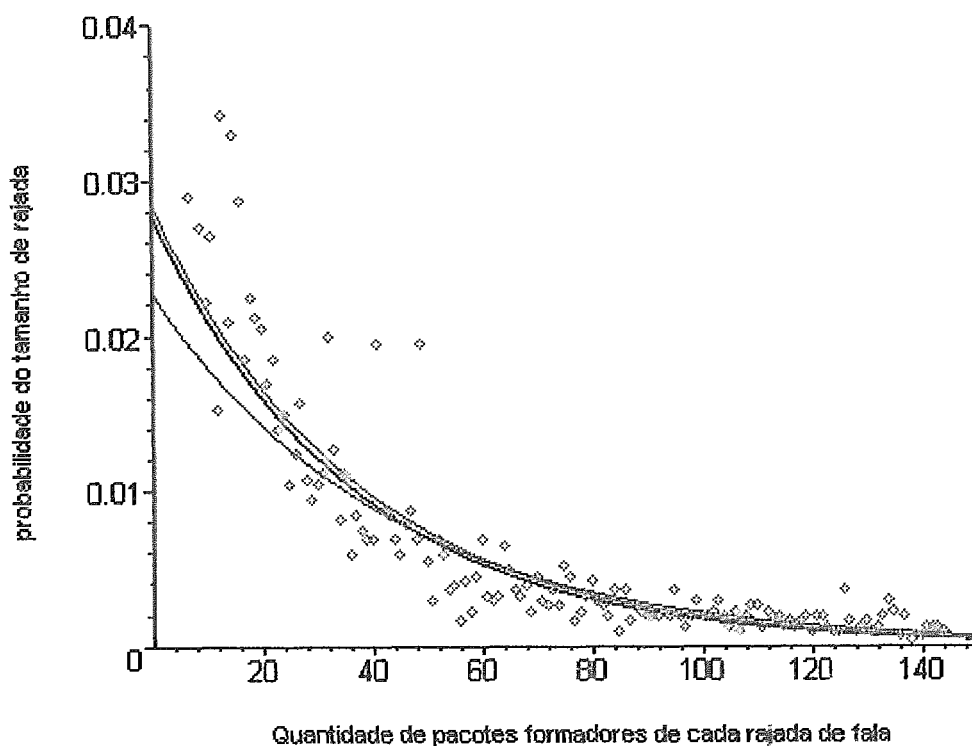


Figura D.4 – Curvas obtidas pelos métodos LS (azul), LMS (vermelho) e cálculo, adequando os coeficientes obtidos do método LMS (preto).

Nesta última figura é fácil notar o quão mais preciso é o método LMS, em relação ao LS tradicionalmente empregado.

## ANEXO E – DISTRIBUIÇÕES EXPONENCIAL E GEOMÉTRICA

Neste anexo será mostrada a relação entre as distribuições exponencial e geométrica.

Seja  $X$  esta variável aleatória (v.a.) contínua, com distribuição exponencial e média  $\lambda^{-1}$ , e  $Y$  uma v.a. em função de  $X$  onde

$$Y = \mathbf{I} ( X ) , \quad (\text{E.1})$$

com  $\mathbf{I} ( . )$  retornando o menor inteiro de seu argumento e ilustrada na figura E.1,  $Y$  assume os valores discretos  $1, 2, 3, \dots, \infty$ , referentes a  $X$ .

A probabilidade de  $Y$  assumir um valor  $k$  é dado por

$$P(Y = k) = P(k \leq X < k+1) ; k \geq 0 \quad (\text{E.2})$$

Conforme [62], a função de distribuição de probabilidade da variável aleatória  $X$ , representada por  $F_X$ , é definida como

$$F_X ( k ) = P(X \leq k) \quad (\text{E.3})$$

e

$$P(k < X \leq k+1) = F_X ( k+1 ) - F_X ( k ) . \quad (\text{E.4})$$

Seja  $X$  contínuo, a probabilidade deste assumir o valor de um ponto específico é zero

$$P(X = x) = 0 , \quad (\text{E.5})$$

logo

$$P(k < X \leq k+1) = P(k \leq X < k+1) . \quad (\text{E.6})$$

Assim, tem-se que

$$P(Y = k) = P(k \leq X < k+1) = F_X ( k+1 ) - F_X ( k ) , \quad (\text{E.7})$$

sendo  $X$  uma v.a. com distribuição exponencial,

$$P(Y = k) = (1 - e^{-\lambda(k+1)}) - (1 - e^{-\lambda k}) = e^{-\lambda k}(1 - e^{-\lambda}) ; k \geq 0. \quad (\text{E.8})$$

Substituindo  $e^{-\lambda}$  por  $p$ , tem-se

$$P(Y = k) = (1-p).p^k ; k \geq 0. \quad (\text{E.9})$$



Comparando a expressão acima com a da probabilidade de uma v.a.  $Y$  discreta, com distribuição geométrica, assumir um valor  $n$ , para  $n \geq 1$

$$P(Y = n) = (1-p).p^{n-1} , \quad (\text{E.10})$$

nota-se a correspondência entre elas, ressaltando a impossibilidade prática do caso  $P(Y = 0)$ .

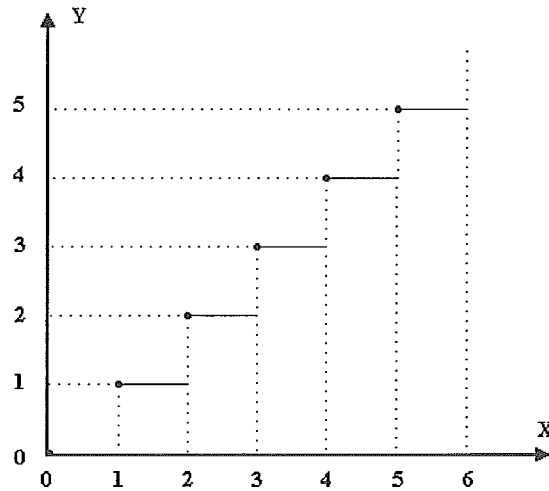


Figura E.1 – Função menor inteiro.