MARCELO DE SOUZA FREITAS

A QUALIDADE DA VOZ EM SISTEMAS DE TELECOMUNICAÇÕES

Dissertação submetida ao Programa de Mestrado em Engenharia de Telecomunicações da Escola de Engenharia da Universidade Federal Fluminense como parte dos requisitos para obtenção do grau de Mestre em Ciências.

Professor Orientador:

Edson Luiz Cataldo Ferreira, D. Sc. (UFF)

Niterói
2009

MARCELO DE SOUZA FREITAS

A QUALIDADE DA VOZ EM SISTEMAS DE TELECOMUNICAÇÕES

Dissertação submetida ao Programa de Mestrado em Engenharia de Telecomunicações da Escola de Engenharia da Universidade Federal Fluminense como parte dos requisitos para obtenção do grau de Mestre em Ciências.

Aprovada em dezembro de 2009.

BANCA EXAMINADORA

EDSON LUIZ CATALDO FERREIRA, D.Sc. - Orientador.
Universidade Federal Fluminense - UFF

MURILO BRESCIANI DE CARVALHO, D.Sc.
Universidade Federal Fluminense - UFF

CARLOS RIBEIRO CUNHA, Ph.D.
Centro de Referência Tecnológica - EMBRATEL

DÉBORA CHRISTINA MUCHALUAT SAADE, D.Sc.
Universidade Federal Fluminense - UFF

MOACYR BRAJTERMAN, Engenheiro.
Universidade Federal Fluminense - UFF

Niterói-RJ

Dedico a presente obra à minha esposa Rosana, as minhas filhas Marcelle e Maria Clara e aos meus pais Mário e Nelcina, pois suas presenças e seus incentivos nos momentos mais difíceis me serviram de inspiração.

Declaração de Originalidade

Esta dissertação foi produzida por mim e relaciona trabalho original de minha própria execução. A menos que de outra forma mencionado, os gráficos e tabelas exibidos foram produzidos a partir de dados obtidos durante a pesquisa. Sempre que materiais, idéias, ou algoritmos computacionais de outros pesquisadores tiverem sido usados ou adaptados, a fonte de informação foi claramente especificada. Esta dissertação não foi submetida para graduação ou qualificação profissional em nenhum outro lugar.

Marcelo de Souza Freitas

Agradecimentos

Este trabalho foi possível devido ao apoio inestimável do meu orientador Prof. Dr. Edson Cataldo. Agradeço, também, aos Engenheiros e técnicos do Centro de Referência Tecnológica da EMBRATEL, especialmente aos engenheiros Marcelo Gomes, Adriano e Anderson, que muito me ajudaram na execução das medições necessárias para a conclusão desta dissertação.

Resumo

Este trabalho compara os principais métodos de medição da qualidade da voz em sistemas de telecomunicações. Inicialmente, são descritos os processos de audição e produção da fala, assim como os fatores físicos e orgânicos que contribuem para uma boa comunicação. São discutidos, então, os conceitos relativos às codificações necessárias para que a voz possa ser transmitida em uma rede de telecomunicações. São apresentados, ainda, os tipos de transmissão, o processo de codificação e os codificadores, assim como as redes de telefonia convencional e VoIP. Para melhor compreender os métodos de medição da qualidade, são descritos os parâmetros que interferem no cálculo dos fatores representativos númericos da qualidade. Discutem-se, então, os parâmetros orgânicos e físicos (*pitch*, *loudness* e etc) que influenciam a qualidade da voz, além de parâmetros específicos de redes de telecomunicações (*jitter*, latência e etc). Os métodos de medição são apresentados, divididos em duas classes: métodos subjetivos e objetivos. Dentre os métodos objetivos são apresentadas diversas opções, separadas em três grupos: os intrusivos baseados em sinal, os não intrusivos baseados em sinal e o método baseado em parâmetros. Por fim são apresentados experimentos nos ambientes TDM e VoIP, com os métodos intrusivos baseados em sinal *PESQ* e *PAMS* realizados com o equipamento *Performer* da *RADCOM*, disponível no CRT (EMBRATEL) e feita uma análise comparativa dos resultados obtidos,

incluindo o modelo E, que é um método baseado em parâmetros da infraestrutura de rede. Conclusões sobre as vantagens e desvantagens de cada método usado nas experiências para avaliar a qualidade da voz transmitida são apresentadas. Uma instrução técnica detalhada do equipamento *Performer* está também inserida nesse trabalho.

Palavras-chave: Qualidade da voz, Medidas de qualidade da voz.

Abstract

This work compares the main methods for measuring the quality of voice in telecommunications systems. Initially, the processes of hearing and speech production are described, as well as physical and organic factors that contribute to good communication. Then, the concepts related to the needed coding to ensure the voice transmission in a telecommunications network are discussed. The types of transmission, the coding process and the coders are presented, as well as the conventional telephone networks and the VoIP. In order to better understand the methods of measuring quality, the parameters involved in the computation of the numerical values related to the voice quality are described: the organic and physical parameters (pitch, loudness, etc.), and specific parameters of telecommunication networks (jitter, latency, etc.). The measurement methods are then presented, divided into two classes: subjective and objective methods. Among the objective methods several options are presented, separated into three method groups: the signal-based intrusive, the signal-based non-intrusive and the parameter-based. Finally, experiments are presented in TDM and VoIP environments with the PESQ and PAMS signal-based intrusive methods using the *Performer* equipment from *RADCOM*, available at CRT (EMBRATEL). A comparative analysis of the results is performed, including the E-model, which is a network parameter-based method. Conclusions about the advantages and disadvantages of each

method used in the experiments to evaluate the transmitted voice quality are presented. A detailed instruction of the *Performer* equipment is also inserted in this work.

Keywords: Voice quality, Voice quality measure.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

A evolução tecnológica tem levado os seres humanos a cada vez mais buscar meios de comunicação mais eficientes. Os sistemas de telecomunicações têm acompanhado esta evolução, permitindo uma maior e melhor comunicação entre as pessoas, independente da distância. Soma-se, ainda, o crescente desenvolvimento de *hardware* e *software* para o processamento da voz humana, permitindo que a fala permaneça como o meio mais adequado de interação. A telefonia tradicional conseguiu, com o passar dos anos, elevados níveis de qualidade na comunicação de voz. Porém, com a introdução da telefonia celular e da telefonia através das redes de comutação de pacotes (Voz sobre IP), os níveis de qualidade não mantiveram os padrões que os usuários esperavam. Perceber a diferença de qualidade entre uma chamada realizada pela telefonia convencional e outra feita pela tecnologia VoIP ou celular não é dificil, porém, a dificuldade está em mensurar o nível de qualidade em comparação com a percepção da voz no "velho"telefone fixo. Algumas perguntas são feitas, como, por exemplo: qual método utilizar para fazer a medição? É preciso utilizar algum equipamento específico? Pode-se fazer a medição com o sistema em operação? As respostas a esses questionamentos estão no cerne desta dissertação e pretendem tornar mais objetivo um conceito tão abstrato: a qualidade da voz.

1.1 Objetivo da dissertação

O objetivo principal deste trabalho é discutir os métodos de medição da qualidade da voz humana em sistemas de telecomunicações, comparando-os entre si, utilizando-se das tecnologias de transmissão e comutação atualmente disponíveis no mercado. Analisam-se, também, os fatores que interferem na percepção da qualidade, seja na voz natural, seja na voz codificada, destacando os aspectos que podem ser aprimorados para melhoria da percepção dos usuários dos sistemas de telecomunicações.

1.2 Contribuições da dissertação

A primeira contribuição desta dissertação aparece como consequência da comparação entre métodos, o presente trabalho reuniu, em único texto, diversos conceitos e métodos relativos à medição da qualidade da voz em sistemas de telecomunicações, assunto que se encontra disperso em diversas publicações de forma superficial e que ressente de uma bibliografia específica e detalhada. Outra contribuição relevante diz respeito ao alerta para o mundo acadêmico e empresarial, ou seja, empresas provedoras e contratantes de serviços de voz, da existência de técnicas que permitem aferir níveis de qualidade da voz em sistemas contratados. Desta forma, hoje já é possível especificar a aquisição de um determinado produto/serviço que apresente uma qualidade específica e exigir a sua efetivação. No que diz respeito ao convênio existente entre a UFF e a EMBRATEL, foi estabelecido um programa de trabalho, onde por parte da EMBRATEL foi autorizado o uso, na pesquisa, dos laboratórios e equipamentos do seu Centro de Referência Tecnológica e do lado da UFF, a contrapartida foi a confecção de uma instrução técnica de utilização do equipa-

mento *Performer* do fabricante *RADCOM*, assim como um relatório dos testes e medidas realizadas.

1.3 Organização da dissertação

No capítulo 2, são apresentados o princípio de funcionamento do sistema auditivo humano, assim como os fenômenos auditivos relacionados à percepção da voz. No capítulo 3, são abordados a constituição do aparelho fonador humano, a mecânica envolvida no processo de produção da fala, as suas fontes geradoras, assim como diversos fatores contribuintes na caracterização da voz humana. No capítulo 4, são abordados os tipos de transmissão utilizados nos sistemas de telecomunicações, ou seja, a transmissão analógica e digital e os tipos de codificadores de voz. São também discutidos os tipos de redes hoje em operação: rede de telefonia comutada tradicional e as novas redes VoIP, com suas características e conceitos básicos de funcionamento. No capítulo 5, são discutidos os parâmetros de medição da voz e os mesmos são divididos em dois grupos: os parâmetros físicos e orgânicos da voz e os parâmetros importantes em redes de telecomunicações. São apresentadas as diversas grandezas que influenciam e permitem quantificar a qualidade da voz, seja na sua forma natural, analógica, ou convertida para códigos digitais de transmissão. No capítulo 6, foram estudados os dois tipos de qualidade percebidas: a qualidade no sentido acústico e a qualidade da voz em sistemas de telecomunicações. No caso dos sistemas de telecomunicações são discutidos os métodos subjetivos e os métodos objetivos de medida de qualidade de voz. No capítulo 7, são apresentados os ensaios realizados no sentido de discutir os métodos de avaliação da qualidade, nas tecnologias TDM e VoIP. Foram abordados os métodos PESQ e PAMS, disponíveis no equipamento de medição utilizado, o

VoIP *Performer* do fabricante *RADCOM*, além do modelo E que mede a qualidade através dos parâmetros de rede. Finalmente, no capítulo 8, são apresentadas as conclusões e os trabalhos futuros.

1.4 Revisão bibliográfica

Nos últimos tempos, tem havido um avanço na convergência entre as tecnologias de redes de comunicação de dados e de voz, ou seja, a tendência atual é de que em uma mesma rede trafeguem serviços de telefonia e informática. No entanto, a qualidade das comunicações de voz utilizando este conceito de unicidade de meios de comunicação não tem apresentado níveis satisfatórios. Apesar das dificuldades inerentes à arquitetura, serviços integrando a telefonia com a rede de dados têm se tornado cada vez mais comuns.

A motivação para a adesão da nova tecnologia não é somente econômica, pois os equipamentos utilizados na tecnologia VoIP(*Voice over IP* ou Voz sobre IP) são ainda muito caros. Diversos novos serviços, que antes não existiam, passaram a ser oferecidos utilizando as redes de telefonia e dados de forma integrada, como por exemplo o *Voice Mail*, onde o usuário recebe o recado deixado por uma ligação não atendida, que é enviado para o seu email . Contudo, a telefonia convencional é uma tecnologia consolidada e de elevada qualidade. É um desafio fazer com que os serviços de VoIP alcancem o mesmo patamar de qualidade da telefonia convencional.

Com a sensível degradação da qualidade nas chamadas através desta nova tecnologia, foram surgindo métodos que permitissem a medição da qualidade de forma objetiva, de modo a aferir a qualidade dos sistemas e serviços. O primeiro método a ser utilizado para avaliar a qualidade de voz foi o *Mean Opinion Score* (MOS)[4], utilizando uma avaliação

subjetiva. Nesse método, um grupo de pessoas escuta e avalia individualmente as ligações, gerando valores médios estatísticos que podem quantificar a qualidade da voz. Publicado em 1996, foi inicialmente desenvolvido para medir a qualidade dos codificadores de voz. A forma de medição e a infraestrutura necessária para sua utilização são descritas na recomendação ITU P.800 [4]. O método MOS não atende as necessidades das tecnologias atuais, pois não é possível, a cada vez que se queira fazer um teste de qualidade, reunir um grupo de pessoas com características específicas, em um local adequado fisicamente para realizar medições.

Com o advento de novas tecnologias para a transmissão de voz, surge a necessidade da criação de métodos de medição e aferição da qualidade de voz de forma automática, estimando como esta qualidade seria percebida por uma avaliação humana. Surgiram diversos modelos de avaliação objetiva, chamados intrusivos baseados em sinal, como por exemplo, PSQM (*Perceptual Speech Quality Measurement*)[6], PAMS (*Perceptual Analysis Measurement System*)[22], PESQ (*Perceptual Evaluation of Speech Quality*)[7]. Nestes métodos um sinal de referência é utilizado e comparado após a degradação do mesmo ao passar pela rede.

Outra iniciativa relevante foi o método baseado em parâmetros E-model, originalmente proposto pelo ETSI (*European Telecommunications Standards Institute*), através do relatório técnico ETR250, e depois transformado em padrão pelo ITU como G.107 [54]. O objetivo do modelo é determinar a qualidade de uma transmissão através da análise de parâmetros coletados na rede, como atrasos e perdas. Mais recentemente surgiram os métodos não intrusivos, onde um pseudo-sinal-referência é construído, através da reconstituição de trechos de fala que estão trafegando nos enlaces de comunicações. Esta referência

gerada é comparada com o a fala degradada na rede. Como exemplos podemos citar o INMD(*In service, non intrusive measurement device*)[21], CCI(*Call Clarity Index*)[44], o NINA(*Non-Intrusive Network Assessment*).e a recomendação P.563 [48].

No que diz respeito aos trabalhos relacionas sobre este tema, encontramos aa literatura convencional o livro *Beyond VoIP Protocols Understanding Voice and Networking Techniques for IP telephony* de Olivier Hersent Et-Al [68], que trata das técnicas e codificadores utilizados no VoIP e que enfoca o modelo E como ferramenta para medição de qualidade da voz. No Brasil, encontram-se como referência a dissertação de mestrado(2001)[25] e a tese de doutorado(2004) de Jaime Garcia Arnal Barbedo, da Universidade Estadual de Campinas, que abordam propostas de melhorias em métodos intrusivos. Outro texto de grande qualidade é o trabalho de Marcelo Nascimento dos Santos, dissertação de mestrado (UFPR - 2006)[58], que apresenta uma metodologia para medição da qualidade de voz utilizando redes simuladas através de software. Na Universidade Federal Fluminense, encontramos um trabalho muito interessante de Bruno de Azevedo Vianna, intitulado "Proposta e análise de um algoritmo adaptativo de ajuste de taxa de transmissão para sistemas VoIP"e que propõe um algoritmo para o ajuste adaptativo da taxa de transmissão de fontes VoIP, com base na qualidade da voz estimada pelo receptor. O ajuste é realizado através da utilização de diferentes codificadores, dependendo das condições da rede e da qualidade de voz observada[73].

A idéia central deste trabalho é comparar os métodos objetivos intrusivos baseados em sinal, PAMS, PESQ e o método baseado em parâmetros E-model. No que diz respeito à comparação entre métodos de medição da qualidade da voz, encontram-se alguns artigos de Alan Clark [23] (PHD - Leicester Polytechnic, UK.), fundador da empresa Psytechnics,

especializada em programas de medição de qualidade da voz. Os temas são o aprimoramento do modelo E e o desenvolvimento de métodos não intrusivos baseados em sinal. Seus trabalhos buscam sempre realizar comparações com os métodos intrusivos realçando a evolução do modelo E e dos métodos não intrusivos. Outra pesquisa que versa sobre comparação entre métodos de medição da qualidade é a dissertação de mestrado de Fabio David, "Ferramentas de Monitoração Ativa e Passiva para Avaliação da Qualidade de Redes VoIP", do Núcleo de Computação Eletrônica da UFRJ, que apresenta duas arquiteturas de ferramentas de monitoração e avaliação de qualidade VoIP: uma atuando no modo intrusivo , onde o tráfego é gerado pelo próprio aplicativo, e outra no modo não-intrusivo, onde o tráfego de ligações reais é capturado para avaliação em tempo real[74].

Normalmente, a referência de qualidade dos diversos métodos de medição é uma comparação a valores pré-estabelecidos pelo método MOS. A idéia de comparar métodos entre si é interessante, pois provedores de serviço e fabricantes de equipamentos apresentam valores de qualidade medidos em diferentes métodos e torna-se difícil interpretar medidas oriundas de processos e algoritmos diferenciados. A medição de um sistema analisado por um determinado método pode indicar boa qualidade, enquanto outra medida no referido sistema por outro método pode apresentar qualidade apenas razoável e fica a pergunta: qual dos dois métodos seria o mais adequado para a medição da qualidade da voz?

Capítulo 2

O sistema auditivo humano

2.1 Introdução

Não podemos discutir qualidade de voz sem antes discutir o que seja, talvez, o melhor aparelho usado para medir a qualidade: o sistema auditivo humano.

2.2 Princípio de funcionamento

O sistema auditivo humano consiste de três partes anatômicas: o ouvido externo, o ouvido médio e o ouvido interno. Graças a ele podemos discriminar várias características sonoras do ambiente como a intensidade (se o som é forte ou fraco), a frequência (se o som é agudo ou grave) e o timbre (se a mesma nota musical é de um violão ou de um piano). A Fig 2.1 mostra um esquema do sistema auditivo humano.

O ouvido externo, Fig 2.1, é constituído pelo pavilhão auricular e pelo meato acústico. O pavilhão funciona como um pré-amplificador natural e, quanto maior a sua área, maior a energia mecânica captada. Os idosos que apresentam deficiência auditiva colocam a mão em forma de concha aumentando a área efetiva do pavilhão auricular e, assim, otimizam a captação do som. O meato funciona como se fosse um tubo ressonador para sons da fala

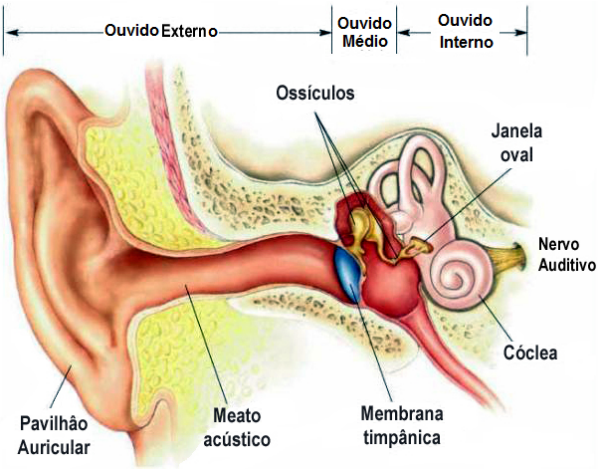


Figura 2.1: O sistema auditivo humano.

Fonte: http://www.ibb.unesp.br/nadi/audicao.htm

humana (2 a 5 KHz). Quando a onda sonora chega na membrana timpânica, ela ressona e transfere a energia mecânica para o ouvido médio.

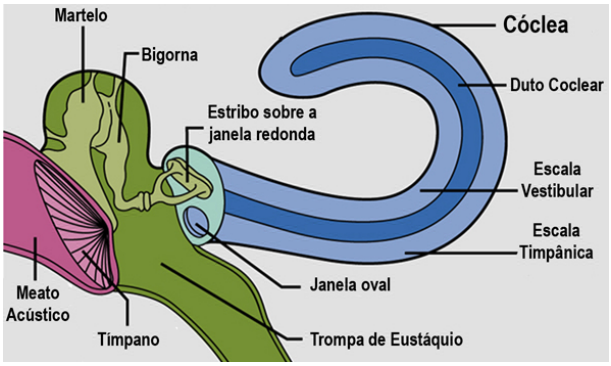


Figura 2.2: O ouvido médio.

Fonte: http://www.ibb.unesp.br/nadi/audicao.htm

O ouvido médio, mostrado na Fig 2.2, faz o limite com o ouvido externo através da membrana timpânica que está em contato com três ossículos articulados entre si (martelo, bigorna e estribo). Quando o tímpano vibra, os ossículos também vibram e funcionam como um sistema de alavancas amplificadoras do som. Adicionalmente, o fato de a área do

tímpano ser maior do que a base do estribo colabora para que a pressão sonora incidente dentro do líquido coclear seja ainda maior e, assim, a impedância é superada.

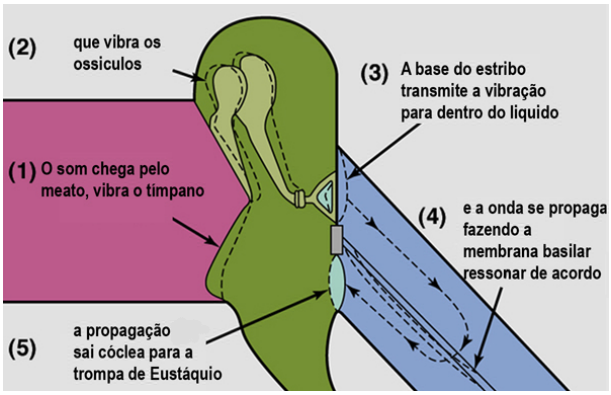


Figura 2.3: O funcionamento do sistema auditivo.

Fonte: http://www.ibb.unesp.br/nadi/audicao.htm

Conforme esquematizado na Fig 2.3, quando as ondas mecânicas chegam na cóclea, propagam-se pelo líquido coclear e deformam mecanicamente a membrana basilar onde as células sensoriais estão assentadas. A deformação mecânica da membrana repercute nas células sensoriais como uma reação elétrica correspondente e serve de estímulo para o nervo auditivo que envia os impulsos nervosos para o cérebro. Finalmente, no córtex auditivo (Fig 2.4), as informações acústicas são interpretadas permitindo que reconheçamos as diferenças entre os sons da fala e de um instrumento musical; um ruído de avião ou de um latido de cão e etc.

Cada região da membrana basilar ressona com determinada frequência sonora: à medida que se afasta da janela oval, a membrana ressona com freqüências mais baixas e as fibras do nervo auditivo também são tonotopicamente específicas.

O ouvido humano está adaptado para escutar sons entre 20 e 20.000 Hz e os sons cujas intensidades estão acima dos 90dB (decibéis) são prejudiciais à saúde auditiva. Os

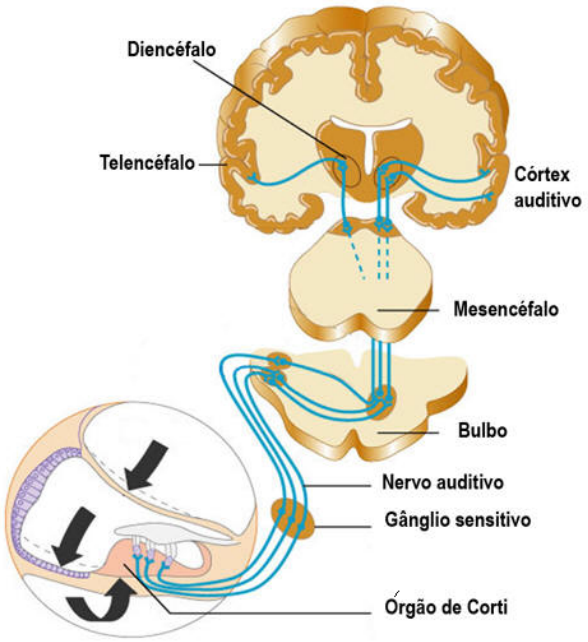


Figura 2.4: O sistema nervoso e a audição.

Fonte: http://www.ibb.unesp.br/nadi/audicao.htm

sons muito intensos causam danos permanentes às células sensoriais, acarretando déficits sensoriais.

2.3 Fenômenos auditivos - escala de Bark

A Escala de Bark é uma escala psicoacústica proposta por Eberhard Zwicker em 1961 e foi nomeada após Heinrich Barkhausen ter proposto a primeira medição subjetiva de intensidade sonora. Para um ponto específico da membrana basilar, a curva de resposta à frequência de vibração presente na janela oval é equivalente a de um filtro passa-faixa com fator de qualidade aproximadamente constante, resultando em uma melhor resolução nas baixas frequências[25]. Assim, as fibras basilares localizadas na região de altas freqüências

características respondem em uma maior faixa de freqüências do que as fibras na região de baixas frequências características. Observa-se no esquema da Fig. 2.5 a curva de resposta ao longo da membrana basilar para um tom em uma frequência específica. Para cada frequência, há um ponto da membrana basilar em que a vibração é máxima [25].

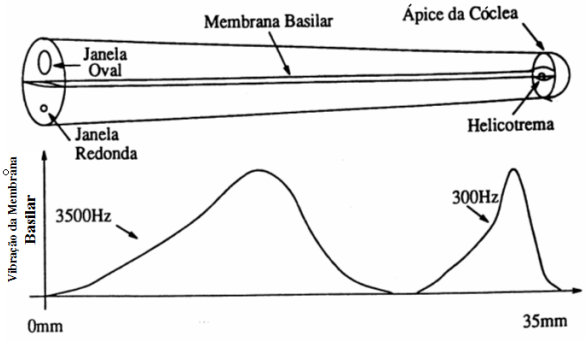


Figura 2.5: Bandas críticas na membrana basilar.

Fonte: http://www.ibb.unesp.br/nadi/bark.htm

2.3.1 Limiar absoluto de audibilidade em silêncio

O limiar absoluto de audibilidade em silêncio é o menor nível, em função da frequência, para o qual um tom se torna audível. Este limiar pode ser aproximado pela seguinte expressão analítica[25]:

$$lim = 3,64f^{-0,8} - 6,5e^{-0,6(f-3,30)^2} + 10^{-3}f^4 \tag{2.1}$$

O limiar é, em geral, expresso em dB_{SPL}, onde SPL significa *sound pressure level*, ou seja, $lim = -20log(P/Pref)$, onde P é pressão sonora do sinal (microbar) e $Pref = 0,0002$ microbars que é a pressão sonora no limiar da audição.

A equação 2.1 apresenta três termos: o primeiro descreve o corte nas baixas frequências;

o segundo descreve o aumento de sensibilidade do ouvido para a faixa de frequências em torno de 3 kHz e o último descreve o corte nas altas frequências. O primeiro termo, ou pelo menos parte dele, é interpretado como um resultado do ruído interno (causado por atividade muscular, fluxo de sangue etc.), ao passo que os dois últimos termos são interpretados como a característica de transferência de ouvido médio para o interno. Consequentemente, em modelos perceptuais, esta equação é frequentemente dividida em duas partes: uma chamada função de ruído interno e outra chamada função de transferência do ouvido médio.

2.3.2 Bandas Críticas

A cada ponto da membrana basilar temos uma frequência associada. Uma banda crítica define uma faixa de frequências audíveis em torno de cada frequência associada a cada ponto da membrana basilar. Assim, a cada ponto da membrana basilar é possível definir uma banda crítica.

Quando dois sinais se situam dentro de uma mesma banda crítica, o de maior energia poderá dominar a percepção e mascarar o outro estímulo sonoro. Portanto, dependendo dos níveis, dois tons distintos somente serão distinguidos um do outro quando estiverem em bandas críticas diferentes. Este é o fenômeno responsável pelo mascaramento simultâneo. A resolução para a distinção entre uma freqüência e outra varia de 100 Hz, nas frequências mais baixas, a mais de 6 kHz, nas frequências mais altas. Sinais com uma largura de banda suficiente para extrapolar os limites de uma banda crítica sempre proporcionarão uma intensidade perceptual maior que aqueles cujas componentes espectrais estejam limitados a uma única banda crítica, ainda que o nível de pressão sonora e a frequência central sejam

equivalentes.

A largura de faixa das bandas críticas corresponde a um espaçamento uniforme de 1,5 mm ao longo da membrana basilar, o que corresponde a aproximadamente 100 Hz para frequências abaixo de 500 Hz e de aproximadamente 20 por cento da frequência central da banda para frequências acima de 1000 Hz (em direção à janela oval). Portanto, a resposta de amplitude em frequência, para cada banda crítica, pode ser modelada como a de um filtro passa-faixas com largura de faixa crescente com a frequência. Tais filtros possuem cortes acentuados: 65 dB/oitava para as bandas críticas em torno de 500 Hz e 100 dB/oitava em torno de 8 kHz.

Embora exista uma banda crítica ao redor de cada frequência, convencionou-se (com algumas pequenas variações) a adoção dos valores mostrados na Tabela 2.1. Os valores apresentados na primeira coluna da tabela correspondem à escala Bark. Uma distância de 1 Bark corresponde à largura de uma banda crítica. A faixa de frequências audíveis corresponde a, aproximadamente, 24 Barks.

Bandas Críticas							
B. Crítica	Freq. Inferior	Freq. Superior	Faixa	B. Crítica	Freq. Inferior	Freq. Superior	Faixa
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

Tabela 2.1: Tabela de bandas críticas.

Para possibilitar o entendimento das escalas perceptuais de frequência é necessário definir o conceito de *pitch*. O termo *pitch* tem sido usado com dois sentidos diferentes, dependendo de sua área de aplicação: na área de processamento de voz, o termo é frequentemente utilizado para designar a frequência de oscilação da glote (vibração das cordas vocais), ou conforme estabelece o ANSI (American National Standards Institute), *pitch* é o atributo auditivo de acordo com o qual os sons podem ser ordenados, em uma escala de frequências, de baixo a alto. No campo da psico-acústica é usado como um atributo da sensação auditiva. Nesta seção será utilizado o conceito da área de processamento de voz.

Devido ao fenômeno das bandas críticas, a resolução espectral da audição não é linear.

Assim, escalas de frequência lineares não modelam adequadamente a percepção de *pitch*, nem são apropriadas para explicar os efeitos auditivos no domínio da frequência.

A principal vantagem de se usar uma escala de frequência auditiva ao invés de uma simples escala de frequência linear ou logarítmica é a facilidade que ela confere à modelagem dos efeitos no domínio da frequência. A escala Bark é a mais utilizada nos métodos objetivos de avaliação da qualidade de áudio por fornecer os resultados mais consistentes para este tipo de aplicação.

Uma aproximação muito simples para a relação entre uma frequência f e seu valor correspondente z na escala Bark é dada pela Eq. 2.2:

$$f = 600sinh(z/6) \tag{2.2}$$

onde f é dado em kHz. Esta aproximação foi projetada para uma faixa de frequências relevante para a codificação de voz, isto é, ela só é válida para frequências abaixo de 5 kHz.

A Eq.2.3 relaciona f e z para toda a faixa audível de frequências:

$$z = 13arctan(0,76f) + 3,5arctan[(f/7,5)^2] \tag{2.3}$$

2.3.3 Formato das bandas críticas

O sistema auditivo humano processa os sons em sub-bandas, chamadas bandas críticas. A largura de cada banda crítica difere de acordo com a faixa de frequência: abaixo de 500 Hz as bandas são constantes e têm largura de 100 Hz. Acima de 500 Hz a largura da próxima banda crítica é 20% maior que a da anterior. Um Bark corresponde à largura de uma banda crítica. O formato atribuído às bandas críticas, em que dois picos adjacentes ficam, por

definição, espaçados de 1 Bark, é denominado função espalhamento da membrana basilar, e sua formulação varia consideravelmente entre os autores.

Por exemplo, para cada uma das "b"bandas críticas[25]:

$$F(b) = 0, \quad \text{para} \quad b - bl < 1,3 \tag{2.4}$$

$$F(b) = 10^{2,5(b-bl+0,5)}, \quad \text{para} \quad -1,3 < b - bl < -0,5 \tag{2.5}$$

$$F(b) = 1, \quad \text{para} \quad -0,5 < b - bl < 0,5 \tag{2.6}$$

$$F(b) = 10^{1,0(0,5-(b-bl))}, \quad \text{para} \quad 0,5 < b - bl < 2,5 \tag{2.7}$$

$$F(b) = 0, \quad \text{para} \quad b - bl > 2,5 \tag{2.8}$$

Onde "l"se refere à l-ésima banda crítica e "bl"é a frequência central, em Bark, dessa banda. Segundo esta função, duas bandas críticas adjacentes se encontram nas extremidades dos valores de pico, como mostra a Fig.2.6.

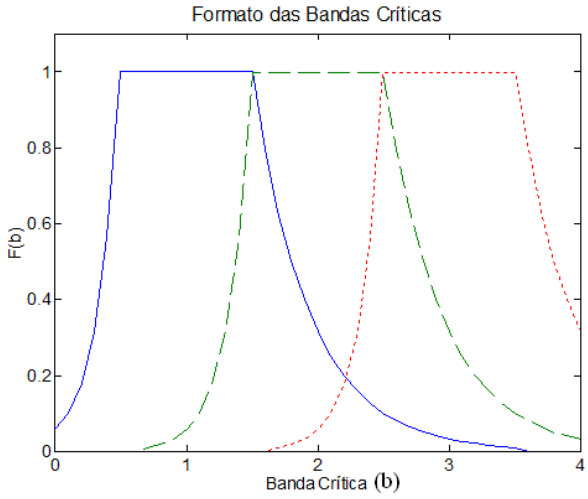


Figura 2.6: Formato das 3 primeiras bandas críticas

Fonte: http://www.decom.fee.unicamp.br

Capítulo 3

O mecanismo de produção da fala

3.1 Introdução

Após termos discutido um pouco sobre o funcionamento da audição humana, discutiremos a seguir o mecanismo de produção da fala.

3.2 O aparelho fonador

O aparelho fonador é composto externamente pela língua, lábios e dentes e internamente por três subsistemas que atuam de modo sucessivo na produção da fala: respiratório, laringeal e supralaringeal.

3.2.1 Subsistema respiratório

O subsistema respiratório é o responsável pela passagem da corrente de ar dos pulmões para a traquéia e a laringe.

3.2.2 Subsistema laringeal (ou laríngeo)

O subsistema laringeal (ou laríngeo), ou simplesmente laringe, situado na parte superior da traquéia, é o mais importante subsistema do aparelho fonador. Nele estão localizados a glote, a epiglote (válvula elástica que obstrui a glote durante a deglutição) e as cordas vocais.

A parte mais importante deste subsistema é a glote, que consiste de uma pequena abertura de forma triangular situada próxima ao pomo-de-adão. Graças à chegada do fluxo de ar vindo dos pulmões, a glote pode abrir-se ou fechar-se, bastando que as bordas das cordas vocais se afastem ou se aproximem.

Com a glote aberta, o ar passa livremente sem fazer vibrar as cordas vocais produzindo um fonema surdo ou não vozeado. Com o movimento cíclico das cordas vocais, causado pelo aumento da pressão do ar sub-glotal vindo dos pulmões, e fechamento, causado pela força de recuperação elástica e pelo efeito de Bernoulli, as cordas vocais vibram em uma frequência fundamental característica, e o fonema produzido, então, é dito sonoro ou vozeado.

A taxa na qual as cordas vocais vibram é controlada pela pressão de ar imposta pelos pulmões, pela tensão e rigidez das cordas vocais e pela área da abertura glotal em condições de repouso. Resumindo, o subsistema laringeal é o responsável pela passagem da corrente de ar que pode provocar ou não a vibração das cordas vocais.

3.2.3 Subsistema supralaringeal (ou supralaríngeo)

Este subsistema completa o mecanismo da produção da fala, com a passagem dos pulsos ou da corrente de ar pela faringe, sujeitos a obstruções ou constrições em vários pontos de articulação (nas cavidades nasal e bucal).

O trato vocal, região situada desde as pregas vocais até as extremidades da cavidade nasal (as narinas) e da cavidade bucal (os lábios), compreende os subsistemas laringeal e supralaringeal.

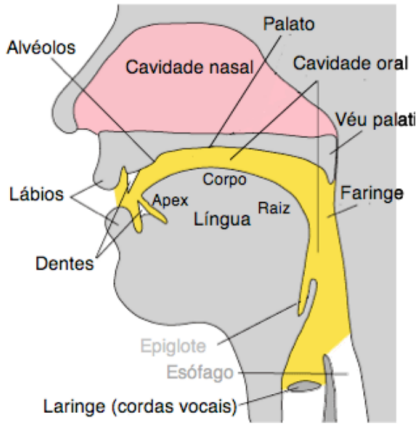


Figura 3.1: Componentes do processo de produção da voz.

Fonte: http://www.fl.ul.pt/guiao04

3.3 O processo de produção da fala

Chamamos de *vocalização* ao som produzido pela vibração das cordas vocais, porém, muitas vezes pode ser produzido um som sem palavras (som produzido pelos animais, o aquecimento de um cantor antes do espetáculo e etc). Chamamos de *fonação* ao processo físico e fisiológico de vibração das cordas vocais, que produz a voz.

O aparelho fonador humano, apresentado na Fig 3.1, é o primeiro bloco na cadeia da comunicação falada. Os sinais de fala são produzidos durante a fase de exalação (a produção de fala durante a fase de inalação é extremamente rara). O fluxo de ar permite a vibração das cordas vocais, Fig 3.2, situadas na laringe e excita o trato vocal constituído pela faringe, cavidade bucal, língua, lábios e dentes. Para produção de sons nasalados, o

véu palatino abre, e o ar, depois de passar pela cavidade nasal, é radiado pelas narinas.

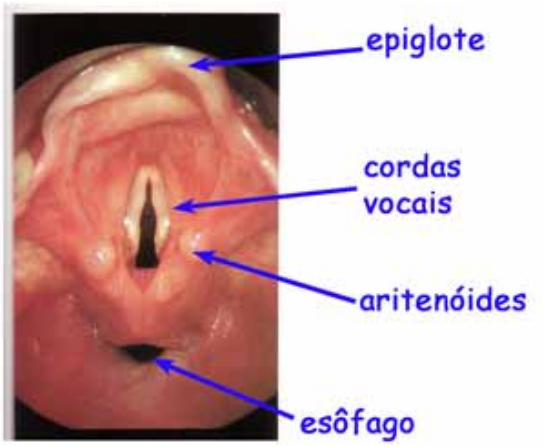


Figura 3.2: As cordas vocais

Fonte: http://www.viaaereadificil.com.br/anatomia/anatomia.htm

Cada locutor produz o seu próprio padrão sonoro através do total controle sobre a alteração da posição dos músculos e orgãos responsáveis pela fala, aliados à passagem de ar pelas cavidades e tubos existentes.

No processo de produção da fala, considera-se como *disfônica* a pessoa que possui vibração anormal de suas cordas vocais. Quando as cordas vocais de uma pessoa não vibram ela é chamada de *afônica*.

3.4 Periocidade das fontes sonoras

A fonte sonora pode ser classificada como periódica ou aperiódica. Quando o som produzido é vocálico, obrigatoriamente ocorre a vibração das cordas vocais, produzindo um fluxo de ar periódico. Quando o locutor profere uma consoante, o fluxo de ar pode ser pu-

ramente aperiódico (consoantes surdas) ou pode haver combinação de fluxo de ar periódico e aperiódico (consoantes sonoras).

3.5 A anatomia da laringe

A laringe está situada em uma zona média e anterior do pescoço, abaixo do osso hióide, entre a faringe (situada um pouco atrás e acima) e a traquéia e compreende o conjunto de cordas vocais, a epiglote, as cartilagens e músculos que a suportam, protegem e fazem movimentar, e uma mucosa de revestimento. Fazem parte da sua anatomia as membranas, ligamentos, nervos e vasos sanguíneos. Sua forma é semelhante a de uma pirâmide invertida, conforme ilustrado na Fig 3.3. É constituída por cartilagens articuladas entre si por músculos, e revestida por mucosa. A laringe, na parte superior, abre-se para a faringe, e na parte inferior, abre-se para a traquéia.

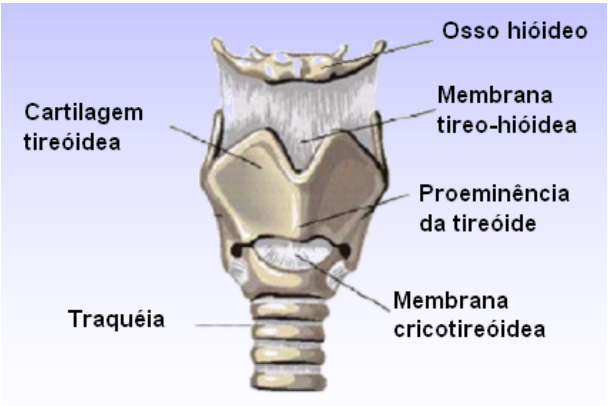


Figura 3.3: Cartilagens da laringe.

Fonte: http://www.viaaereadificil.com.br/anatomia/anatomia.htm

A laringe, que é revestida por mucosa, começa na epiglote e termina na borda inferior da cartilagem cricóide. No homem, a laringe tem, em média, comprimento de 4,50 cm e,

na mulher, 3,50 cm.

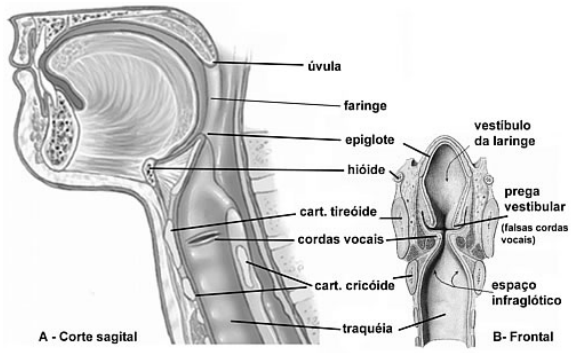


Figura 3.4: A laringe em cortes.

Fonte: http://www.viaaereadificil.com.br/anatomia/anatomia.htm

No relevo interno do cilindro interno da laringe, Fig 3.4, observam-se duas saliências: as pregas ventriculares (falsas cordas vocais) e as cordas vocais propriamente ditas ou verdadeiras. As bandas ventriculares e as cordas vocais dividem a laringe em 3 patamares: zona supra-glótica, glótica e infra-glótica.

A chamada zona supra-glótica, está localizada acima da glote, e é constituída de epiglote, prega ariepiglótica, aritenóide, pregas vestibulares e ventrículo. É separada em duas sub-regiões: (1) Epilaringe (porção supra-hióidea) e (2) Supraglote (porção infra-hióidea).

A zona glótica(média) ou glote propriamente dita, é constituída de cordas vocais, comissuras anterior e superior. Pode ser dividida em duas partes: (1) Interligamentosa (cordas vocais) e (2) Intercartilaginosa (faces internas das cartilagens aritenóides e, posteriormente, ao músculo aritenóideo).

A Zona infra-glótica vai desde as cordas vocais até ao primeiro anel da traquéia e a subglote, vai do limite inferior da glote até a borda inferior da cricóide.

3.6 A produção de voz

A oscilação é um repetido movimento de *vai e volta*. O que faz as cordas vocais oscilarem é um fato interessante, pois o movimento de ida e volta se auto sustenta por todo o tempo. O fenômeno é chamado de oscilação de fluxo induzido ou auto-sustentada.

As cordas vocais são mantidas juntas por uma pressão negativa de Bernoulli na glote. De acordo com a conservação de energia do fluxo, esta ação é possível se a glote é suficientemente estreita, o fluxo de ar é suficientemente alto, e a parede da glote é macia o suficiente para ceder. O colapso da glote é então seguido de uma subida da pressão glotal durante o fechamento, causando o início do movimento lateral das cordas vocais e a abertura da glote. O movimento lateral continua até as forças elásticas no tecido retardarem o movimento e em último estágio revertê-lo. O tecido então se move de forma mediana novamente, e outro ciclo de colapso inicia.

Fisicamente, o som é gerado sempre que existir uma perturbação do equilíbrio da densidade (ou pressão) de um gás, líquido ou sólido. Se um distúrbio na pressão local em um meio contínuo contém frequências na faixa de 20 até 20.000 Hz (faixa audível), o som é produzido. Frequências abaixo de 20 Hz são chamadas infrassônicas, e frequências acima de 20.000 Hz são chamadas ultrassônicas. Se o distúrbio de pressão é positivo (acima da pressão média), haverá uma condensação no meio, e um aumento na densidade do ar. Se o distúrbio de pressão for negativo (abaixo da pressão média), haverá uma rarefação no meio, e uma diminuição na densidade do ar.

A fonte glotal é formada por pulsos de ar, na forma de jatos, efetivamente aspergido dentro de uma coluna de ar no trato vocal (porção que vai da glote à boca) em intervalos de tempo regulares[1]. A fonte glotal surge após a passagem do fluxo de ar pelas cordas

vocais, ilustrada na Fig 3.4. Dois ciclos típicos de fluxos glotais são mostrados na Fig 3.5. Esta forma de onda representa a combinação de um fluxo transglotal e de um fluxo de deslocamento e a criação de um distúrbio de pressão na entrada do trato vocal.

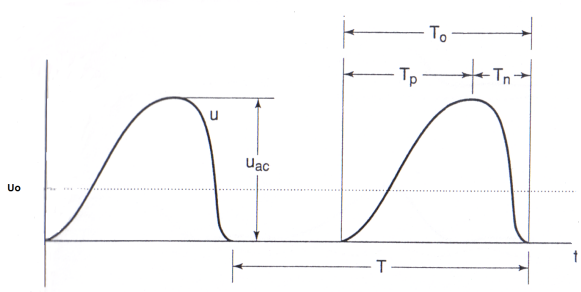


Figura 3.5: Intervalos de tempo do fluxo glotal.

Fonte: Livro "Principles of Voice Production"[1].

A turbulência no fluxo de ar glotal é uma fonte adicional de som. Tem a qualidade de um assovio, que é chamado aspiração quando combinado com a vibração das cordas vocais e um sussurro com a ausência de vibração das cordas vocais. A aspiração é um importante componente percentual na voz respirada.

Em adição aos sons secundários criados pela turbulência do ar, existem sons transientes produzidos pelo repentino começo e término da fala. Este sons transientes são chamados paradas glotais.

Finalmente, sons secundários podem ser gerados pelo movimento de líquidos nas das cordas vocais ou próximo das mesmas. Este som é caracterizado por uma sonoridade "molhada". Uma pequena quantidade deste som é tolerada na fala normal, porém a sonoridade molhada demais é considerada um distúrbio.

3.7 O espectro de frequências das fontes de som

Geralmente, uma coleção infinita de senóides são necessárias para construir um tom complexo, como o da voz humana. Por definição, um espectro de amplitudes é um gráfico da amplitude relativa pela frequência de todos os componentes. Para preservar todas as informações contidas em uma forma de onda, um espectro de fase deve ser construído. Neste caso, os pares de números fase-frequência deverão ser plotados.

Qualquer forma de onda complexa pode ser transformada em combinação dos espectros de amplitude e fase. Se os componentes de frequência são inteiros múltiplos de outro; a forma de onda é periódica e os componentes são chamados harmônicos. A frequência de qualquer harmônico pode ser calculada como nF_0, onde n é número do harmônico e F_0 é a frequência fundamental.

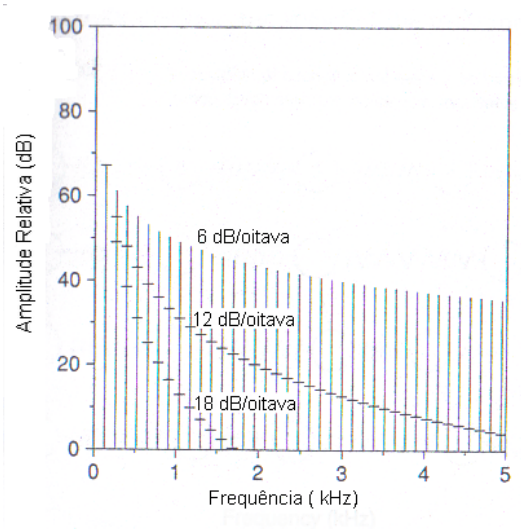


Figura 3.6: Espectro de frequências da voz humana

Fonte: Livro "Principles of Voice Production"[1].

A inclinação do espectro é a medida de como as amplitudes sucessivas dos compo-

nentes decresce com o aumento do número do harmônico. A inclinação do espectro é normalmente medida em dB/oitava, onde uma oitava é o dobro (ou a metade) da frequência. A inclinação do espectro é relacionada com a qualidade (timbre) do som. Há muito tempo foi estabelecido que sons com muitas altas frequências são percebidos como metalizados e aqueles com poucas altas frequências são percebidos como flautados (assoviados)[1].

Embora a produção da voz não especificamente signifique articulação da fala, toda vocalização humana requer alguma configuração do trato vocal. A menos que um locutor esteja deliberadamente sussurando ou tossindo, a configuração é normalmente com o trato vocal aberto, como nas vogais. As propriedades acústicas das vogais têm sido tradicionalmente descritas na teoria fonte-filtro das vogais[42].

3.7.1 Banda de frequências formantes

O trato vocal, constituído pelas cavidades oral e nasal, pode ser modelado por um filtro LIT (Linear Invariante no Tempo) apenas com pólos [15]. Quando ocorre qualquer mudança de disposição geométrica no trato vocal, é gerado um modelo diferente de cavidade ressonante e, consequentemente, são alteradas as frequências de ressonância, chamadas de formantes, designadas de F_1 a F_n, ordenadas do menor para o maior valor, conforme mostrado na Fig 3.7.

É interessante ressaltar que a maior parte dos fonemas é percebida pois sua resposta em frequência é denotada pela convolução do filtro que representa o trato vocal com o filtro que representa a glote. Espectralmente, o movimento de lábios, dentes, língua e maxilares torna a fonte colorida[15], ou seja, variando em ampla faixa do espectro. Sob o aspecto pericial, diferentes locutores apresentam diferentes configurações geométricas do

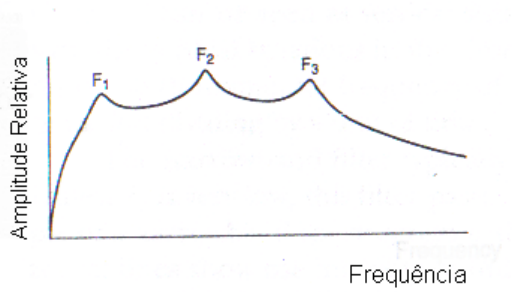


Figura 3.7: Gráfico das frequências formantes da voz humana

Fonte: Livro "Principles of Voice Production"[1].

trato vocal para produzir um mesmo fonema e isso pode auxiliar em sua identificação. Os formantes de maior ordem normalmente F_3 e F_4 são mais dependentes da geometria do trato e, portanto, são de maior valia para reconhecimento de locutor [15].

Uma forma simples de quantificar o significado da ressonância resultante da perda de energia é definir a banda de frequência de formante. Dois pontos são identificados na inclinação da curva de ressonância, onde a resposta é 3 dB menor que o pico. A diferença de frequências entre os pontos de 3 dB define a banda de frequência de formante. Percebe-se que formantes com faixas estreitas possuem sons "metálicos" e formantes com faixas largas possuem som "abafado".

Diferenças fisiológicas no tamanho da laringe, tamanho do trato vocal e composição do músculos são combinados com diferenças sociais (personalidade e atividades profissionais) e afetam sensivelmente a geração da voz humana.

Uma das mais importantes variáveis acústicas para classificação da voz é a frequência fundamental F_0. Em termos gerais, a F_0 do som produzido por um ser humano é inversamente proporcional ao seu tamanho.

Bebês choram com F_0 próximo dos 500 Hz, crianças falam na faixa de 250 a 400 Hz.

A média de F_0 das mulheres adultas é de 200 Hz e os homens por volta de 125 Hz. Parece intuitivo o desenvolvimento de uma relação empírica entre alguns aspectos do tamanho do corpo e F_0. Ao examinarmos as dimensões próximas da fonte da voz, parece razoável admitir que o tamanho da laringe afeta a frequência fundamental mais decisivamente. A frequência fundamental da voz é baseada no comprimento das cordas vocais(L), de acordo com a equação:

$$F_0 = \frac{\sqrt{\sigma/\rho}}{2L} \tag{3.1}$$

Onde: σ = Densidade do tecido e ρ = Tensão longitudinal.

Baseado nesta fórmula, F_0 é inversamente proporcional ao comprimento das cordas vocais e diretamente proporcional à raiz quadrada da tensão longitudinal.

A classificação da voz é concernente não somente com a média de F_0, mas também com a qualidade da voz (timbre). O tamanho do trato vocal necessita ser incluído em um esquema de classificação completa. As chamadas vozes escuras, ou seja, mais graves de acordo com o tipo de timbre de voz, tendem a ser classificadas diferentemente das vozes com brilho apesar de suas faixas de F_0 se sobreporem.

Existe uma relação inversa entre as frequências formantes e o comprimento do trato vocal. Para uma dada laringe, uma voz pode parecer mais escura (grave) se o trato vocal for longo.

A percepção de sons da voz é algumas vezes reduzida a quatro fatores: pitch, loudness, vogais (ou consoantes vozeadas) e qualidade. O último destes fatores, a qualidade, é um termo definido pobremente que inclui todo o restante de percepções após pitch, loudness e categorias fonéticas que têm sido identificadas. Aspereza, voz cansada, chiada, nasalizada e registro são somente umas poucas descrições que se referem a qualidade da voz. Uma

destas qualidades é o registro, sendo percebidamente a mais saliente e pedagogicamente uma das mais problemáticas. Quando se canta com facilidade de uma nota grave até uma aguda sem esforçar as cordas vocais, canta-se no registro médio da voz. Quando a voz ocasionalmente "quebra", este fenômeno é atribuído a uma mudança repentina no registro. Por exemplo, no estilo de canto *yodel*, caracterizado por modificações de tonalidade entre a voz normal e um *falsete*, a "quebra" é intencionalmente exagerada e cultivada como uma forma de arte. Variações no registro médio da voz são observados na fala e no canto.

Capítulo 4

A voz em sistemas de telecomunicações

4.1 Tipos de transmissão

4.1.1 Transmissão analógica

A rede telefônica iniciou sua jornada com a transmissão analógica, com canais de voz analógicos e usando multiplexação em frequência (FDM - *Frequency Division Multiplexing*), onde cada canal ocupa uma largura de faixa de 4KHz.

4.1.2 Transmissão digital

Para que o sinal de voz analógico possa trafegar em uma rede digital, ele necessita ser convertido. Para convertê-lo em um sinal digital, foi adotada a técnica denominada PCM (*Pulse Code Modulation*) - Modulação por Código de Pulso. Esta técnica baseia-se no princípio de amostragem do sinal analógico, seguido da quantização (ajuste e definição do valor) e representação na forma binária do sinal amostrado.

De acordo com o teorema de Nyquist, quanto maior o número de amostras, maior será a fidelidade na recuperação do sinal original no seu destino. O referido teorema permite correlacionar a taxa efetiva de dados a serem transmitidas em um canal a partir da banda passante deste mesmo canal, ou seja, se um sinal é transmitido através de um canal com largura de banda W (Hz), o sinal pode ser reconstituído pelo receptor através da amostragem do sinal transmitido, a uma freqüência no mínimo igual 2W vezes por segundo.

A capacidade do canal"C", medida em bit por segundo (bps), é dada por:

$$C = 2Wlog_2L \tag{4.1}$$

onde "W"é banda passante, medido em Hertz e "L"é o número de níveis utilizados para representar o sinal.

A transmissão digital apresenta como vantagens: a facilidade de regeneração, em comparação com sistemas analógicos; maior imunidade à interferência; circuitos mais confiáveis e baratos. Além disso, a combinação de sinais usando TDM (*Time Divison Multiplexing*) é mais simples que a combinação de sinais analógicos usando FDM (*frequency division multiplexing*). Como desvantagem, a transmissão de um sinal digital requer maior largura de banda.

4.1.3 A conversão analógica - digital

Inicialmente, para que a voz seja trasmitida em um sistema de telecomunicações, a energia sonora é convertida em energia elétrica, através de um transdutor, como por exemplo, um microfone.

Esse sinal elétrico resultante é um sinal analógico, ou seja, pode assumir qualquer valor ao longo do tempo. Porém, devido à introdução da tecnologia digital, há de se converter o sinal analógico em digital. Isso é feito através de um circuito eletrônico chamado conversor Analógico-Digital.

A informação $m(t)$ gerada pela fonte analógica é uma função contínua do tempo e não se encontra em uma forma adequada para ser transmitida digitalmente. Se a fonte é tal que a sua frequência é $f \geq \frac{1}{2T_a}$, podemos, de acordo com o teorema da amostragem, representar o sinal analógico $m(t)$ por suas amostras $m(\ell T_a)$, ℓ inteiro, igualmente espaçadas de T_a segundos.

Ainda assim esta informação, as amostras $m(\ell T_a)$, não se encontra em uma forma adequada para serem transmitidas através do sistema de comunicações digital, já que $m(\ell t)$ pertence ao conjunto dos Reais, não enumerável, e irá exigir um número infinito de bits para representar cada número.

A alternativa é realizar um processamento denominado *quantização* que consiste em limitar o número de possíveis valores que o usuário irá receber. Vamos designar este conjunto de valores por $A = \alpha_0, \alpha_1, ..., \alpha_{j-1}$, onde α_i é inteiro. Desta forma, quando o valor $m(\ell T_a)$ é gerado pela fonte, a informação enviada ao usuário é quantizada e o índice i_ℓ correspondente ao valor $\alpha_{i_\ell} \in A$ que mais se aproxima da amostra $m(\ell T_a)$ produzida pela fonte e é a informação que será enviada. Vamos denominar o processamento que realiza o mapa $m(\ell T_a) \rightarrow \alpha_{i_\ell} \rightarrow i_\ell$ por DIGITALIZADOR.

Supondo que $|A| = J$ e que $J = 2^L$, cada índice a ser enviado pode então ser representado com L bits. O conversor analógico digital, de forma bem simplificada é então formado pela concatenação do AMOSTRADOR, QUANTIZADOR e DIGITALIZADOR.

A cadeia de processamento que, nas premissas do transmissor, transforma:

1. $m(t)$ (sinal contínuo no tempo) na sequência $m(t_\ell)$ (sinal discreto no tempo), 2. A sequência $m(t_\ell)$ na sequência i_ℓ e, subsequentemente, 3. Esta última, na sequência de bits d_ℓ na entrada do modulador[68].

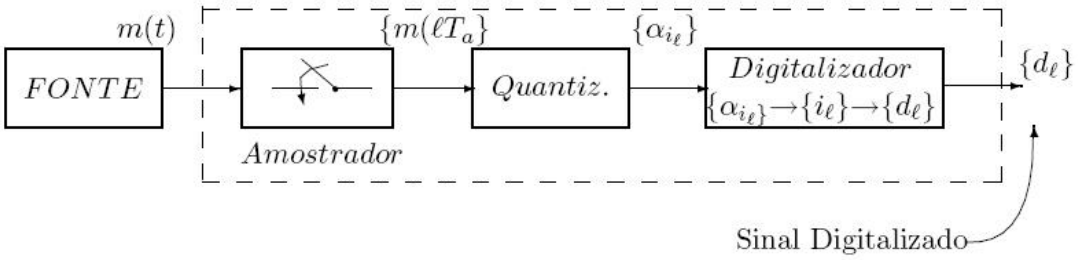


Figura 4.1: Esquema gráfico da conversão analógico/digital

Fonte: Representação digital de sinais analógicos[68].

O número de representações diferentes usados para codificar o sinal também influencia na qualidade da voz digitalizada. Por exemplo, com um número de representações maior, teremos uma melhor fidelidade do sinal digitalizado ao ser feito o processo inverso. Essas representações são limitadas pelo número de bits usados para representar as amostras. A qualidade do sinal pode ainda ser melhorada se utilizarmos intervalos não uniformes de quantização. Usando uma quantificação linear, os sinais de maior amplitude são quantizados com o mesmo intervalo que os sinais de mais baixa amplitude, ou seja, o espaço de codificação foi mal utilizado[63]. Por isso são utilizadas curvas contínuas de supressão logarítmica que obedecem a duas leis: a Lei μ, nativa do padrão T1, usada na América do Norte e Japão e a Lei A, nativa do padrão E1, usada na Europa e no resto do mundo.

Por exemplo, para um sinal com valor 1 podemos dizer que sua representação digital é 00000001, usando oito bits. Qualquer forma de compressão dos dados a serem transmitidos,

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

Neste ponto, devem ser gerados os pacotes de dados e enviados pela rede. Em primeira instância, teremos de esperar até que a informação binária de saída do conversor alcance o tamanho definido para a parte de dados do pacote. Essa informação de voz, pode ser definida como *frames* de voz e será encapsulada em três "envelopes"no protocolo RTP (*real-time protocol*), utilizando o protocolo de transporte UDP (*user datagram protocol*). Em seguida, na camada de rede, o endereço de rede IP (*Internet Protocol*) é associado ao pacote no processo de encapsulamento e assim, ele está pronto para ser enviado.

4.2 Codificadores de voz

O objetivo de todos os sistemas de codificação de voz é transmitir a voz com a maior qualidade possível empregando a mínima capacidade de canal, porém, no ambiente de Telecomunicações, a necessidade de economia de banda é uma realidade.

Há cerca de 60 anos atrás, no início da pesquisa na área da codificação da voz, a intenção dos pesquisadores era o desenvolvimento de um sistema que possibilitasse a transmissão de voz através da estreita largura de banda dos cabos telegráficos. Pioneiro neste trabalho, Homer Dudley, do *Bell Telephone Laboratories*, demonstrou a redundância existente nos sinais de voz e criou o primeiro método de codificação de voz. A idéia básica por trás do codificador de voz era a de analisar a voz em termos da sua frequência fundamental e do seu espectro, e sintetizá-la pela excitação de um banco de filtros passa-banda analógicos. A implementação original de Dudley comprimia as formas de onda em

sinais analógicos com uma banda passante total de 300 Hz [62].

De tempos em tempos, novos codificadores são apresentados com melhor qualidade, utilizando menores taxas. Atualmente, com o crescimento das aplicações da tecnologia de voz sobre IP, a utilização dos codificadores têm crescido enormemente, seja no ambiente corporativo ou para o usuário comum através da Internet.

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

A diferença básica entre a codificação em formato de onda e a codificação paramétrica está na filosofia básica utilizada para efetuá-la. Codificadores em formato de onda transmitem aproximações do sinal de entrada, estes codificadores tentam reproduzir o sinal original, amostra por amostra, com base nas suas características estatísticas, espectrais ou temporais; enquanto que os codificadores paramétricos realizam a digitalização das principais características do sinal analógico e transmitem uma série de parâmetros resultantes das operações anteriores.

Em linhas gerais, o resultado da comparação das soluções adotadas é um tanto intuitivo: os codificadores em formato de onda, por transmitirem amostras do sinal de voz original, acabam por apresentar uma qualidade muita boa. Porém, devido à ausência de

compressão do sinal, as larguras de banda utilizadas para transmissão apresentadas por tais codificadores são bastante grandes. Em contrapartida, os codificadores paramétricos, por realizarem uma manipulação matemática do sinal de voz a ser transmitido e transmitirem apenas os parâmetros resultantes de tal manipulação apresentam larguras de banda para transmissão muito reduzidas, mas geram no receptor sinais de qualidade precária.

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

4.2.1 Codificadores por forma de onda

Estes codificadores tentam reproduzir o sinal original, amostra por amostra, com base nas suas características estatísticas, espectrais ou temporais. Os codificadores por forma de onda são de baixa complexidade e produzem um pequeno retardo na voz além de produzirem um sinal de alta qualidade, podendo citar como exemplo de codificador por forma de onda a Modulação por Código de Pulso (*Pulse Code Modulation - PCM*).

Tais codificadores utilizam para a codificação do sinal de entrada propriedades de caráter temporal e/ou espectral do mesmo. Porém, isso não pressupõe analisar o que realmente o sinal está representando, o que significa que tais codificadores apenas se concentram na necessidade de reconstruir a forma de onda a ser codificada. Dessa maneira, qualquer tipo de sinal pode vir a ser codificado utilizando este método. Entre tais codificadores podemos citar o PCM e o ADPCM (*adaptative pulse code modulation*). Tais codificadores apresentam uma qualidade muito boa, mas acompanhada de grande largura de banda no processo de transmissão/recepção.

O ideal é que se procurem algumas formas de codificação de sinais de voz que ocupem a menor largura de banda possível. Para realizarmos tal operação com os codificadores por forma de onda, exploramos as características estatísticas dos sinais de voz e também as características do sistema auditivo humano. Mais especificamente, devemos ter dois objetivos em mente: A máxima remoção de redundâncias do sinal de voz e a alocação dos bits disponíveis para codificar a parte não-redundante do sinal de voz.

Para uma mesma qualidade de sinal de voz, cada vez que dividimos por dois o número de bits a serem utilizados, a complexidade computacional cresce aproximadamente de uma ordem de grandeza. Ocorre o mesmo quando tentamos baixar as taxas de codificação, os sistemas utilizados para remoção de redundância e eficiente atribuição dos bits passam a ter sua complexidade computacional muito elevada[26].

O codificador G.711

O G.711 [27] foi aprovado no início da década de 70, mais precisamente em 1972. Ele consiste na codificação PCM e apresenta uma taxa de codificação de 64 kbits/s. A norma não definiu um escopo preciso, mas a mesma foi concebida de modo a ser uma forma de digitalizar o sinal de voz para o mesmo ser tratado da forma mais eficiente possível pelos sistemas de comunicação digital.

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

Como sabemos, o princípio de Nyquist deve ser levado em consideração em qualquer sistema que sofra amostragem de seus dados contínuos. Dessa maneira, levando-se em conta

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

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

O codificador G.726

O G.726[28] é uma recomendação que data de dezembro de 1990. O seu princípio de funcionamento se baseia na utilização da codificação ADPCM, para a redução do número de bits a serem codificados e transmitidos no canal. O funcionamento do G.726 pode ser ilustrado através da Fig 4.2, que mostra os blocos de codificação e de decodificação do sistema. Primeiramente, verificamos que o sinal PCM de entrada é convertido, através da

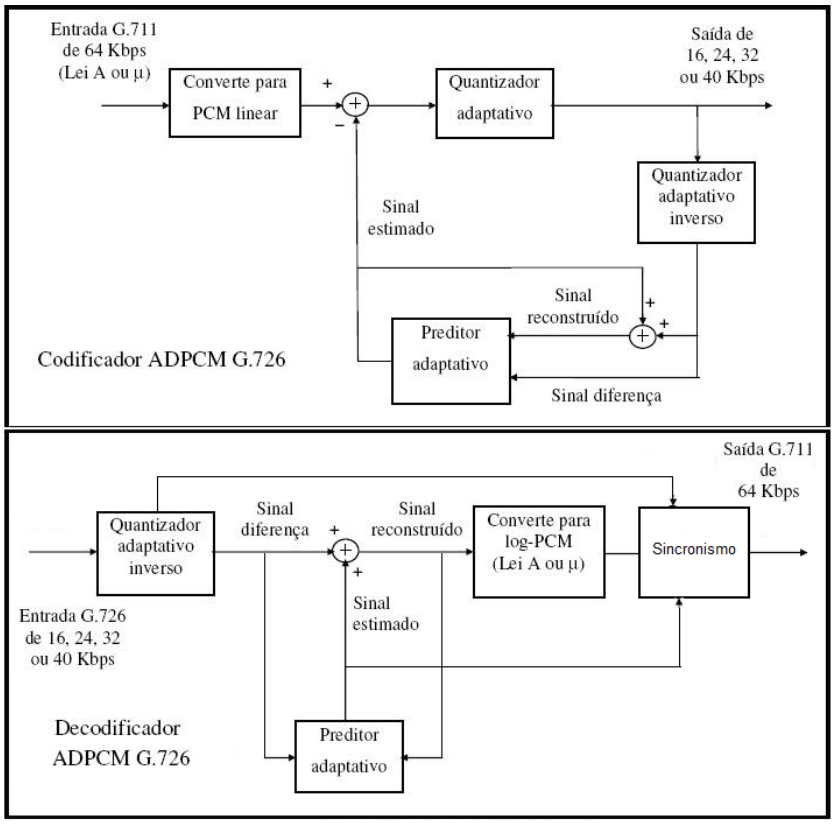


Figura 4.2: Codificador e decodificador G.726

Fonte: Recomendação ITU-T G.726 [28].

Lei A ou da *Lei* μ, para PCM uniforme. Em seguida, o sinal convertido é inserido em uma malha de realimentação e sofre os efeitos combinados da predição e da quantização adaptativa e em seguida é transmitido.

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

o G.726 foi concebido para realizar a codificação ADPCM com diversas taxas de codificação, dependendo da aplicação a ser utilizada.

Podem ser utilizados de 2 a 5 bits para a quantização, sendo que sempre 1 bit é reservado para o sinal da amplitude. Dessa maneira, taxas de 16 a 40 kbits/s podem ser utilizadas. O valor mais utilizado é de 4 bits para quantização, que é identificado como G.726 com taxa de compressão 2 para 1, ou de 32 kbits/s.

4.2.2 Codificadores Paramétricos

A voz humana é um sinal não estacionário, mas se levarmos em consideração pequenos segmentos de voz com duração entre 10 e 30 ms, podemos considerar a voz estacionária por partes[64]. Para estes segmentos pode-se modelar o processo de geração da voz humana. Nos codificadores paramétricos, também chamados de *vocoders*, a informação transmitida são os parâmetros obtidos no modelo, que são atualizados de tempos em tempos e determinados com a segmentação do sinal inicial em intervalos periódicos chamados quadros, onde o sinal de voz pode ser considerado estacionário. Os vocoders operam com baixas taxas de transmissão, quase sempre inferiores a 4 kbps, com atrasos e complexidade elevados e baixa qualidade da voz digitalizada, soando de forma sintética.

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

Dessa maneira, ao contrário dos codificadores de forma de onda, os codificadores

paramétricos fazem uso de aspectos específicos, intrínsecos à voz humana, o que os torna específicos para cada tipo de sinal a ser codificado. Entre tais codificadores podemos destacar o LPC (do inglês Linear Predictive Coding).

A Codificação LPC

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

Nesse caso, processos estocásticos estacionários podem ser interpretados como a saída de um filtro digital onde a entrada é um modelo para o ar que sai dos pulmões e atravessa as cordas vocais e o filtro propriamente dito é dado pelo trato vocal (trecho que vai desde a glote até os lábios, passando pela epiglote, laringe, faringe, garganta, língua e dentes). Uma vez definidos os elementos que compõem o nosso sistema (entrada, sistema propriamente dito e saída) podemos classificar os sinais de voz (sinais de saída) em dois tipos: sinais vozeados e sinais não-vozeados. Os trechos vozeados são aqueles que, quando emitidos, geram vibração das cordas vocais, sendo os mesmos modelados por um trem de pulsos praticamente periódico.

O período relacionado com a sua frequência fundamental é denominado período de *pitch*, um parâmetro determinante no projeto de codificadores atuais. Tais sons podem ser exemplificados através daqueles relativos às vogais. Os sons não-vozeados são aqueles que não geram vibração das cordas vocais. Dessa maneira, são modelados como ruído branco, apresentando como característica marcante a taxa de cruzamento por zero significativa. Sons não-vozeados podem ser exemplificados como som da letra “s”na palavra “sapo”ou

como o som do encontro consonantal “ch”na palavra “chocalho”.

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

Para a geração de sons sonoros no codificador, utiliza-se um trem de pulsos com período igual ao período de *pitch* da voz. Para sons não-vozeados utilizamos como excitação uma sequência de ruído branco, sem nunca deixarmos de estimar um ganho para o modelo. Dessa maneira, deixamos de transmitir as amostras do sinal de voz quantizadas, passando a ser transmitidos os coeficientes determinados pela análise LPC, um sinal de *flag* que indica se o som é sonoro ou surdo, o ganho estimado e, caso o som seja sonoro, o *pitch*. Com isso, passamos a transmitir uma quantidade extremamente inferior de parâmetros com relação aos codificadores de forma de onda, o que diretamente resulta em uma taxa de codificação bem inferior. Se considerarmos o PCM padrão, cuja frequência de amostragem é de 8 kHz, e janelas de 20ms, teríamos que transmitir 160 valores por janela, enquanto que no LPC, supondo um filtro de ordem 10, transmitiríamos 13 valores (10 correspondentes aos coeficientes, 1 ao ganho, 1 ao *flag* e 1 ao *pitch*)[69].

No entanto, apesar de esse tipo de codificador apresentar uma redução de taxa de codificação, ele possui uma grande restrição. Como existe uma decisão do tipo binária com relação ao sinal ser ou não vozeado, não há espaço para transmissão de sons intermediários. O resultado disso é o sinal codificado de voz ser um tanto “robótico”. Porém, sua baixa taxa de codificação é uma grande vantagem. Como os codificadores por forma de onda apresentam um comportamento antagônico em relação aos vistos nessa seção, devemos

procurar sistemas que reúnam e maximizem os pontos fortes de cada uma das famílias. Tais sistemas se encontram nos codificadores híbridos.

Codificador Residual Excited Linear Prediction (RELP)

O codificador RELP funciona quase da mesma forma como o codificador LPC. Para analisar o sinal, os parâmetros do filtro do trato vocal são determinados e o inverso do resultado do filtro é aplicado ao sinal. Isto nos dá o sinal residual. O codificador LPC então verifica se o sinal foi falado ou mudo e utiliza esta informação para modelar o sinal de excitação. No codificador RELP no entanto, o residual não é analisado de forma alguma, mas vai ser utilizado diretamente como a excitação para o sinal de voz síntese. O resíduo é compactado usando técnicas de codificação de forma de onda para reduzir os requisitos de banda. Codificadores RELP podem proporcionar boa qualidade de voz com taxas na região de 9,6 kbps[59].

Codificadores *Multi-Band Excitation* (MBE)

Os codificadores MBE são codificadores no domínio da frequência que incorporam uma inovação para melhorar o modelo de excitação. O modelo de excitação é misto, permitindo que ambas componentes, harmônicas e aleatórias, participem de um mesmo quadro da fala. Para sinais sonoros da fala, uma sequência periódica de impulsos de excitação corresponde no domínio da frequência a uma sequência periódica de impulsos no domínio da frequência, espaçadas em harmônicas do *pitch*.

O modelo MBE divide o espectro e as sub-bandas em múltiplos da frequência fundamental (ou frequência de *pitch*). A maneira na qual o *vocoder* MBE representa as informações de frequência do trato vocal pode ser pensada como um *vocoder* de canal que

tem todos os canais centrados em harmônicas da frequência fundamental (ou frequência de *pitch*). O modelo MBE permite separar a decisão sonoro/surdo para cada canal (ou grupo de canais) em cada quadro do sinal da fala. Isto permite uma representação do sinal de excitação mais fiel do que a decisão simples dos *vocoders*. Um codificador *Improved Multi-Band Excitation coder* (IMBE) operando a uma taxa de 4,15 kbits/s foi selecionado pelo Inmarsat como padrão para a comunicação de voz por satélite[60].

4.2.3 Codificadores híbridos

Os codificadores híbridos são os de maior complexidade, trabalham com taxas de transmissão de 4 a 16 kbps. Estes codificadores são baseados nos modelos de produção de voz e utilizam uma excitação mais apurada para o filtro de síntese.

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

A técnica mais usada neste tipo de codificador é a CELP (do inglês Code Excited Linear Prediction). Esse sistema reúne os mesmos princípios da aná lise LPC para redução de parâmetros a serem transmitidos. Porém, como característica dos codificadores híbridos, o CELP lança mão de um artifício para anular a desvantagem do codificador LPC. De fato, para manipular as excitações da entrada do sistema, o CELP faz o uso de dicionários, tanto

fixos quanto adaptativos (o que será explicitado posteriormente). Sendo assim, um número bem maior de excitações (com relação ao LPC) pode ser utilizado, o que significa que uma gama maior de sinais de voz pode ser reconstituída. Isso torna o sistema bastante mais completo e capaz de gerar um sinal de voz (após a transmissão) de qualidade bem superior àquele gerado pelo LPC. Existe uma série de codificadores baseados na técnica CELP disponíveis no mercado e na literatura, tal grande é a eficiência desse tipo de codificação.

Como já foi citado anteriormente, a voz humana possui uma natureza extremamente não estacionária, o que nos leva à necessidade de segmentá-la, ou melhor, que sejam considerados intervalos relativamente pequenos do sinal de voz, individualmente. No caso do sistema considerado, as janelas possuem 20 ms, o que, a uma taxa de amostragem de 8 kHz, representa 160 amostras, sendo utilizada a janela de Hamming [29]. Por motivos a serem explicados mais adiante, cada um dos blocos de 20 ms será dividido em quatro sub-blocos de 5 ms cada. Com a análise LPC, são calculados (para cada uma das janelas de 20 ms) os coeficientes do filtro digital que melhor representa o trato vocal do usuário no instante em questão.

O Filtro de Síntese

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

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{p} a_i z^i} \tag{4.2}$$

A constante p tem como função representar a ordem do modelo LPC usado e identifica a precisão com a qual o trato vocal será modelado através de H(z). Normalmente, os codificadores CELP utilizam $p = 10$, o que estabelece um bom compromisso entre a qual-

idade do sinal de saída e a taxa de codificação[69]. Os valores representam os coeficientes de predição linear, obtidos através da análise LPC.

O Filtro de Ponderação (ou Perceptivo)

Um estudo realizado de acordo com [69] verificou que o ouvido humano é mais sensível a erros e ruídos em componentes de mais baixa amplitude do que em componentes de alta amplitude, no domínio da frequência. Isso nos induz a dar mais importância às componentes de menor amplitude quando calculamos o erro contido no processo de análise por síntese.

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

$$W(z) = \frac{A(z)}{A(\frac{z}{y})} \tag{4.3}$$

Onde y representa o fator de percepção, que possui valor típico de 0,8. No entanto, nos sistemas CELP, é utilizada uma cascata entre o filtro perceptivo e o filtro de síntese, com a intenção de acelerar o processamento.

Os Dicionários

Dicionários (ou, em inglês, "codebooks") são conjuntos de excitações, ou amostras do sinal de entrada, para o filtro de síntese, que podem ser descritos da seguinte forma:

$$C = X_0(n), X_1(n), X_2(n), ..., X_{k-1}(n) \tag{4.4}$$

Esse modo de representação indica que o dicionário armazena K sequências $x(n)$, onde n indica o índice da sequência.

No caso do CELP, geralmente são utilizados dois dicionários, um com as excitações fixas e o outro com as excitações adaptativas. Em geral, o dicionário fixo tem as suas sequências geradas aleatoriamente, e para reduzirmos os cálculos, as amostras das mesmas sofrem um processo de *clipping* para simplificação dos cálculos, o que significa que as amostras abaixo de um determinado valor são zeradas. Já o dicionário adaptativo tem as suas sequências atualizadas a cada bloco de 5ms, o que é realizado através de uma malha de realimentação. Essas atualizações se dão baseadas na soma das melhores excitações dos dicionários fixo e adaptativo. A cada sub-bloco, a melhor excitação encontrada é incluída no dicionário adaptativo que tem seu conteúdo mais antigo descartado, estando, este dicionário, inicialmente zerado por convenção. Finalmente, cada um dos dicionários deve ter o seu ganho calculado, sendo que ambos podem ser calculados da mesma maneira, segundo a fórmula a seguir:

$$G = \frac{Corr.(Sinal - Resposta.Resposta - Dic)}{Corr.(Resposta - Dic.Resposta - Dic)} \tag{4.5}$$

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

Análise por Síntese

A excitação que deve ser utilizada para reconstituir a voz do usuário na saída é determinada por um processo chamado "Análise por Síntese". Tal processo é realizado a cada

Parâmetro	Faixa	Número de Bits
Ganho do Dicionário Fixo (Gf)	-0,05 a 0,05	5 bits X 4
Índice do Dicionário Fixo (I)	0 a 511	9 bits X 4
Ganho do Dicionário Adaptativo (Ga)	0 a 2	4 bits X 4
Índice do Dicionário Adaptativo (L)	0 a 511	9 bits X 4
Coeficientes do Filtro de Síntese	—-	32 bits
Total	—-	140 bits

Tabela 4.1: Características do sistema CELP[69]

sub-bloco de 5 ms, uma vez que as características da excitação variam mais rapidamente que as características do trato vocal, o que impõe esse requisito ao sistema [69].O procedimento é o seguinte:

- Para cada sub-bloco, cada uma das excitações contida no dicionário passa pelo filtro calculado pela análise LPC, gerando uma determinada resposta;

- Subtrai-se essa resposta do sinal de voz na referência;

- Armazena-se a norma quadrática da diferença obtida no passo anterior;

- A excitação que gerar a menor norma quadrática da diferença será escolhida para reconstituir o sinal na saída.

No sistema CELP, são transmitidos apenas o índice da melhor excitação, o ganho do filtro e seus respectivos coeficientes, ao invés de todas suas amostras quantizadas, como por exemplo, no sistema PCM. A alocação dos bits está demonstrada pela Tabela 4.1.

Assim, podemos verificar um ganho promovido pelo uso do sistema CELP, uma vez que a taxa de codificação é diminuída dos 64 kbits/s do PCM para 7 kbits/s.

O codificador G.728

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

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

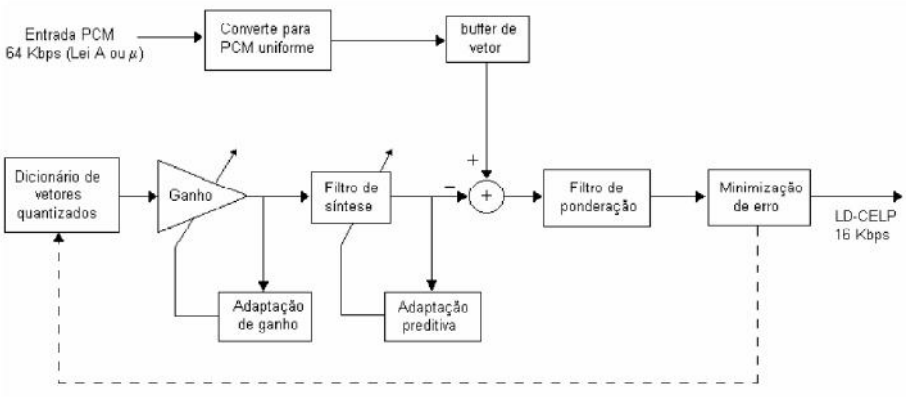


Figura 4.3: Diagrama de blocos simplificado da codificação G.728

Fonte: Recomendação ITU-T G.728 [29].

Com auxílio da Fig 4.3 podemos entender melhor como tal recomendação funciona. Inicialmente, temos uma conversão do sinal de entrada de PCM *lei A* (ou *lei* μ) para PCM uniforme. Em seguida, o mesmo é agrupado em blocos de cinco amostras consecutivas. Cada um desses blocos é submetido a uma comparação com cada um dos 1.024 vetores armazenados no dicionário de vetores quantizados, tendo estes vetores atravessado

as unidades de ganho e o filtro de síntese. A comparação que apresentar o menor erro indicará qual índice do dicionário a ser transmitido. Dessa maneira, sabendo que temos 8.000 amostras por segundo, que as mesmas são agrupadas em grupos de 5 amostras e que o índice do dicionário é formado por 10 bits, chegamos a uma taxa de codificação de 16 kbits/s.

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

O codificador G.729

O G.729 é uma recomendação posterior ao G.728 e também utiliza uma variante do sistema CELP tradicional. A mesma foi aprovada em 19 de março de 1996, lançando mão da codificação CS-ACELP (Conjugate-Structure Algebraic-Code Excited Linear Prediction) e apresenta uma taxa de codificação duas vezes menor que o G.728, ou seja, de 8 kbits/s. Tal recomendação foi concebida para transmitir sinais de voz com qualidade em ambientes onde baixas taxas de codificação são de extrema importância, como por exemplo, aplicações de comunicação sem fio e circuitos transoceânicos. Com relação ao atraso, o G.729 codifica os sinais de áudio em janelas de 10 ms, que é sempre sucedido de um tempo de *look-ahead* de 5 ms, o que resulta em um atraso total de 15 ms para essa recomendação.

O codificador em questão foi projetado para operar com sinais de entrada obtidos através da filtragem de sinais analógicos, conforme exposto na recomendação G.712 do ITU. Depois de uma discretização à taxa de 8.000 amostras em um segundo, o sinal é

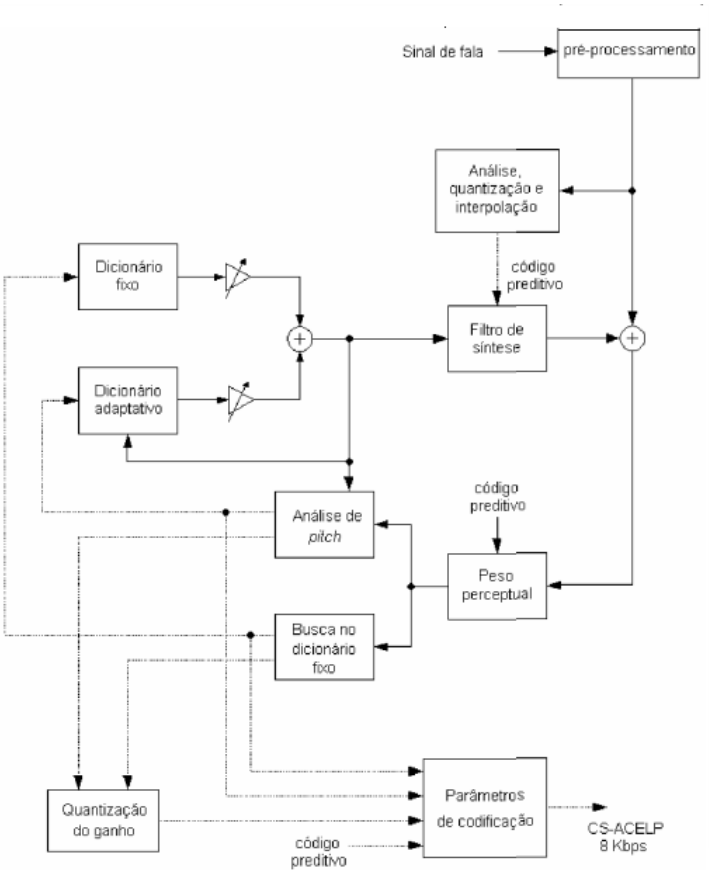


Figura 4.4: Diagrama de blocos da codificação G.729

Fonte: Recomendação ITU-T G.729 [30].

convertido em PCM linear de 16 bits, servindo assim de entrada para o codificador (na saída realiza-se a operação inversa).

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

na taxa de codificação de 8 kbits/s.

Os filtros do CODEC G.729 utilizam coeficientes gerados pelo método de autocorrelação com janelas de observação de 30ms. A cada 80 amostragens a 10ms os coeficientes são atualizados e é feito o deslizamento da janela, realizando assim a análise destes valores levando em conta as 120 amostras dos 4 quadros passados, as 80 amostras do quadro atual e 40 amostras do próximo quadro, onde se pode observas os 5ms de *look-ahead*.

Apesar do G.729 ser uma recomendação extremamente eficiente em termos de taxa de codificação, seus requisitos computacionais são extremamente elevados, inclusive superiores aos do G.723.1, que será visto a seguir. Dessa maneira, em maio de 1996 foi lançada o Anexo A da recomendação, mantendo a operabilidade com o G.729 original e reduzindo a sua complexidade computacional. A operação do G.729A é bastante similar à do G.729, sendo as principais alterações referentes à forma de busca nos dicionários e à forma de operação de cada um dos filtros. Coloquialmente, o G.729 ficou conhecido como G.729 de baixa complexidade. Em outubro de 1996, foi homologado o Anexo B do G.729. Tal anexo descreve o gerador de ruído de conforto e o detector de voz, utilizados na implementação da compressão de silêncio, tanto no G.729 quanto no G.729A.

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

O codificador G.723.1

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

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

Com relação ao atraso, independentemente de qual das duas velocidades está se usando, são necessários 30 ms para a formação de cada uma das janelas, além de 7,5 ms de *look-ahead*, gerando um atraso total de 37,5 ms. O principal requisito dessa recomendação é garantir que as duas velocidades (taxas) estejam disponíveis em qualquer momento. Para tanto podem ser usadas duas operações. A primeira consiste em simplesmente trocar as velocidades entre um quadro e outro, enquanto a segunda baseia-se em utilizar os períodos de descontinuidade de transmissão (nos intervalos de silêncio) para efetuar a troca.

No que diz respeito ao processo de codificação do sinal de entrada, o procedimento adotado é idêntico ao do G.729 e explicitado anteriormente.

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

No que diz respeito ao processo de codificação propriamente dito é bastante similar

àquele descrito na recomendação G.729. No caso do G.723.1, uma janela deslizante de observação com tamanho de 180 amostras é centrada em cada um dos sub-quadros. Dessa maneira, quatro grupos de parâmetros serão gerados; tais parâmetros (referente aos valores de ganho, índices de dicionários e coeficientes de filtros utilizados) serão agrupados, codificados e transmitidos. Quando chegamos ao último sub-quadro de um determinado quadro, a janela considera também o primeiro sub-quadro do quadro subsequente. Como cada sub-quadro contém 60 amostras e cada uma destas leva 125 μs, chegamos a um atraso de *look-ahead* de 7,5 ms. Na codificação MP-MLQ, que possui a maior banda, são transmitidos 189 bits por quadro, enquanto na codificação ACELP são 158 bits por quadro. Dessa maneira, as taxas de codificação são respectivamente 6,3 kbits/s (189 x 8000 / 240) e 5,3 kbits/s (158 x 8000 / 240).

O codificador iLBC

O CODEC iLBC (*Internet Low Bit Rate Codec*)[11] [12] foi desenvolvido pela Global IP Sound (GIPS), sendo projetado para banda estreita e resulta em um fluxo de bits empacotados de 13,33 Kbps com quadros (frames) de 30ms e 15,20 Kbps com quadros (frames) de 20ms. Este CODEC permite boa qualidade na degradação da voz no caso de perda de frames, o qual ocorre em conexões com perda ou atraso de pacotes IP. Quando o CODEC opera com blocos de 20ms, produz 304 bits por bloco (15200 * 0,02 = 304). Similarmente, para blocos de 30ms produz-se 400 bits por bloco (13333 * 0,03 = 400). Os dois modos para diferentes tamanhos de frame operam de maneira muito parecida. O algoritmo descrito resulta em um sistema de codificação de voz amostrado em 8 kHz com uma resposta controlada às perdas de pacotes similares à Modulação por PCM com *Packet Loss Con-*

cealment (PLC), técnica usada para mascarar os efeitos da perda ou discarte de pacotes. O iLBC consiste de um codificador e um decodificador. Para um bloco constituído com 160/240 (20ms/30ms) amostras, as principais etapas são executadas:

(1) - Um conjunto de filtros LPC é computado e o sinal de voz é filtrado através deles para produzir o sinal residual.

(2) - O CODEC usa a quantização escalar da parte dominante, em termos de energia, do sinal residual para o bloco.

(3) - O estado dominante é de amostras com comprimento 57/58 (20ms/30ms) e forma um estado inicial para codebooks (livro de códigos) dinâmicos construídos das partes já codificadas do sinal residual. Estes *codebooks* dinâmicos são usados para codificar as partes restantes do sinal residual.

(4) - Por este método, a independência de codificação entre os blocos é adquirida, tendo por resultado a eliminação da propagação das degradações perceptíveis devido à perda do pacote.

(5) - O método permite que seja realizada Codificação Linear Preditiva (LPC) de alta qualidade.

Se conhecermos uma parcela suficiente de um sinal redundante nós podemos inferir o resto do sinal, ou pelo menos tentar fazer a estimativa mais provável. Em particular, se conhecermos o comportamento passado de um sinal até um determinado ponto no tempo então é possível fazer alguma inferência sobre seus valores futuros. Tal processo de inferência é conhecido como predição. Na Predição Linear uma amostra futura é obtida como uma combinação linear de um conjunto de amostras passadas.

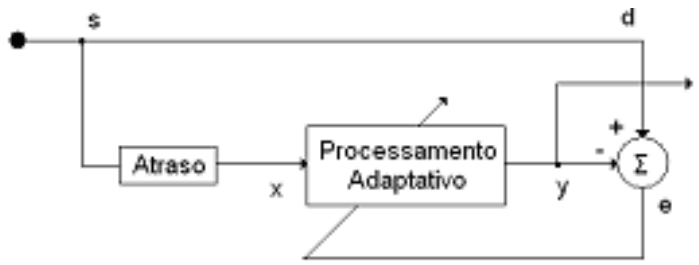


Figura 4.5: Previsão do sinal de entrada através de um Processamento Adaptativo

Fonte: *Paper* "Comparação entre os CODECS de compressão de voz iLBC e G.729". [13]

De acordo com a Fig 4.5, com base num sinal x, que é uma versão atrasada do sinal corrente s, o processador tenta obter uma saída y que se aproxime de d (que é igual a s), minimizando assim e. Por outras palavras, o processador tenta "prever"o sinal de entrada corrente, s.

Codec	Algortmo	Tipo	Banda	Atraso
G.711	PCM	Forma de onda	64 kbit/s	$< 1ms$
G.726	ADPCM	Forma de onda	40/32/24/16 kbit/s	$< 1ms$
RELP	-	Paramétrico	9.6 kbit/s)	4.14
IMBE	-	Paramétrico	4.5 kbit/s)	80 ms
G.728	LD-CELP	Híbrido	16 kbit/s	5 ms
G.729	CS-ACELP	Híbrido	8 kbit/s	10 ms
G.729a	CS-ACELP	Híbrido	8 kbit/s	10 ms
G.723.1	ACELP	Híbrido	5.3 kbit/s	30 ms
G.723.1	MP MLQ	Híbrido	6.3 kbit/s	30 ms
iLBC	-	Híbrido	15.2 kbit/s	37,5 ms

Tabela 4.2: Tabela de codificadores de voz.

4.3 Redes de telefonia

4.3.1 Rede Pública de Telefonia Comutada - RPTC

É o maior sistema eletrônico integrado do mundo. É um sistema muito confiável e com qualidade muito boa.

Os telefones são os equipamentos terminais desse tipo de rede mas também há o FAX e os modens. Eles permitem a interação de seus usuários usando a rede de voz. Para se obter uma linha, ou acesso a essa rede, deve-se fazer uma ligação do local requisitante até a central telefônica local mais próxima. O conjunto dessas centrais interligadas constitui o *backbone* da RPTC.

A rede telefônica é uma rede de comutação de circuitos orientada à conexão na qual nenhuma informação é trocada sem que antes tenha sido estabelecida uma conexão entre origem e destino. Se estes não estiverem na mesma central, é necessário encaminhar o tráfego de voz através do *backbone* até o destino. Geralmente o tráfego do *backbone* é multiplexado, o qual melhora a utilização dos circuitos que interligam as várias centrais.

Junto à essa infra-estrutura existe uma rede de sinalização que estabelece e gerencia as comunicações entre terminais e comutadores de voz ou entre comutadores (faz a ligação da linha de acesso com a linha de destino).

A RPTC garante uma largura de banda reservada para uma determinada ligação e também valores de atraso controlados e dentro de valores definidos devido a utilização de comutação de circuito. A Rede Pública de Telefonia Comutada é também chamada PSTN, do idioma inglês *Public Switching Telecommunnication Network.*

4.3.2 Redes de telefonia voz sobre IP (VoIP)

Antes de falarmos de VoIP teremos de observar a definição de Internet e por conseguinte a de IP(*Internet Protocol*). A Internet pode ser definida como uma rede de computadores mundial, ou seja, uma rede que conecta milhões de equipamentos de computação em todo o mundo. Os equipamentos conectados à Internet: computadores portáteis, pagers, Web TV's etc, são comumente denominados de hospedeiros ou sistema finais. As aplicações, por outro lado, que muitos de nós utilizamos, como *Web* e o e-mail são denominadas de programas de aplicação de rede as quais funcionam nesses sistemas finais.[35]

Para que haja comunicação entre esses sistemas finais na Internet, eles precisam falar a mesma língua, ou seja, precisam usar a mesma regra ou protocolo. Os protocolos definem a maneira de como pode haver envio e recebimento de informação na Internet. O TCP(*Transmission Control Protocol*), o UDP (*User Datagram Protocol*) e o IP são os protocolos mais importantes da Internet.

Os sistemas finais, geralmente, não estão somente conectados entre si por um simples enlace de comunicação. É mais comum que eles estejam ligados indiretamente através de um roteador. Este encaminha a informação que está chegando por um enlace de comunicação para um dos enlaces de comunicação de saída. O protocolo IP especifica o endereçamento da informação que é enviada e recebida entre os roteadores e os sistemas finais. O protocolo TCP controla o fluxo de informações e também é responsável pela verificação do correto recebimento dos pacotes e quando necessário de seu reenvio.

O termo VoIP(Voice over IP) designa a transmissão de voz, em forma de pacotes, através do protocolo IP(Internet Protocol).

A telefonia IP

É preciso esclarecer que o termo voz sobre IP é comumente tido com o mesmo significado de telefonia IP, no entanto, voz sobre IP é mais amplo. A telefonia IP é uma aplicação do VoIP e esta pode ser utilizada igualmente no contexto de videoconferência, audioconferência, transmissão de áudio com alta fidelidade, entre outras diversas aplicações.

Quando falamos de telefonia IP podemos contemplar dois ambientes. No primeiro coexistem a rede telefônica com a rede de dados, a este chamamos de VOIP, trata-se de um sistema híbrido onde coexistem as centrais telefônicas convencionais e a rede de dados. No ambiente VoIP as centrais são interligadas aos *gateways* que fazem a conversão analógica digital e o encapsulamento dos *frames* de voz. A infra-estrutura de cabeamento telefônico e terminais convencionais são totalmente aproveitados. Na telefonia IP pura, somente a rede de dados está presente que, no entanto, presta o serviço de telefonia substituindo a rede telefônica convencional.

O protocolo IP

O Internet Protocol (IP) é designado para o uso em sistemas de computadores interconectados em rede utilizando troca de pacotes. O IP fornece transmissão de blocos de dados, chamados datagramas, entre origem e destino, sendo que origem e destino são computadores ou equipamentos identificados por um endereço de tamanho fixo. Outra função do IP, definido na RFC 791, é a fragmentação e remontagem de datagramas longos, se necessário, para transmissão através de redes que apresentam um desempenho melhor com pacotes menores.

O protocolo IP é especificamente limitado em escopo para fornecer as funções necessárias

para entregar um pacote de uma origem para um destino em uma rede de computadores. Não há mecanismos para aumentar a confiabilidade dos dados, assegurar a sequência dos datagramas, ou outros serviços comumente encontrados em outros protocolos ponto a ponto.

O pacote IP pode ser perdido, reproduzido, atrasar-se ou ser entregue com problemas, mas o serviço não detectará tais condições, nem informará isso ao transmissor nem ao receptor.

O protocolo IP é considerado sem conexão porque cada pacote é independente dos demais. Uma sequência de pacotes enviados de um computador a outro pode trafegar por caminhos diferentes, ou alguns podem ser perdidos enquanto outros são entregues.

O protocolo de transporte UDP

Na maior parte das aplicações que trafegam nas redes de comunicações de dados, o protocolo de transporte utilizado é o TCP. Dentre as várias funções que desempenha, uma delas é garantir a entrega dos pacotes. Em certas situações, o dispositivo de origem não precisa da garantia da entrega de dados ao dispositivo de destino. Nesses casos, o TCP é substituído pelo UDP. No caso da voz a retransmissão até atrapalharia, pois, determinada informação só é valida naquele exato momento, sua repetição dificultaria o entendimento em uma fala contínua. O UDP é um protocolo sem conexão, ou seja, não necessita estabelecer uma conexão entre origem e destino antes de enviar datagramas IP.

Cada mensagem UDP é chamada de segmento e consiste de duas partes: um cabeçalho e uma área de dados. Uma de suas principais funcionalidades é a utilização do conceito de porta de origem e destino. Sem os campos de portas, a camada de transporte não saberia

como diferenciar os pacotes oriundos de diversas aplicações em um mesmo computador. Um programa aplicativo que utiliza o protocolo UDP assume a responsabilidade de lidar com o problema de confiabilidade, inclusive perda de mensagem, duplicação, retardo, transmissão defeituosa e perda de conectividade.

Protocolos de suporte à telefonia IP

Os protocolos RTP/RTCP

Para dar suporte à transmissão multimídia em tempo real na Internet, foi criado o protocolo *Real Time Transport Protocol* (RTP), RFC 3267 do IETF. O RTP implementa o sequênciamento dos pacotes UDP na transmissão da mídia, cada pacote de voz enviado em um fluxo RTP recebe um número uma unidade maior que seu predecessor, para que eles possam ser recuperados no destino na ordem correta, além de permitir ao destino descobrir se algum pacote está faltando. Se um pacote for omitido, a melhor ação que o destino deve executar é fazer a aproximação do valor que falta por interpolação. A retransmissão não é uma opção prática, pois o pacote retransmitido provavelmente chegaria tarde demais para ser útil. Como conseqüência, o RTP não tem nenhum controle de fluxo, nenhum controle de erros, nenhuma confirmação e nenhum mecanismo para solicitar retransmissões. Outra função do RTP é multiplexar diversos fluxos de dados de tempo real sobre um único fluxo de pacotes UDP. O fluxo UDP pode ser enviado a um único destino ou a vários destinos (multidifusão)[70]. Também é descrito em seu cabeçalho o tipo de codificador e a taxa de amostragem que está sendo usada na mídia transmitida. Há um campo nele pelo qual pode-se sincronizar o destino com a fonte.

O protocolo RTCP (*Realtime Transport Control Protocol*), definido através da RFC

1889 do IETF, trabalha em conjunto com o RTP fornecendo informações de controle e sincronismo para as aplicações e usuários participantes da conexão ou chamada, porém não transporta nenhum dado. Uma das funções mais relevantes deste protocolo é fornecer informações às fontes de fluxo sobre o estado atual da rede, tais como: atraso, *Jitter*, banda disponível, congestionamento e outros parâmetros. Esta informação pode ser usada no processo de codificação para aumentar a taxa e dar maior qualidade quando a rede está funcionando bem e reduzir a taxa quando existir problema na rede. O RTCP auxilia no sincronismo entre os fluxos de voz ou vídeo. Diferentes fluxos podem usar diferentes relógios, com granularidades diferentes e diferentes taxas de variação. O RTCP pode ser usado manter os fluxos em sincronismo. O RTCP provê também uma forma de nomear as diversas fontes. Esta informação pode ser mostrada na tela dos receptores indicando quem está se comunicando[70].

Os protocolos H.323 e SIP

Com certeza, umas das dificuldades para a telefonia é determinar uma arquitetura de sinalização. Esta denota a forma de localizar endereços e também o estabelecimento de chamadas na Telefonia. Entre os protocolos mais promissores com esta finalidade está o protocolo de inicialização de sessão, *Session Initiation Protocol*. Outro padrão ainda bastante utilizado é o H.323. Estes protocolos permitem realizar a localização de nomes e/ou números telefônicos e fazer toda a sinalização de chamada, de modo a mapear transparentemente os endereços IP e portas de comunicação de origem e destino das partes envolvidas na comunicação.

O protocolo SIP, definido na RFC 3261, possibilita estabelecer chamadas entre dois interlocutores por uma rede IP. Permite quem chama avise ao que é chamado que quer

iniciar uma chamada. Permite que os participantes concordem com a codificação da mídia e que também encerrem as chamadas. Provê mecanismos que permitem a quem chama identificar o endereço IP corrente de quem está sendo chamado. Suporta gerenciamento de chamadas tais como adicionar novos fluxos de mídia, mudar a codificação, convidar outros usuários a participarem da chamada, e ainda transferir e segurar chamadas[35].

O H.323 é uma especificação guarda chuva que inclui outras especificações. Uma especificação para o modo como os terminais negociam diferentes codificações comuns de audio/vídeo. Como o H.323 suporta uma variedade de padrões de codificadores de audio e vídeo, é necessário que um protocolo permita aos comunicantes chegar a um acordo quanto a uma codificação comun. Outra especificação para o modo como porções de audio e vídeo são encapsuladas e enviadas a uma rede. Em particular para esta finalidade o H.323 impõem o RTP. Outro protocolo define o modo como telefones por INTERNET se comunicam por meio do gateway com os telefones comuns da rede pública por comutação de circuitos[35].

O H.323 vem do ITU (telefonia) enquanto que o SIP vem do IETF e toma emprestado vários conceitos da INTERNET, como resolução de nomes, email e etc.

Vantagens e desvantagens do VOIP

Entre as vantagens podemos destacar:

(1) União dos mundos da telefonia e da Internet em um só, ou seja, compartilhamento da rede de dados com o tráfego de voz. A principal vantagem de Voz sobre IP (VoIP), sobre a Rede de Telefônica Pública Comutada é a facilidade de adição de novos serviços e funcionalidades, assim como a significativa diminuição dos custos de implantação por parte

das empresas, uma vez que a estrutura da Internet já existe.[37]

(2) Otimização da utilização da largura de banda e da infra-estrutura de dados: uma vez que as redes de transporte de dados baseadas em IP são cada vez mais frequentes, a utilização de um sistema de telefonia IP permite utilizar as ligações de dados como suporte do tráfego de voz. Desta forma é possível os dois tipos de comunicação (dados e voz) numa única infra-estrutura e ter uma maior rentabilidade da mesma infra-estrutura;[36].

(3) Contribui para a integração de mais uma tecnologia(som, em especial a voz humana) com a Internet, a qual possibilita a realização de conferências eletrônicas envolvendo componentes audiovisuais e textuais.

(4) Redução dos custos das chamadas telefônicas. Obviamente, isso não ocorre em todos os casos. Mas, por exemplo, em uma empresa cujas filiais estão espalhadas em outros estados e até em outros países a diminuição pode ser significativa. Também entre universidades como é o caso das que participam da rede da RNP(Rede Nacional de Ensino e Pesquisa).

(5) A telefonia IP torna transparente a localização física de seu usuário. Basta este indicar a sua localização atual em registros mantidos em servidores VoIP para que uma ligação seja encaminhada para ele.[38]

Entre as desvantagens do VOIP podemos destacar:

(1) Requer cuidadoso planejamento da rede para garantia da qualidade de Voz.[37]

(2) Dispositivos VoIP ainda são caros. Por exemplo, telefones IP, gateways etc.

(3) Em geral, grande variação de desempenho. Redes de pacotes, diferentemente das redes de telefonia convencional, não garantem uma largura de banda fixa[37]. O que nos dá uma qualidade de som imprevisível.

(4) Requer uma base de conhecimento em redes, telefonia e voz.

Qualidade de Serviço (QoS)

Na Internet e nas intranets atuais, a largura de banda é um assunto importante. Mais e mais pessoas estão usando redes de computadores por motivos comerciais e particulares. O montante de dados que precisa ser transmitido através das redes vem crescendo exponencialmente. Atualmente existe uma tendência de convergência de aplicações em um único meio físico, ou seja, voz, vídeo, dados, imagens, músicas, e tudo que possa ser transformado em bits utilizando o mesmo meio físico. Novos aplicativos de voz e vídeo precisam cada vez de mais largura de banda que os aplicativos antes utilizados, pois trabalhavam somente com texto e imagens estáticas.

Enquanto que aplicativos tradicionais, como WWW, FTP ou email, não toleram perda de pacotes, mas são menos sensíveis aos retardos variáveis, a maioria dos aplicativos em tempo real apresenta exatamente o comportamento oposto, pois podem compensar uma quantidade razoável de perda de pacotes mas são, normalmente, muito críticos com relação aos retardos variáveis[67]. Isso significa que sem algum tipo de controle de largura de banda, a qualidade desses fluxos de dados em tempo real depende da largura de banda disponível no momento. Larguras de banda baixas, ou mesmo larguras de banda melhores, mas instáveis, causam má qualidade em transmissões de tempo real, com eventuais interrupções ou paradas definitivas da transmissão. Por isso, são necessários conceitos novos para garantir uma Qualidade de Serviço(QoS) específica para aplicativos em tempo real. Uma QoS pode ser descrita como um conjunto de parâmetros que descrevem a qualidade (por exemplo, latência, largura de banda, perda de pacotes, prioridades, variação no atraso

e etc.) de um fluxo de dados específico.

A pilha do protocolo IP básica propicia somente uma QoS que é chamada de melhor esforço. Os pacotes são transmitidos de um ponto a outro sem qualquer garantia de uma largura de banda especial ou retardo mínimo. No modelo de tráfego de melhor esforço, as requisições são processadas conforme a estratégia do primeiro a chegar, primeiro a ser atendido. Isso significa que todas as requisições têm a mesma prioridade e são processadas uma após da outra. Não há possibilidade de fazer reserva de largura de banda para conexões específicas ou aumentar a prioridade de uma requisição especial.

A crescente demanda pelos serviços IP *Telephony*, provocou uma corrida frenética dos fabricantes de equipamentos de redes para desenvolver protocolos que garantissem qualidade de serviços fim-a-fim. Para se obter o efeito da qualidade fim a fim, foram criadas técnicas que atuam basicamente em quatro aspectos:

Marcação dos pacotes: O protocolo IP, através de um campo específico, permite que cada pacote possua uma marcação própria que pode identificar um tipo de aplicação ou dado específico. No caso das aplicações de voz todos os pacotes de voz são marcados com um identificador único que permite a distinção entre os demais pacotes de dados. Normalmente um dispositivo VoIP gera pacotes com uma marcação padronizada que permite sua identificação na rede, ou então, o primeiro dispositivo de rede o qual o pacote passa realiza a marcação mediante programação prévia.

Criação de classes de serviços: De acordo com as marcações realizadas nos pacotes, as mesmas são agrupadas e identificadas e dão origem às filas de pacotes, sendo que

cada uma poderá ter tratamento diferenciado dependendo das características da aplicação que ela representa.

Mecanismos de filas: Existem várias formas de tratar a concorrência entre pacotes de filas diferentes, ou mesmo de uma única fila, ou seja, como é organizado o sequenciamento de envio de pacotes por uma interface. Nesta fase define-se a prioridade de uma classe ou fila em relação a outra. Várias técnicas foram desenvolvidas buscando atender as diversas aplicações de voz e vídeo que atualmente estão se consolidando nas redes, por exemplo: *First In First Out, Class Based Queue, Low Latency Queue* e etc[37].

Reserva de banda: para cada classe de serviço que tem um mecanismo de fila adequado para a característica de suas aplicações, pode ser reservada uma banda específica que garanta o pleno funcionamento daquela aplicação. Por exemplo pode-se separar 128 Kbits/s de um link de 512 Kbits/s somente para aplicações de voz.

A aplicação das técnicas acima citadas permitem garantir valores míninos em algums parâmetros essenciais para uma eficiente comunicação.

Capítulo 5

Parâmetros de medição do sinal de voz

Observando as característas da voz humana, podemos encontrar diversos parâmetros que influenciam a medição da voz e que podem aferir sobre sua qualidade. Dentre os parâmetros apresentados, alguns se referem a medidas aplicáveis a grandezas ligadas à parte física e orgânica dos aparelhos fonador e auditivo, como *pitch, loudness*, volume e etc e, no que diz respeito à percepção de qualidade de voz por um usuário de sistemas de telecomunicações são discutidos parâmetros tais como latência, *jitter*, perda de pacotes e etc.

5.1 Parâmetros físicos e orgânicos da voz

5.1.1 *Pitch*

A *pitch* é a frequência fundamental das cordas vocais e é determinada pela taxa de vibração das mesmas. Quanto maior for o número de vibrações por segundo, maior será a taxa de *pitch*. A taxa de vibração, por sua vez, é determinada pelo comprimento e espessura das cordas vocais e pelo tensionamento ou relaxamento destas cordas. Normalmente, nas vozes

femininas as frequências de pitch são mais elevadas que nas vozes masculinas.

5.1.2 Potência acústica

É uma medida física da quantidade de energia produzida e irradiada no ar por segundo. É o fluxo de energia acústica por unidade de tempo. Ao contrário do que acontece com a intensidade e a pressão sonora, a potência não depende do ambiente nem da distância da fonte, pois refere-se à energia emitida pela fonte. É medida em WATTS.

O Nível de Potencia Sonora (NPS) ou *Sound Power Level*(SPL) é expresso em decibéis tomando-se como referência o valor $W_0 = 10^{-12}W$ (1 picowatt).

5.1.3 Volume x Intensidade sonora

O volume é uma propriedade ou característica do som. O termo "volume"se aplica popularmente quando desejamos diminuir ou aumentar a quantidade de som existente em um ambiente. Não devemos confundi-lo com a intensidade sonora, pois esta é uma prática musical.

O que difere a intensidade do volume, é que o ato de um indivíduo dirigir-se a um aparelho de som e "aumentar o volume"não significa que ele esta tocando "forte"ou "fraco"naquele trecho musical que foi alterado. A intensidade sonora "I"é a medida de potência por unidade de área, medida em watts/m^2.

$$I = \frac{P}{4\pi.R^2} \tag{5.1}$$

Onde P= Potência e R= raio da superfície esférica.

O Nível de intensidade sonora corresponde ao logaritmo da razão entre a Intensidade sonora e um nível de referência.

$$NIS = 10log\frac{I}{I_0}dB \tag{5.2}$$

O nível de referência padrão I_0 é 10^{-12} watt/m^2.

A definição mais comum de volume, é a de quantidade de som, medida em relação a um nível relativo. Alguns controles de volume são calibrados em decibéis (dB). Normalmente, a cada marcação de um controle de volume a potência acústica dobra.

5.1.4 Nível de pressão sonora

A variação média (RMS - root mean square) de pressão em relação à pressão atmosférica; medida em Pascais (Pa) ou Newtons por metro quadrado (N/m2). Corresponde ao logaritmo da razão entre a Pressão sonora e um nível de referência.

O Nível de Pressão Sonora - NPS (Sound Pressure Level - SPL) em um determinado ponto é expresso em decibéis e tem como nível de referência padrão $P_0 = 0,00002Pa$.

$$NPS = 20log\frac{P}{P_0}dB \tag{5.3}$$

5.1.5 *Loudness*

O *Loudness* tem relação com a percepção do volume da voz e com o tipo de ambiente em que a voz está sendo emitida. Pode ser classificado em forte, fraco e adequado. Na voz cantada, deve-se lembrar do ambiente e da utilização da aparelhagem de amplificação sonora. O cantor pode não possuir uma boa aparelhagem de som ou retornos eficientes,

fazendo com que produza uma voz cantada com muito volume, e solicitando ao seu corpo que produza um apoio muito potente para que não ocorra sobrecarga das cordas vocais. O cantor popular normalmente não precisa utilizar um volume maior para cantar pois possui um bom sistema de amplificação. O cantor de coral emprega muitas vezes excesso de volume para poder se escutar dentro do coro. O cantor lírico procura explorar todas as suas caixas de ressonância, todo o seu potencial respiratório para atingir o quarto formante e assim ser ouvido junto com a orquestra que o acompanha. A voz transforma-se, neste momento, em mais um instrumento.

Por que quando aumentamos os graves e agudos no equalizador temos uma sensação de melhor definição do som de uma música? Por que quando apertamos a tecla *loudness* de um aparelho de áudio, o som fica mais "presente"? A intensidade sonora, também chamada popularmente de "volume do som", é percebida pelo ouvido humano de acordo com a frequência do som. Sons cujas frequências são muito graves ou muito agudas são percebidas pelo ouvido como tendo menos intensidade do que as frequências médias, próximas da voz humana. Esse fenômeno foi estudado por Fletcher e Munson que fizeram experimentos com indivíduos submetidos a escutarem sons senoidais. A intensidade desses sons era mantida constante enquanto sua frequência variava desde muito grave até muito agudo. Os indivíduos eram então interrogados a cada passo dessa variação sobre a sensação de sua intensidade sonora. Os resultados médios desse experimento foram expressos no gráfico da Fig 5.1.

A unidade do *loudness* é o *phon*. As curvas do gráfico mostram que a intensidade sonora necessária para se ter o mesmo *loudness* varia de acordo com a frequência do som. Observe, por exemplo, que um som na região entre 4KHz a 6KHz começa a ser percebido

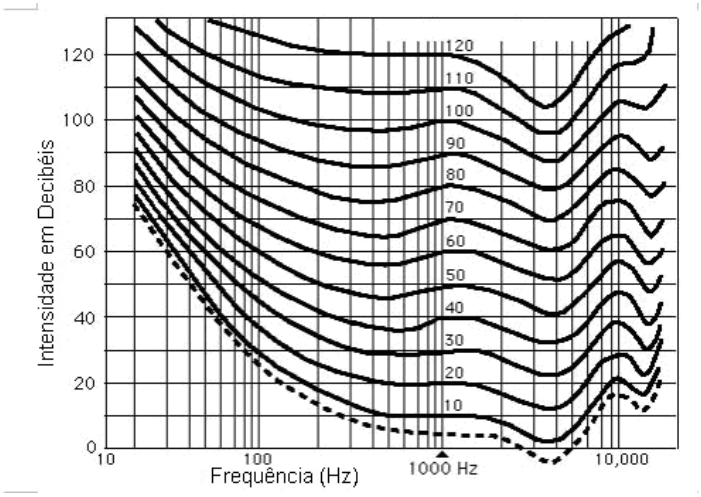


Figura 5.1: Gráfico de Loudness

Fonte: http://www.nics.unicamp.br/nicsnews/curiosidades.php

com aproximadamente 0dB enquanto que o mesmo som na frequência de 20Hz começa a ser percebido com 70dB, ou seja, com intensidade sonora muito superior. Esta primeira curva pontilhada, de 0 *phon*, representa o limite da percepção sonora, onde os sons começam a ser percebidos pelo ouvido. A última curva de loudness, de 120 phon, é a curva do limiar da dor, onde o som está tão intenso (alto) que provoca a sensação de dor nos ouvidos.

A tecla *loudness* encontrada em muitos aparelhos de áudio procura compensar a perda natural de percepção do ouvido nas frequências baixas e altas. Da mesma forma que um equalizador regulado para aumentar essas frequências, o *loudness* aumenta a intensidade das frequências extremas corrigindo a curva de "*equal loudness*" (vista acima). Assim tem-se maior definição dos graves e agudos de uma música, o que torna a sua audição bem mais nítida e confortável.

Sone (*Specific Loudness Sensation*)

O Sone é uma escala subjetiva de audibilidade, representa numericamente o *loudness* de acordo com a sensação auditiva humana. As grandezas acústicas costumam ser indicadas em decibéis. Por exemplo, para dobrar o volume de um som de 1 kHz, aumenta-se a pressão acústica em 10 dB-SPL. No entanto, ao invés de se falar "aumentar 10 dB", alguns podem preferir a frase "dobrar o volume". O Sone é uma maneira de expressar uma grandeza em número de vezes o valor de referência, estabelecida em 40 dB-SPL ou 40 *phons*. Como a cada 10 dB o volume dobra (a 1 kHz), podemos construir uma escala em Sones, onde o volume de som a 40 dB seja igual a 1 Sone, com 50 db teremos 2 Sones e assim sucessivamente[40]. A equação 5.4 permite construir esta escala, onde SPL(dB)representa a pressão acústica medida a ser transformada.

$$Sone = 2^{\frac{SPL(dB)-40(dB)}{10}} \tag{5.4}$$

5.1.6 *Shimmer*

No processo de geração da voz no aparelho fonado humano, o termo *shimmer* significa perturbação ou variabilidade da amplitude ciclo a ciclo.

5.2 Parâmetros importantes em redes de telecomunicações

Com o advento da tecnologia VoIP alguns parâmetros típicos de redes de computadores passaram a ter vital importância na avaliação da qualidade de uma chamada telefônica.

Novos conceitos foram criados, pois a concorrência de aplicações de voz e dados em um mesmo meio exigiu uma regularidade maior no tempo de entrega dos pacotes e uma coordenação quando se transmite mídias diferentes. Os fatores que interferem na qualidade da voz que transita por uma rede de pacotes são muito mais numerosos e complexos que os existentes na telefonia convencional.

5.2.1 Latência

Em redes de telecomunicações, mais especificamente em redes de pacotes, latência é o tempo que leva para a voz sair da boca do originador e chegar até os ouvidos dos destinatário, ou seja, é o tempo que o pacote leva da origem ao destino. Caso esse atraso seja muito grande, prejudica uma conversação através da rede, tornando difícil o diálogo e a interatividade necessária para certas aplicações. Segundo experimentos, um atraso confortável para o ser humano fica na ordem de 100ms. As redes de telefonia convencional possuem uma latência de 30 ms ou menos.

Para VoIP, o objetivo é manter a latência, de uma única via, em 100ms ou menos, podendo chegar a 150ms ainda com qualidade razoável. Latências muito longas farão com que o originador tenha que fazer pausas durante sua fala, caso contrário, ele não saberá quando o outro lado da chamada tiver terminado de falar, ou eles podem sobrepor a fala um do outro. Em sistemas de voz sobre IP atrasos que ultrapassam certos limites são considerados inaceitáveis, pois a qualidade da comunicação cai excessivamente. Para a recomendação ITU-T G.114 o limite extremo deste atraso é de 400ms. Na Fig. 5.2 observa-se um gráfico, onde foi representada a percepção de qualidade da voz, por parte de usuários, em a relação latência, de acordo com o modelo E[57]; método objetivo de medição

da qualidade baseado em parâmetros, que será apresentado no capítulo 6 deste trabalho. Verifica-se que a insatisfação começa a crescer a partir de valores de latência na ordem de 400 ms.

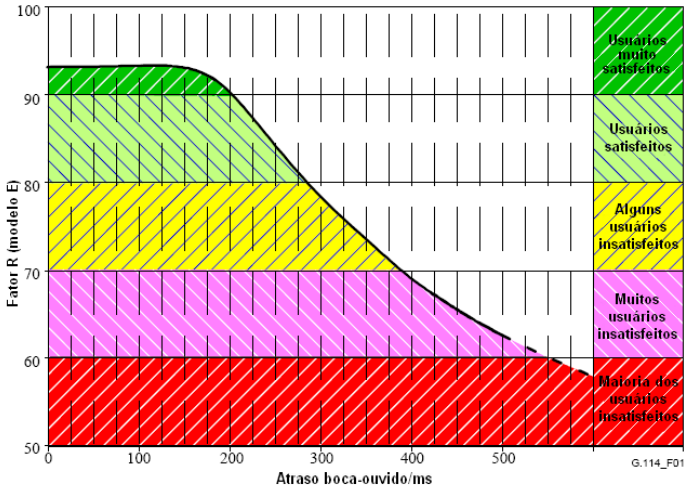


Figura 5.2: Determinação do efeito absoluto do atraso na qualidade da voz pelo Modelo E

Fonte: Recomendação ITU-T G.114

Suponha duas pessoas conversando através da Internet. À medida que o atraso aumenta, as conversas tendem a se entrelaçar, ou seja, uma pessoa não sabe se o outro a ouviu e continua falando. Após alguns milissegundos vem a resposta do interlocutor sobre a primeira pergunta efetuada, misturando as vozes. Num atraso muito grande, as pessoas devem começar a conversar utilizando códigos, tipo "câmbio", quando terminam de falar e passam a palavra ao outro.

Existe também o conceito de atraso de ida-e-volta, que corresponde ao tempo transcorrido entre o momento em que o usuário fonte emite uma conversação, que o usuário de destino recebe essa conversação e então emite uma resposta e o momento em que o usuário fonte recebe essa resposta.

Os principais responsáveis pela latência são o atraso de transmissão, de codificação e de empacotamento, que podem ser definidos da seguinte forma:

Atraso de transmissão

Esse tempo envolve uma série de fatores, como o atraso no meio físico (por exemplo, fibra ótica, cabo de par metálico, rede sem fio ou comunicação via satélite), processamento em cada dispositivo intermediário (roteadores, *proxies , firewalls* e etc), fila de espera em cada equipamento intermediário, e assim por diante. No caso de dois usuários utilizando o VoIP pela internet, este tempo inicia após a placa de rede ter transmitido o pacote, e vai até ele chegar na placa de rede do computador de destino.

No que se refere ao percurso dos pacotes IP pela rede de dados. Existem dois tipos de fontes para esse atraso:

- Atraso de roteamento: Devido à politica utilizada nos roteadores (*best effort*) o pacote nem sempre ganha a preferência na fila do roteador, causando o atraso. A politica usada para solucionar esse problema é a inserção no cliente de um *buffer*, o qual tenta suavizar as variações do atraso.

- Atraso em *firewalls e proxies*: por vezes o pacote encontra pela frente *firewalls e proxies* os quais vão introduzir mais atraso devido as suas filas e também ao seu processamento, visto que eles precisam verificar o conteúdo interno do pacote.

Atraso de codificação e decodificação

A voz humana, nativamente analógica, precisa ser convertida em digital para ser usada na transmissão. E devido a esta ação, realizada no processador, são introduzidos mais atrasos. São, basicamente, os processos de amostragem, quantização e compressão do sinal. Sinais

como voz normalmente são codificados em um padrão, tipo PCM (G.711 a 64Kbps). Essa codificação gasta um tempo de processamento na máquina. Alguns protocolos gastam menos, como o G.711, que ocupa menos de 1ms de codificação, porém, requer 64Kbps de banda. Alguns protocolos de voz, como o G.729, requerem 25ms de codificação, mas ocupam apenas 8Kbps de banda;

Atraso de empacotamento e desempacotamento

Duas causas básicas influenciam no atraso de empacotamento e desempacotamento: o atraso no empacotamento IP e o atraso de reprodução ou de supressão de *jitter*.

- Atraso no empacotamento IP: os quadros de voz são montados em uma série de protocolos: RTP - UDP - IP. Processo esse realizado pela CPU ou pelo gateway. Após codificado, o dado deve ser empacotado na pilha OSI a fim de ser transmitido na rede, e isso gera um atraso. Por exemplo, numa transmissão de voz a 64Kbps, tem-se que, para preencher um pacote de dados contendo apenas 100 bytes, vai levar 12,5ms. Mais 12,5ms serão necessários no destino a fim de desempacotar os dados.

- Atraso de reprodução ou de supressão de *jitter*: Devido a característica das redes TCP/IP os pacotes chegam ao seu destino com atrasos variados. Esse atraso variável é denominado de *jitter*. No destino, o pacote não pode ser reproduzido imediatamente visto o risco de ter um corte no som por motivo de um atraso maior. Por isso é mantido um buffer para a organização desses pacotes. O armazenamento no lado do cliente compensa a variação do atraso provocada pela rede.

5.2.2 Perdas de pacotes

Como foi dito anteriormente, as redes TCP/IP possuem a característica de *best effort*. Elas fazem o maior esforço para entregar o pacote mas não garantem nada. As perdas de pacote podem acontecer quando a rede está congestionada e o pacote é descartado nos roteadores ou pode ocorrer devido a erros ocorridos na camada de enlace (PPP - *Point-to-Point Protocol, Ethernet, frame relay, ATM, ...*).

É preciso lembrar que em certas aplicações,como em tempo real, é considerado "perda de pacote"atrasos superiores a certos limites. As aplicações que exigem um fluxo continuo, ou em tempo real, sofrerão com isso. Na verdade é estabelecido uma tolerância em que a perda de pacotes é aceitável. Ou seja, podemos dividir as perdas de pacote e os seus mecanismos de correção em dois casos:

(1) Quando ela é aceitável - os especialistas dizem que para toda comunicação essa perda deve ser no máximo de 5% [56]. Pode-se usar o método de repetir o último pacote nessa situação. Caso a perda seja consecutiva, a aplicação ou o gateway deve repetir apenas o último pacote recebido somente uma vez.

(2) Quando está acima do aceitável - nesse casos pode-se usar técnicas mais avançadas como o FEC(*Foward Error Correction*) que pode ser implementada em gateways VoIP. Essa técnica tem dois níveis:

a) Intra-pacote, no qual são adicionados bits extras permitindo a aplicação descobrir quais bits do pacote foram perdidos podendo ser corrigido o erro.

b) Extra-pacote, no qual são adicionadas informações extra a cada pacote permitindo que o gateway extrapole a informação do último pacote correto e reconstrua a informação que está corrompida.

Para perdas de 10 a 20%, esse mecanismo consegue absorver satisfatoriamente as perdas de pacote. No entanto, ele produz um consumo de banda extra sendo uma questão de escolha para a sua implementação.

No caso específico da tecnologia voz sobre IP, a perda de pacotes com a voz digitalizada implica numa perda de qualidade eventualmente não aceitável para a aplicação. Do ponto de vista da qualidade de serviço da rede (QoS) a preocupação é normalmente no sentido de especificar e garantir limites razoáveis (Taxas de Perdas) que permitam uma operação adequada da aplicação.

Nas aplicações de dados convencionais, o protocolo de transporte TCP, automaticamente retransmite os pacotes perdidos. Devido a sua característica de tempo real, as aplicações de VoIP utilizam os protocolos UDP e RTP, e o UDP não efetua a retransmissão dos pacotes perdidos (e também não faz sentido a retransmissão, pois somente atrapalharia a conversação).

A Tabela 5.1 relaciona a perda de pacotes com a qualidade da voz usando o codec G.711 e valores do documento TIPHON4[51](*Telecommunications and Internet Protocol Harmonization Over Networks*), grupo do ETSI (*European Telecommunications Standards Institute*). O termo MOS significa *Mean Option Score* e representa um valor númerico de 1 a 5 que vem a ser a média do valor atribuído por diferentes ouvintes para a qualidade da voz percebida.

Porcentagem de Perda	Qualidade de Voz	MOS
3%	Boa	4.2
15%	Média	3.8
25%	Pobre	3.0

Tabela 5.1: Relação da Perda de pacote e MOS

5.2.3 Perdas devido à codificação

Quando ocorre digitalização do sinal de voz é inevitável que haja perdas no sinal resultante deste processo. A fase de quantização do sinal não deixa dúvida desse fato, pois seriam necessários muitos bits para representar o sinal analógico, o que não é o caso. Ou seja, é inevitável a distorção do sinal. No momento da reprodução tem-se, então, o ruído de quantização.

Um outro fato importante, ocorre no uso dos codificadores os quais fazem a compressão do sinal que será transmitido. Ganhando-se, assim, no tamanho da banda necessária para transmissão, mas, perdem-se algumas características da voz como o timbre.

Ou seja, o uso dos codificadores apesar da diminuição da largura de banda, causam, além do atraso uma perda da qualidade de voz. Essa perda é percebida, por parte dos participantes da comunicação VoIP, pelo som "metálico".

5.2.4 *Jitter*

Utilizar somente a latência não é suficiente para definir a qualidade de transmissão, pois as redes não conseguem garantir uma entrega constante de pacotes ao destino, esta variação ocasiona o *jitter*, que nada mais é do que uma flutuação na latência, ou variação estatística do retardo. Conforme mostra a Fig 5.3, a latência dos pacotes não é fixa durante a

transmissão de dados em uma rede de computadores. Observa-se neste caso um exemplo que retrata bem a ocorrência de *Jitter*, tratam-se de pacotes de uma mesma origem para um único destino que apresentam um pico no tempo t1, vê-se que apesar dos pacotes terem a mesma origem e destino possuem atrasos diferenciados. Uma variação de atraso elevada produz uma recepção não regular dos pacotes e baixa qualidade na voz percebida pelos usuários.

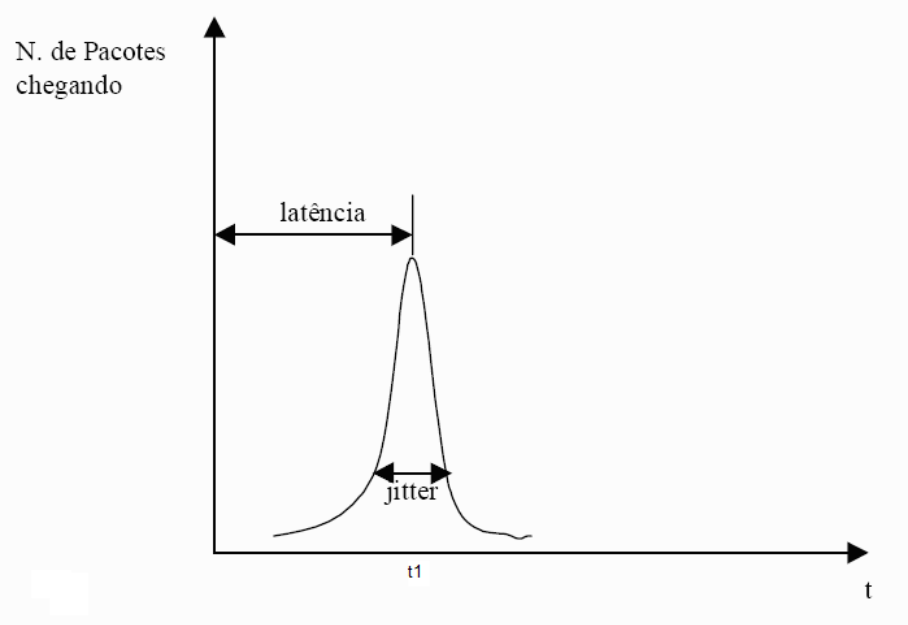


Figura 5.3: Comparação entre Latência e Jitter.

Fonte: http://www.ipg.pt/user/ sduarte/rc/Trabalhos2005/QoS/necessidadesQoS.htm

A consequência do *jitter* é que a aplicação no destino deve criar um *buffer* cujo tamanho vai depender do *jitter*, gerando mais atraso na conversação. Esse *buffer* vai servir como uma reserva para manter a taxa de entrega constante no interlocutor. Daí a importância de latência e *jitter* baixos em determinadas aplicações sensíveis a esses fatores, como, por exemplo, VoIP e videoconferência.

A variação do atraso *(jitter)* muda, em tempo real, em função do congestionamento do

tráfego na rede. Corresponde, essencialmente, à soma dos atrasos na fila e na transmissão, em cada comutador ou roteador intermediário na rede. A fila tem um impacto significativo no atraso se a voz estiver competindo com outras aplicações. O controle de *buffer* também incrementa o atraso total na rede. As redes IP podem ser projetadas para minimizar o atraso, acrescentando-se banda e reduzindo-se as aplicações que competem entre si, por exemplo. O *jitter* é um fator importante para a qualidade de serviço. No caso, o *jitter* é importante para as aplicações cuja operação adequada depende de alguma forma da garantia de que as informações (pacotes) devem ser processadas em períodos de tempo bem definidos. Este é o caso, por exemplo, de aplicações de voz e fax sobre IP (VoIP) e outras aplicações de tempo real, etc. *Jitter* severo em transmissões de VoIP causa distorção da voz .

A Tabela 5.2 relaciona os níveis de degradação da rede baseada no *jitter* usando o codec G.711 e valores do TIPHON4[51].

Categoria de degradação da rede		
Nivel de qualidade	*Jitter* médio (ms)	*Mean Option Score* (MOS)
Perfeita	0	4.5
Boa	75	4.0
Média	125	3.5
Pobre	225	3.0

Tabela 5.2: Níveis de degradação da rede baseada no jitter[51]

No estudo do processo de geração da voz no aparelho fonado humano, o termo *jitter* significa perturbação ou variabilidade da Frequência Fundamental ciclo a ciclo.

5.2.5 *Skew*

O *skew* é um parâmetro utilizado para medir a diferença entre os tempos de chegada de diferentes mídias que deveriam estar sincronizadas. Em muitas aplicações existe uma dependência entre duas mídias, como áudio e vídeo, ou vídeo e dados. Assim, numa transmissão de vídeo, o áudio deve estar sincronizado com o movimento dos lábios (ou levemente atrasado, visto que a luz viaja mais rápido que o som, e o ser humano percebe o som levemente atrasado em relação à visão). Outro exemplo é quando tem-se uma transmissão de áudio explicativo e uma seta percorrendo a figura associada.

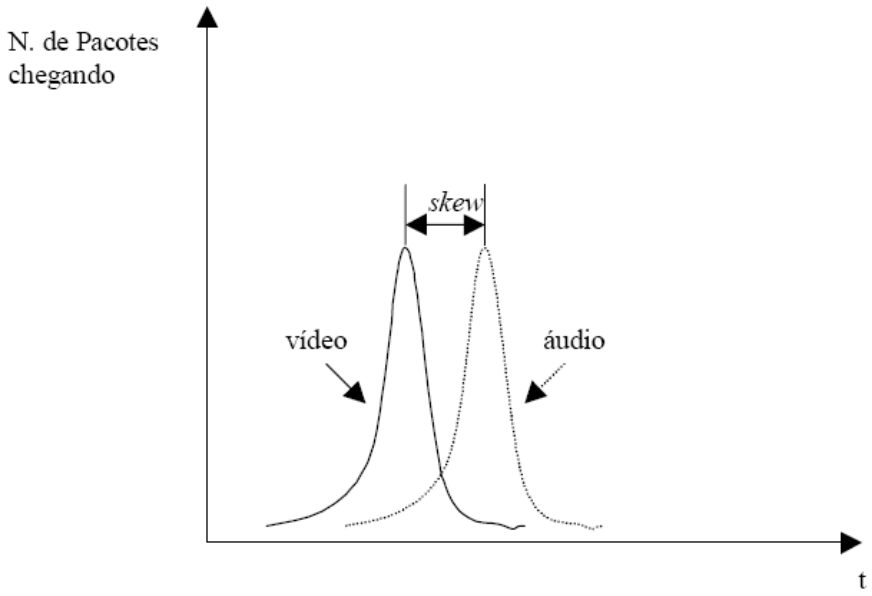


Figura 5.4: Definição de *Skew* entre vídeo e áudio.

Fonte: http://www.ipg.pt/user/ sduarte/rc/Trabalhos2005/QoS/necessidadesQoS.htm

5.2.6 Vazão (*Throughput*)

Vazão (*Throughput*) é uma medida absoluta da quantidade de dados que é transferida através de uma conexão dentro de um período de tempo especificado. Nas redes atuais, o tráfego oferecido para um serviço IP fim-a-fim não é checado por sua conformidade a um padrão de tráfego acordado. Além disso, essas redes podem limitar a taxa em que os pacotes são oferecidos por uma fonte simplesmente descartando esses pacotes. Normalmente não fazem qualquer compromisso para entregar qualquer tráfego oferecido. No entanto, é usual caracterizar o desempenho em termos de parâmetros relativos a vazão, que avaliam a capacidade de redes ou seções IP de transportar quantidades de pacotes IP. Um parâmetro caracterizando a vazão para um serviço IP deveria relacionar o total de pacotes IP transportados com sucesso por uma rede ou seção IP ao total de pacotes IP que foram entregues a esta rede ou seção. Alguns parâmetros relativos a fluxo ou vazão tentam caracterizar a capacidade de vazão de uma rede IP, ou seja, sua capacidade de sustentar uma determinada taxa de transferência de pacotes IP. É recomendado que quaisquer desses parâmetros devem cumprir os seguintes requisitos:

1) O padrão de tráfego oferecido para a rede ou seção IP deveria ser descrito, já que a habilidade da rede ou sessão IP para entregar de forma bem sucedida esses pacotes depende desse padrão de tráfego.

2) A taxa em que o tráfego é oferecido não deve exceder a capacidade (em bits por segundo) do enlace que conecta as sessões em teste.

Em qualquer declaração individual a respeito de desempenho da vazão, o tipo de pacote IP considerado deve ser declarado. Há dois tipos principais de parâmetros de vazão. Um deles mede a vazão em termos de taxa de pacotes IP transmitidos com sucesso. O outro é

baseado em octetos e mede a vazão em termos dos octetos que foram transmitidos nesses pacotes[53].

1 - IPPT (IP Packet Throughput)

Para um dado número de pacotes enviados de uma origem a um destino determinados, a vazão é o número total de pacotes IP transmitidos com sucesso ao destino, durante um intervalo de tempo especificado, dividido pela duração do intervalo de tempo.

2 - IPOT (Octet based IP packet throughput)

Para um dado número de pacotes enviados de uma origem a um destino determinados, a vazão é o número total de octetos em pacotes IP que foram transmitidos com sucesso ao destino, durante um intervalo de tempo especificado, dividido pela duração do intervalo de tempo. Existem vazões típicas[52] para cada aplicação, como apresentado na Tabela 5.3.

Vazão por aplicação	
Aplicações Transacionais	1 kbps a 50 kbps
Voz	10 kbps a 120 kbps
Aplicações Web	10 kbps a 500 kbps
Transferência de Arquivos (grandes)	10 kbps a 1 Mbps
Vídeo (stream)	100 kbps a 1Mbps
Vídeo MPEG	1 Mbps a 10 Mbps
Imagens Médicas	10 Mbps a 100 Mbps
Realidade Virtual	80 Mbps a 150 Mbps

Tabela 5.3: Vazão típica para diferentes tipos de aplicação

Capítulo 6

Técnicas de medição da qualidade

6.1 A qualidade vocal no sentido acústico

O timbre da voz é o fator que define a qualidade no nível das cordas vocais. O timbre costuma ser suficiente para a avaliação perceptual da voz falada. O timbre identifica o tipo de voz: soprano, mezzo soprano, contralto, tenor, barítono e baixo. O timbre de uma pessoa não é escolhido aleatoriamente, ele existe por razões anatômicas: o tamanho da laringe.

No canto, isso ocorre de forma diferente, pois na voz cantada avalia-se a qualidade em relação direta com o estilo de música adotado e com a forma pessoal de interpretação. Não é possível dizer que todo cantor de bossa-nova possui uma voz patológica soprosa, pois esta soprosidade faz parte de um estilo de cantar, de tornar a voz mais um instrumento dentro da música e não o principal deles.

O cantor de rock, normalmente, tem a voz rouca e áspera, mas pode-se utilizar outro modo de cantar este estilo de música. A qualidade da voz cantada abre discussão para as características vocais mais frequentes em cada estilo de música. É impossível se discutir a qualidade sem falar de estilo, ao contrário da voz falada, na qual a qualidade vocal tem

relação direta com a patologia.

Há várias maneiras de se nomear as qualidades e características vocais. Utilizamo-nos da voz rouca (moderada/suave), suave, fluida, áspera, soprosa, com quebra de sonoridade, tensa, pastosa, trêmula, estrangulada, infantil e diplofônica (bitonal).

Há outros nomes diferentes dos utilizados acima, mas estes combinados entre si quando necessários são suficientes para definir precisamente uma qualidade vocal no sentido acústico.

6.2 A qualidade da voz em sistemas de telecomunicações

Nos últimos anos, tem crescido o interesse pela utilização de redes IP para transporte de voz, em substituição às redes de telefonia tradicionais. Entre as principais motivações para tal estão redução de custos, simplificação da infra-estrutura decorrente da convergência de redes, além da possibilidade de se propiciar novos tipos de serviços. Para que a nova tecnologia de Voz sobre IP (VoIP) possa substituir a já consolidada tecnologia de comutação de circuitos, ela deverá ser capaz de transmitir a voz com boa qualidade e confiabilidade. Portanto, métricas de qualidade de conversação através de redes VoIP são necessárias para projetar, testar, manter e expandir tais sistemas.

Enquanto a Rede Pública de Telefonia Comutada (RPTC ou PSTN) aloca um circuito reservado durante toda a duração de uma ligação e emprega codificadores de baixa compressão, as redes IP baseiam-se na técnica de melhor esforço. Em consequência, o atraso ponto-a-ponto não pode ser seguramente garantido num determinado patamar de valores, e a variação do atraso (conhecida como *jitter*) ao longo de uma ligação e a perda de pacotes introduzem distorções não típicas da RPTC. Além disso, codificadores utiliza-

dos para reduzir a taxa de transmissão de bits também introduzem distorções e são mais sensíveis à perda de pacotes, por terem grande razão de compressão.

Essas mesmas características que comprometem a entrega de voz com qualidade via sistemas VoIP também dificultam a medida da qualidade de voz através deles. As métricas tradicionais não se aplicam precisamente às comunicações baseadas em VoIP, mas grupos de pesquisas na área de telecomunicações e a ITU (Internacional Telecommunication Union) têm se esforçado para estabelecer métricas confiáveis e que englobem os diversos elementos que as afetam.

Neste trabalho foi realizado um estudo comparativo entre os principais métodos já desenvolvidos para a medida de qualidade de voz, caracterizando como cada um deles obtém um parâmetro de medida e abordando os respectivos méritos e deficiências. Os métodos de medida da qualidade de voz baseados em uma escala de opinião dada por um grupo de ouvintes são classificados como **métodos subjetivos**. Aqueles cuja intenção é predizer o resultado dado por um grupo de ouvintes sem, no entanto, utilizá-los, são chamados **métodos objetivos** de medida de qualidade de voz.

Como método subjetivo será apresentada a técnica *Mean Opinion Score* (MOS). No que se refere à abordagem objetiva serão discutidas diversas técnicas, dentre elas, *Perceptual Speech Quality Measurements* (PSQM), *Perceptual Analysis Measurement System* (PAMS), *Perceptual Evaluation of Speech Quality* (PESQ), *E-model* e alguns métodos que são aprimoramentos dos anteriores.

6.3 Métodos subjetivos de medida da qualidade da voz

O método subjetivo existente para medição da qualidade da voz é chamado *Mean Opinion Score* (MOS) e é um padrão numérico usado para mensurar e reportar a qualidade da voz após a compressão e/ou transmissão. Os valores de MOS vão de uma faixa máxima de 5 pontos, que é considerado o mesmo que estar falando pessoalmente próximo ao ouvido de uma pessoa, até o valor de 1, que é considerado como qualidade inaceitável para todos os usuários. O MOS pontua apenas qualidade da voz e do som.

Os padrões P.800[4] e P.830[5] do ITU-T (*International Telecomunication Union*) definem as técnicas de medição do MOS. Aproximadamente 30 pessoas ou mais são submetidas a 8 ou 10 segundos de fala, em condições controladas. É solicitado a elas que opinem sobre as chamadas, como sendo muito boas até terríveis, pontuando-as de 5 a 1, ou seja, várias pessoas são recrutadas para ouvir um exemplo de conversa e convidadas a avaliá-la de acordo com um procedimento de classificação definido.

A recomendação ITU-T P.800 define as seguintes classificações:

- Classificação por categoria absoluta (ACR), cujo resultado é a pontuação de opinião média (MOS);

- Classificação por categoria de degradação (DCR), cujo resultado é a pontuação de opinião média de degradação (DMOS);

- Classificação por categoria de comparação (CCR), cujo resultado é a pontuação de opinião média de comparação (CMOS).

No procedimento de ACR, os ouvintes-avaliadores escutam amostras de conversação

na saída de um sistema de comunicação avaliado, sem comparar com as amostras de referência [45]. A opinião de cada ouvinte sobre a qualidade absoluta da voz e sobre o esforço exercido para a compreensão da fala é expressa numa escala de pontuação que varia entre um e cinco.

Este método utiliza três escalas de opinião:

Qualidade de audição (*Listening-Quality*): nesta escala, um sistema de pontuação define a qualidade de pequenos grupos de sentenças descorrelacionadas, cada uma submetida ao processo sob teste. O MOS é calculado pelo processamento estatístico dos resultados individuais. A tabela 6.1 apresenta este esquema de avaliação na coluna "Qualidade".

Valor	Qualidade	Esforço
5	excelente	relaxamento completo nenhum esforço é necessário
4	boa	atenção necessária não é preciso muito esforço
3	regular	um certo esforço é necessário
2	pobre	muito esforço é necessário
1	péssima	ininteligível apesar de qualquer esforço empregado

Tabela 6.1: Tabela de qualidade e esforço.

Esforço de audição (*Listening-Effort*): avalia os níveis de degradação. Neste caso, preocupa-se mais com a inteligibilidade do sinal, do que com qualidade, fato aceitável em algumas aplicações, como nas comunicações militares. A tabela 6.1 apresenta esta avaliação na coluna "Esforço".

Preferência de sonoridade (*Loudness-Preference*): define o grau de sonoridade (volume) percebido pelos ouvintes. Sua escala de degradação é mostrada na tabela 6.2.

Quando o critério não é mencionado (qualidade da fala ou esforço requerido), a pon-

Valor	Preferência de sonoridade
5	Muito mais alto que o ideal
4	Mais alto que o ideal
3	Ideal
2	Mais baixo que o ideal
1	Muito mais baixo que o ideal

Tabela 6.2: Tabela de percepção do volume da voz.

tuação MOS se refere à qualidade da fala. Apesar de o procedimento parecer simples, ele deve seguir algumas regras, a fim de produzir resultados confiáveis e reproduzíveis [19]:
- o número total de ouvintes deve ser suficientemente grande para se obter uma margem de segurança estatística;

- pessoas que trabalham diretamente envolvidas com avaliação de desempenho de sistemas de transporte de voz não devem estar entre os ouvintes-avaliadores;

- os ouvintes devem ser corretamente instruídos a respeito da metodologia dos testes, não podendo ter conhecimento prévio das amostras que ouvirá durante os mesmos;

- as amostras de voz a serem reproduzidas nos testes devem ser diversificadas em sexo, idade e sotaque;

- as condições dos experimentos devem estar controladas (volume físico da sala de testes, isolamento de ruídos externos, condições do equipamento utilizado, entre outros).

No procedimento de degradação (***Degradation Category Rating* - DCR)**, o teste avalia a degradação do material processado em relação ao material original, o que o torna mais sensível à distinção de qualidade, em contraste com os testes tipo ACR. A escala para esta modalidade é apresentada na tabela 6.3.

O teste de Comparação *Comparison Category Rating* - CCR se distingue do teste

Valor	Degradação
5	A degradação é inaudível
4	A degradação é audível, porém não incomoda
3	A degradação incomoda um pouco
2	A degradação incomoda
1	A degradação incomoda muito

Tabela 6.3: Tabela de degradação - DMOS.

tipo DCR apenas pela ordem em que as amostras são apresentadas aos ouvintes. Neste método, a ordem das amostras é escolhida aleatoriamente. Portanto, neste tipo de teste, os ouvintes têm de responder a duas perguntas: qual dos sinais é melhor e quanto ele é melhor, segundo a escala da tabela 6.4.

Valor	Comparação
3	Muito melhor
2	Melhor
1	Ligeiramente melhor
0	Aproximadamente igual
-1	Ligeiramente pior
-2	Pior
-3	Muito pior

Tabela 6.4: Tabela de comparação - CMOS

A vantagem do método CCR em relação ao DCR está na possibilidade de se poder avaliar não apenas processamentos de voz em que a qualidade é degradada, como também os casos em que a qualidade é melhorada. A deficiência deste tipo de teste, assim como no DCR, é que apenas desempenhos relativos podem ser obtidos. O método mais utilizado,

para a maioria das aplicações, tem sido o *Absolute Category Rating* (ACR), usando a escala qualidade de audição. Este método está bem estabelecido e tem sido aplicado a conexões telefônicas digitais e analógicas de dispositivos de telecomunicações. Vários laboratórios em diferentes países realizaram testes subjetivos utilizando este método, nas mesmas condições e com sistemas de transmissão idênticos, conseguindo resultados com alto grau de consistência. A média aritmética dos pontos atribuídos é denominada de *Mean Opinion Score* (MOS). As escalas de degradação e de comparação entre elementos de um par, usando os métodos de classificação DCR e CCR, também têm sido largamente utilizadas, e sua média aritmética é denominada *"Comparative Mean Opinion Score"* (CMOS).

Percebe-se, assim, que testes subjetivos são caros, demorados e de grande complexidade na sua execução. Métodos subjetivos são impraticáveis para executar testes frequentes tais como os necessários para projeto, monitoração e mudanças em redes de comunicação. Apesar de suas desvantagens intrínsecas, o MOS é a medida de referência para os testes de avaliação objetivos, pois ele espelha diretamente a opinião dos usuários finais de um sistema de comunicação.

A tabela 6.5 lista o MOS relacionado com alguns codecs comumente utilizados. A avaliação foi realizada em um ambiente de telecomunicações onde não havia fatores intervenientes negativos para a qualidade da voz, ou seja, foram proporcionadas excelentes condições para todos os parâmetros (latência, *jitter*, banda, etc). O objetivo foi analisar a qualidade da codificação, ou melhor, em condições ideais, identificar o melhor desempenho de um determinado CODEC.

Codec (taxa)	*Mean Opinion Score* (MOS)
G.711 (64 kbit/s)	4.30
iLBC (15.2 kbit/s)	4.14
AMR (12.2 kbit/s)	4.14
G.729 (8 kbit/s)	3.92
G.723.1 (6.3 kbit/s)	3.90
GSM EFR (12.2 kbit/s)	3.80
G.726 ADPCM (32 kbit/s)	3.80
G.729a (8 kbit/s)	3.70
G.723.1 (5.3 kbit/s)	3.65
GSM FR (12.2 kbit/s)	3.50

Tabela 6.5: Tabela de qualidade dos codificadores de voz [58].

6.4 Métodos objetivos de medida de qualidade

Os métodos objetivos são métodos nos quais a predição da qualidade da voz é realizada através de meios matemáticos que podem ser facilmente computadorizados e automatizados. Atualmente, os métodos objetivos podem ser separados em três grupos distintos, são estes: métodos intrusivos baseados em sinal, métodos não-intrusivos baseados em sinal e métodos baseados em parâmetros.

Conforme mostra a figura 6.1, a diferença entre os métodos objetivos intrusivos e os não-intrusivos está na forma de se obter o sinal de voz a ser medido ao passar por um sistema de comunicações. No método intrusivo o sinal de entrada ou de referência, assim como o sinal de saída, necessitam ser comparados em um avaliador intrusivo. No método não-intrusivo apenas o sinal de saída é utilizado na medição da qualidade.

Os métodos objetivos, também chamados perceptuais de medida de qualidade de voz,

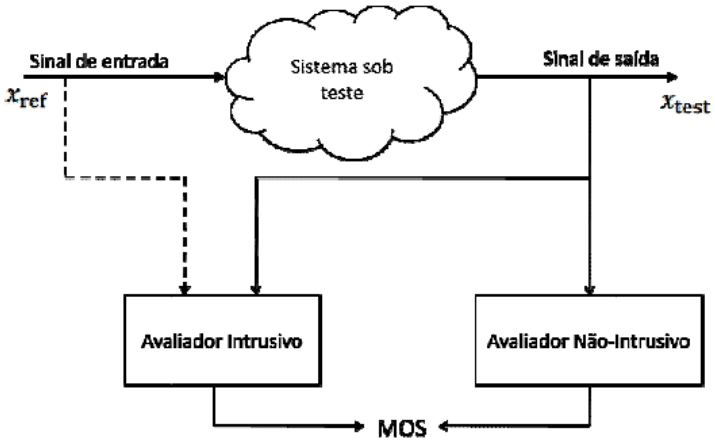


Figura 6.1: Comparação entre métodos intrusivo e não-intrusivo.

Fonte: Artigo do Congresso da Sociedade de Engenharia de Áudio: Avaliação subjetiva de qualidade de áudio: fala vs. Música.

utilizam o conhecimento do funcionamento do sistema auditivo humano para compor uma medida de distorção da voz através de um sistema de comunicação. As distorções mais significativas ao ouvido humano são computadas com maior peso do que aquelas que são quase inaudíveis [20].

6.4.1 Métodos intrusivos baseados em sinal

Os métodos intrusivos, baseados em sinal, são métodos que basicamente comparam o sinal de voz de entrada (sinal de referência) do sistema com seu sinal de saída (sinal degradado) trabalhando em cima das divergências entre os dois sinais (entrada e saída).

Uma amostra gravada de voz humana (sinal de referência) é submetida ao sistema de comunicação a ser inspecionado (VoIP, telefonia celular, por exemplo). Nesse sistema, pré-seleciona-se um codec (G.711, G.721, G.722, G.723.1, G.726, G.728, G.729a são os mais utilizados em VoIP). O sinal na saída do sistema (sinal degradado) é gravado e sincronizado

no domínio do tempo com o sinal de entrada antes de serem comparados, a fim de se evitar uma falsa descorrelação devida ao atraso.

A comparação entre o sinal degradado e o de referência baseia-se em fatores da percepção do ouvido humano tais como sensibilidade a certas frequências e amplitudes do sinal de voz, ao invés de uma simples comparação de densidades espectrais de potência ou outros parâmetros lineares da forma de onda dos sinais. Isso se deve ao fato de que codecs de baixa taxa de bits também se utilizam dessa modelagem da percepção do ouvido humano para poder transmitir somente as informações mais importantes através dos sistemas de comunicação.

Ainda nesta fase de inicialização, os sinais de entrada e saída são equalizados a fim de compensar o ganho total do sistema sob teste. Na segunda fase, os dois sinais são submetidos, separadamente, a modelagem perceptual, o qual varia de acordo com o método utilizado. Em linhas gerais, ambos os sinais são transformados do domínio do tempo para um domínio bidimensional de tempo-frequência, dividindo-se as amostras em vários *frames* de certa duração. A escala de frequência é convertida de Hertz para a escala Bark, que representa melhor como o ouvido humano percebe níveis de áudio diferenciadamente, de acordo com a faixa de frequência, onde: $f = 600senh(b/6)$[21], sendo f a frequência em hertz e b a frequência em bark.

A seguir, a escala de intensidade, baseada na densidade de potência dos sinais, é transformada para uma escala de sonoridade, que representa a sensibilidade sonora humana. Isso é feito porque o sistema auditivo humano percebe distorções de acordo com a sonoridade (volume) do sinal de áudio no qual o ruído ocorre [3]. O resultado desta fase de modelagem perceptual é conhecido também como uma representação matemática interna

dos sinais de entrada e saída.

A última fase, a modelagem cognitiva, é onde ocorre efetivamente a comparação entre os sinais de entrada e saída e uma pontuação é gerada. Novamente, o que diferencia os métodos perceptuais entre si é a maneira como cada um faz o cômputo do erro entre os sinais parametrizados e como cada um apresenta sua escala de pontuação. A regra geral é que há sempre uma relação entre pontuação apresentada e o MOS. Entre os algoritmos mais difundidos na literatura, encontram-se a medida perceptual de medida de qualidade (PSQM)[3], o sistema de medida de análise perceptual (PAMS)[22] e a avaliação perceptual da qualidade de conversação (PESQ)[7]. As características comuns a todos eles são levantadas a seguir.

Medida perceptual de qualidade da fala (PSQM)

O PSQM (*Perceptual Speech Quality Measure*) é um método perceptual de medida objetiva da qualidade de voz desenvolvido em 1994 por J. G. Beerends e J. A. Stemerdink, ambos da KPN Research, na Holanda, e encontra-se definido na recomendação P.861 da ITU-T[6]. Como todo método perceptual, seu processo de avaliação pode ser dividido em três passos: inicialização dos sinais, modelagem perceptual e modelagem cognitiva. A primeira observação sobre o PSQM é que ele, em si, não provê o alinhamento de tempo entre o sinal de referência e o degenerado durante a fase de inicialização. É necessário, pois, o uso de um algoritmo de alinhamento de tempo separado para alinhar os sinais antes de serem entregues à avaliação do PSQM. Daí, técnicas diferentes de alinhamento de tempo podem produzir pontuações diferentes de PSQM [3].

Durante a fase de modelagem cognitiva, o PSQM leva em conta a assimetria da

percepção auditiva humana. Pequenas distorções aditivas causadas por codecs são mais bem percebidas pelo sistema auditivo do que pequenas distorções atenuantes. Da mesma forma, o ruído que afeta os períodos de silêncio durante uma chamada telefônica é tratado com maior indulgência pelo ouvido humano do que em períodos de conversação. Assim, a pontuação do PSQM, que varia de zero (qualidade perfeita) a infinito, dá um peso maior para distorções aditivas do que para distorções atenuantes. Similarmente, a modelagem cognitiva aplica pesos diferentes para a influência do ruído no cômputo da pontuação do PSQM [3].

A pontuação final dada pelo PSQM indica o grau de degradação de qualidade subjetiva devido à codificação da fala, ao invés de medir diretamente a qualidade. Uma pontuação zero indica qualidade perfeita. Valores mais altos indicam níveis crescentes de degradação. O pior índice é de 6,5 [46].

O valor PSQM é uma estimativa da medida subjetiva de qualidade numa escala de degradação. Assim, ele não precisa ser transformado quando esta escala é suficiente, como, por exemplo, na comparação ou otimização de codecs. A transformação é necessária apenas quando se deseja expressar a avaliação em escalas subjetivas padronizadas, como o MOS ou CMOS. Os valores da escala MOS dependem da língua e laboratório em que os arquivos de voz foram gravados. Assim, não existe uma única função de transformação entre PSQM e MOS. Ao contrário, curvas devem ser definidas para cada língua, laboratório e, em geral, para o tipo de teste pretendido. E, portanto, dado que os valores MOS estimados via PSQM através de uma dada curva de mapeamento dependem deste contexto, toda vez que os valores MOS obtidos forem apresentados, a curva de transferência utilizada deve ser também apresentada[25]. A equação abaixo é um exemplo da transformação de um valor

medido pelo método PSQM para o MOS[46].

$$MOS = \frac{4}{1 + e^{0.66PSQM - 0.2}} + 1. \tag{6.1}$$

Entre as desvantagens do PSQM, podemos citar que ele não mede variação do atraso (*jitter*), nem a influência da perda de pacotes e de entrecortes de tempo, pois ele foi desenvolvido para medir especificamente a qualidade de codificadores, sem levar em conta a arquitetura da rede de comunicação. Seu mérito se encontra no fato de ter sido uns dos primeiros métodos perceptuais a ganhar grande aceitação. O PSQM foi posteriormente revisado, as modificações concentram-se principalmente na fase da modelagem cognitiva, pois essas versões oferecem um processamento assimétrico mais robusto, levando em consideração a influência da perda de pacotes, da variação brusca do atraso e do entrecorte de tempo.

Measuring Normalizing Blocks **(MNB)**

O MNB (*Measuring Normalizing Blocks*) foi desenvolvido pelo Instituto de Ciências de Telecomunicações do Departamento de Comércio Americano em 1997. Foi baseado no relatório[21] de Stephen D. Voran do *Institute for Telecommunications Science* e na Contribuição 24 (COM 12-24-E)[47] do Grupo de Estudo 12 do ITU-T. O método MNB foi publicado como uma proposta de anexo a P.861 (PSQM). Esse anexo foi aceito em 1998 como apêndice II da P.861. Descreve uma técnica alternativa ao PSQM para medir a distância perceptual gerada pelas transformações perceptíveis dos sinais de entrada e saída. Tal técnica é conhecida como Medida Normalizada em Bloco. Ele modela o julgamento humano da qualidade através da análise no tempo e na frequência. Essas duas análises são combinadas de forma que seja obtido um valor chamado *Auditory Distance* (AD), que

mede a distorção. O "AD"representa uma medida da distância perceptível entre os sinais de entrada (referência) e saída (teste) no sentido de predizer a qualidade subjetiva.

A medida é feita através do alinhamento entre o sinal original e o sinal a ser medido. São então eliminados os componentes DC, e posteriormente é estimada a potência média dos dois sinais, para que sejam normalizados. O passo seguinte é a transformação do sinal para o domínio da frequência utilizando uma FFT com janela *Hamming* de 128 amostras (16 ms) e overlap de 50%. Os quadros resultantes são comparados entre os dois sinais, de forma que são eliminados os quadros cujas diferenças estão abaixo de um determinado valor, e também os quadros com componentes de frequência com potência zero. Os quadros que não foram eliminados são transformados de acordo com uma escala de loudness, e comparados no domínio do tempo e da frequência. São então obtidos valores que representam as diferenças entre os sinais para diversos intervalos de frequências e é feita uma combinação desses valores, que representa o valor medido [45]. O método MNB, por possuir um algoritmo diferente do PSQM, é mais adequado para medir o impacto na clareza da voz levando em consideração outros fatores como, por exemplo, erros nos canais de comunicação ou codecs com taxas de transmissão inferiores a 4 kbps. Tais fatores não são tratados pelo PSQM original. O MNB pode ser considerado como uma técnica complementar ao algoritmo do PSQM [21].

Sistema de medida de análise perceptual (PAMS)

O PAMS (Perceptual Analysis Measurement System) é um método perceptual de avaliação objetiva da qualidade de voz desenvolvido em 1998 por Michael P. Hollier, do grupo Psytechnics da British Telecommunications[22]. Representou em sua época um avanço

na avaliação da qualidade da fala, pois foi o primeiro modelo a oferecer uma avaliação confiável de uma boa gama de redes de comunicações, incluindo VoIP . O PAMS oferece, em sua fase de inicialização, alinhamento dos sinais de entrada e saída do sistema sob teste, evitando o problema de pontuação dúbia, como no PSQM. Ele não somente alinha grandes seções da fala, como também seções mais curtas, removendo os efeitos de atraso e de variações lentas de atraso.

As variações bruscas no atraso são conservadas e medidas na fase de modelagem cognitiva. Durante a fase de modelagem perceptual, o PAMS utiliza um banco de filtros Sekey para escalonar os sinais audíveis para o domínio perceptual e separá-los em bandas de frequência na escala Bark, ao invés de utilizar a transformada rápida de Fourier, como no PSQM. O resultado é uma representação em tempo e frequência da sonoridade captada, conhecida como "superfície de sensação", a qual é calculada para o sinal de referência e para o degenerado.

Na fase de modelagem cognitiva, a diferença entre as duas superfícies de sensação é computada, determinando-se uma superfície de erro. Erros positivos correspondem a distorções devido ao codec ou ruído, enquanto erros negativos representam perda de energia do sinal. A superfície de erro é avaliada de várias maneiras, gerando diversos parâmetros de erro, os quais são comparados com um banco de dados de resultados de testes subjetivos. O PAMS retorna uma pontuação de qualidade em duas diferentes escalas de opinião, a de qualidade auditiva e a de esforço de compreensão, as mesmas utilizadas nos testes de ACR especificados na Recomendação ITU-T P.800 [22].

Há diversas diferenças importantes entre o PAMS e os modelos objetivos anteriores de avaliação da qualidade de voz.

- Atraso variável: o PAMS foi o primeiro modelo objetivo a levar em consideração o atraso variando no tempo que é uma característica do serviço voz sobre o IP e de outros serviços baseados em rede de pacotes. A versão 2, incorporando a habilidade de identificar o tipo mais comum de mudança de atraso - em períodos silenciosos - está disponível comercialmente desde dezembro 1998. A versão 3, de dezembro 1999, adicionou a capacidade para identificar variações do atraso durante a fala, mesmo que estas sejam muito menos comuns. Um perfil total do atraso e dos pontos de variação do atraso é retornado para o PAMS.

- Filtro nas interfaces analógicas: o PAMS foi projetado para uso em redes reais. As interfaces híbridas e as analógicas introduzem filtragem que faz com que os modelos objetivos anteriores como o PSQM(P.861) forneçam predições não confiáveis da qualidade. Para lidar com isso, o PAMS é capaz de identificar automaticamente uma ampla faixa de tipos de filtro. A função de transferência do sistema é retornada ao usuário para o diagnóstico.

- Robustez: o método usado para projetar o PAMS garante que sempre haverá uma relação de um para um entre a quantidade de distorção e a escala de qualidade. Outros métodos tais como a regressão linear ou redes neurais não podem garantir isto. Foi feito então um novo tratamento matemático para usar o conhecimento adquirido tal que se qualquer coisa piorar, a qualidade deve cair. Este processo torna o modelo mais exato em predizer a qualidade para as condições da rede que não fizeram parte de seus dados de treinamento, e dá uma confiança maior de que continuará a desempenhar com confiabilidade quando usado em campo.

Entre as desvantagens do PAMS, estão algumas relacionadas a todos os métodos perceptuais de avaliação da qualidade da fala, tais como dificuldade de acesso às duas pontas de um sistema de comunicação e o efeito do ruído de fundo. Além disso, PAMS assume que a qualidade da fala é relativamente constante durante a chamada.

Avaliação perceptual da qualidade de conversação (PESQ)

O PESQ (*Perceptual Evaluation of Speech Quality*) é um padrão de medida objetiva da qualidade de voz desenvolvido por J. G. Beerends e J. A. Stemerdink, ambos da KPN Research, e por A. W. Rix e M. Hollier, do grupo Psytechnics da British Telecommunications. Combina os méritos do PSQM99 e do PAMS e encontra-se definido na recomendação P.862 da ITU-T[7].

O PESQ foi desenvolvido com o objetivo de medir precisamente as distorções causadas por codecs de diversos tipos, transcodificação (conversão de um formato digital em outro), erros de transmissão, perda de pacotes, entrecorte de tempo, entre outros. Contudo, sua precisão ainda é desconhecida para parâmetros e situações tais como *delay*, ruído de fundo, falantes simultâneos, codecs com taxa de transmissão de bits abaixo de 4 kbps e fala artificial como sinal de referência [24].

Como apresenta a figura 6.2, o funcionamento do método PESQ ocorre de acordo com as seguintes fases: alinhamento no tempo, modelagem perceptual e modelagem cognitiva.

Na fase de inicialização dos sinais, o alinhamento de tempo é feito tal como descrito para o PAMS. Na fase de modelagem perceptual, o sinal de referência e o degenerado são separadamente transformados via FFT para o domínio de tempo-frequência, tal qual no PSQM. Em seguida, suas escalas de frequência e de sonoridade são convertidas respecti-

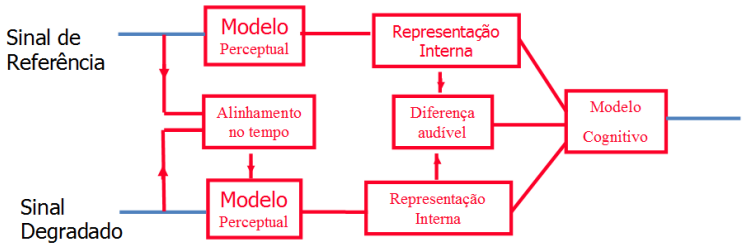


Figura 6.2: Esquema de funcionamento do método PESQ

Fonte: Recomendação ITU-T P.862.

vamente para a escala de Bark e de Sone, a fim de se representar melhor a sensibilidade auditiva humana.

Durante a fase de modelagem cognitiva, o PESQ calcula dois tipos de valores de distorção. Um deles se refere à variação brusca do *delay* detectado no processo de alinhamento de tempo. O outro valor se refere ao processamento assimétrico que é efetuado no PSQM. Esses valores são combinados ao final desta fase e produzem uma pontuação MOS que varia de 0,5 a 4,5.

Novamente, as desvantagens do PESQ residem nas que são comuns a todos os métodos perceptuais, tais como dificuldade de acesso às duas pontas de um sistema de comunicação e o efeito do ruído de fundo. Além disso, a modelagem de como o cérebro humano julga a qualidade de voz ainda não está totalmente definido. Um aspecto de interesse é o efeito da recentidade, pois em testes de MOS há uma tendência de ocorrer uma pontuação maior para amostras de conversação onde o ruído se apresenta no início delas e uma pontuação mais baixa quando o ruído já aparece ao final. Em experimentos, chegaram-se a diferenças de 0,64 e 0,68 [23]. Sua importância reside no fato de representar a consolidação do trabalho de grupos de pesquisa na área de modelagem perceptual do sistema auditivo humano. É o

padrão da ITU-T atualmente em vigor, tendo superado o ITU-T Recommendation P.861 (PSQM).

Em 2007 foi publicada pelo ITU-T a extensão P.862.2. Este trabalho permite a aplicação do método PESQ quando os ouvintes utilizam *headphones* de banda larga, a diferença básica da versão original, que agora é denominada P.862.1, é que o sinal de áudio varia na faixa de 50 a 7000 Hz, enquanto a anterior trabalha de 300 a 3100 Hz, banda padrão dos aparelhos telefônicos convencionais.

SQUAD(*Speech Quality Detector*)

O método SQUAD é um método de predição da qualidade da voz que utiliza o PESQ como base, adicionando mais um atributo chamado análise integrada das causas de qualidade. O método SQUAD, desenvolvido e patenteado pela empresa *SwissQual*, pode ser separado em duas partes diferentes, uma que utiliza o PESQ e uma parte da análise que consiste em um detector de degradação seguido pela análise dos valores das causas, onde o detector contém sensores para identificar diferentes tipos de deterioração que depois são avaliados separadamente [43]. São avaliados adicionalmente ao PESQ os fatores: ruído, eco, longas latências, controle de ganhos de sinal e falas inintelegíveis.

6.4.2 Métodos não intrusivos baseados em Sinal

Os métodos não-intrusivos baseados em sinal, diferente dos intrusivos, analisam os sinais de voz degradados sem compará-los com um sinal de referência, não afetando o tráfego da rede. Para isto, capturam parâmetros da rede tais como perdas de pacotes, jitter e atrasos. Estes métodos são os métodos mais utilizados atualmente, por serem razoavelmente precisos e

menos custosos do que os outros.

INMD(*In service, non intrusive measurement device*)

O método de predição da qualidade da voz não intrusivo (INMD), da recomendação P.561[41], foi originalmente desenvolvido para medir parâmetros de uma rede de comutação de circuitos.

Dois tipos de medições são abrangidas pelo INMD: expressão e caracterização do ruído e caracterização do eco. Nas primeiras implementações do INMD as medições incidiam apenas sobre caracterização do eco. Uma vez que a presença de eco era detectado, os dispositivos de medição apresentavam um relatório em tempo real do nível detectado de eco e do atraso gerador do eco. O INMD detecta eco pelo princípio da correlação cruzada, usa como parâmetros os sinais "INMD-escravo" e "INMD-MASTER", ambos apresentam uma sinalização interna que tem uma janela longa de 256ms (2048 amostras em 8000Hz taxa de amostragem).

A presença de eco é declarada quando todas as seguintes condições forem satisfeitas: O eco sinal lateral (lado "INMD-MESTRE") analisado na janela de captura apresenta sinal (E (n)), cujo nível de potência (Pe) é superior a -60 dBm.
A referência sinal lateral (lado "INMD-SLAVE") analisado na janela também captura alguns sinais r(n), cujo nível de potência (Pr) é maior do que Pe.

A correlação circular cruzada é feita entre r(n) e E(n):

$$cor(n) = \sum r(n)E(n) \tag{6.2}$$

Com a seguinte variação: n = 0, 1, 2,. . . , 2047.

Nas implementações atuais a correlação circular é obtida através de duas FFTs diretas e uma FFT inversa. A abordagem baseada em FFT é muito mais eficiente que o cálculo computacional. Atualmente, este método foi melhorado para também levar em consideração redes de comutação de pacotes e visa encontrar e analisar os efeitos que podem afetar o desempenho da transmissão da voz [43]. Os parâmetros medidos pelo INMD são fortemente relacionados às variáveis envolvidas no cálculo de qualidade do *E-Model*[54], podendo ser considerados como uma medida complementar do *E-Model*, que é um método não intrusivo baseado nos valores dos parâmetros da rede pela qual trafegam os pacotes de voz.

O método INMD normalmente realiza medições no meio da rede, criando um problema na medida, em tempo-real, na predição da qualidade da voz, visto que os dados recebidos pelo método diferem dos dados percebidos pelo usuário. Este método ainda pode ser ilusório, pois trata separadamente todos os parâmetros medidos, não levando em conta seus efeitos combinados [43].

CCI(*Call Clarity Index*)

Pensando na falta de garantia dos resultados medidos pelo INMD, foi elaborada uma extensão sua chamada CCI, recomendação P.562 do ITU-T[39], criada pelos mesmos desenvolvedores do método intrusivo PAMS. A Fig 6.3 ilustra o funcionamento do método de predição da qualidade da voz CCI, onde [44]:

O modelo de pressuposição adiciona informações sobre a rede que não estão presentes no método INMD, tais como tipos de rede, telefones e até indivíduos presentes em cada lado da chamada;

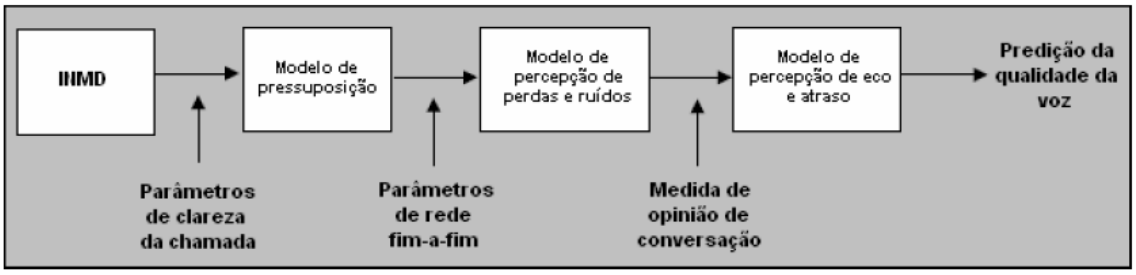


Figura 6.3: Esquema gráfico do funcionamento do método CCI

Fonte: http://www.psytechnics.com/site/sections/resources/whitepapers.php

O modelo de percepção de perdas e ruídos leva em consideração fatores perceptíveis à audição humana tais como seletividade de frequências, máscara de ruído e ruído local;

O modelo de percepção de eco e atraso calcula, através de parâmetros da rede, os efeitos de eco e atraso modificando o valor recebido do modelo de percepção de perdas e ruídos para um valor de saída, representado pela pontuação subjetiva do método MOS.

NINA(*Non-Intrusive Network Assessment*)

O método de predição da qualidade da voz NINA, desenvolvido pela empresa *SwissQual*, tem o seu algoritmo subdividido em 7 partes [43]:

1. O discurso, ou seja, a voz propriamente dita é separada do sinal de entrada, que contém voz, silêncio e ruído;

2. Um sinal da referência é aproximado separando o sinal degradado do sinal de entrada;

3. Cortes na fala são detectados e os quadros perdidos são recuperados. Após ter terminado estas etapas, o modelo contém um sinal degradado e um sinal de referência criado;

4. Os dois sinais são comparados em um método intrusivo;

5. Os impactos dos parâmetros ruído, eco e atividade de discurso são calculados;

6. Degradações criadas pelo codec são avaliadas;

7. Os resultados são finalmente processados para se obter uma pontuação satisfatória da predição da qualidade do discurso, correlacionada com o MOS subjetivo.

Como visto, o método objetivo de predição da qualidade da voz não intrusivo NINA basicamente cria um sinal de referência e, a partir deste sinal criado e do sinal degradado recebido, calcula o valor da predição da qualidade através de um método intrusivo, comparando os dois sinais, também leva em consideração parâmetros como ruído, eco e atividade de discurso.

Recomendação P.563

Esta recomendação do ITU-T é o resultado de um trabalho colaborativo entre a Psytechnics, SwissQual e Opticom, três fornecedores europeus de software e hardware para qualidade de voz, que anteriormente à P.563[48] tinham suas próprias versões proprietárias de software de análise da qualidade de voz (passiva, não intrusiva). *Non-intrusive speech Quality Assessment* (NiQA), desenvolvido pela Psytechnics; *Non-intrusive Network Assessment* (NiNA), desenvolvido pela SwissQual e *Perceptual Single Sided Speech Quality Measure* (P3SQM), desenvolvido pela Opticom. Lançada em maio de 2004, a P.563 faz uma análise extensiva da forma de onda e é, consequentemente, computacionalmente custosa. Como mostra a figura 6.4, nos métodos não-intrusivos baseados em sinal, a medição é tipicamente implantada nos *gateways e probes* das operadoras[49]. Os gateways são dispositivos já fabricados com as funcionalidades de monitoração da qualidade. Os probes são

sondas específicas para cópia dos pacotes de voz e envio para o software de gerenciamento de qualidade. A coleta de parâmetros é realizada através do protocolo RTCP (*Real-time Control Protocol*), com tipo de pacote XR (*Extended Reports*), definido na RFC 3611[72].

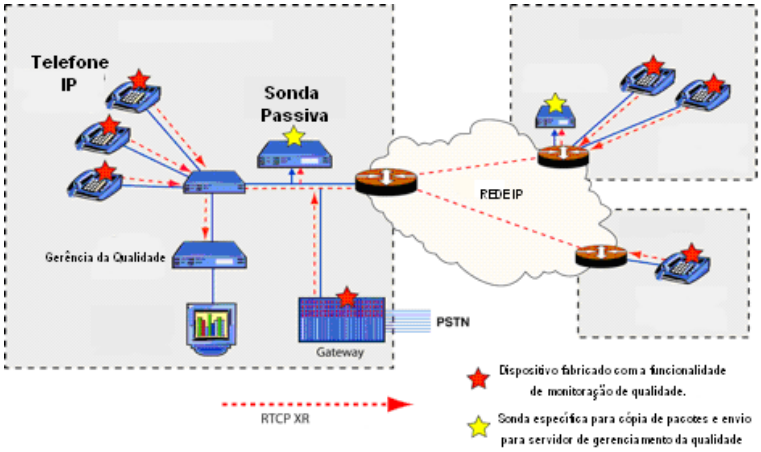


Figura 6.4: Sistema de medição não intrusivo baseado em sinal

Fonte: http://www.telchemy.com/Measuring VoIP Performance

A recomendação P.563 é considerada "monitoramento passivo". Não injeta qualquer dado na rede e usa software para analisar o áudio a partir das chamadas em curso e dá um valor MOS que prediz o que um grupo (humano) de peritos de teste daria. Neste método é criada uma pseudo-referência baseada no próprio sinal de voz que está sendo medido. A fala é extraída do sinal de entrada e então reconstituída para servir de referência na comparação. É diferente de outros algoritmos, pois não há arquivo de áudio de referência para comparar. A correlação entre as avaliações P.563 e aquelas dos usuários reais varia tipicamente de 0,85 a 0,9. Embora estes números mostrem que a P.563 não seja tão exata quanto o PESQ, é estatísticamente significativo o suficiente para ser uma ferramenta útil[50]. A correlação entre as avaliações subjetivas no método MOS e as medidas realizadas no método objetivo PESQ é menor que 0,05[71].

6.4.3 Métodos baseados em parâmetros

Diferente dos outros métodos objetivos, os métodos baseados em parâmetros não analisam os sinais de voz, baseando-se somente nos parâmetros capturados da rede, o mais conhecido método baseado em parâmetros é o chamado E-Model.

Modelo E

O Modelo E foi desenvolvido pelo ETSI (Instituto Europeu de Padronização em Telecomunicações) e encontra-se definido em detalhes no Relatório Técnico do ETSI 250 (ETSI ETR 250, 1996) e nas recomendações G.107[54] e G.108[55] da ITU-T.

Ao contrário dos métodos perceptuais, o modelo E não compara diretamente um sinal degradado com um de referência. Ele estabelece um método computacional de avaliação da qualidade subjetiva da fala através de um sistema de comunicação, onde cada elemento contribuinte para a degradação na qualidade da fala percebida é associado a um valor numérico denominado fator de perda. Os fatores de perda são computados pelo modelo E, fornecendo um fator de avaliação R, de valor entre 0 e 100, que pode ser relacionado a um valor MOS. Uma pontuação próxima de cem indica ótima qualidade de voz, ao passo que pontuações próximas de zero indicam qualidade péssima.

O fator R é obtido pela seguinte fórmula:

$$R = Ro - Is - Id - Ie + A \tag{6.3}$$

"Ro" representa os efeitos da relação sinal ruído (SNR); "Is" representa as perturbações simultâneas na transmissão da voz; "Id" representa as perdas associadas ao atraso boca-a-ouvido; "Ie" representa as perdas associadas à tecnologia utilizada (o tipo de CODEC, no

caso de VoIP); "A"corresponde ao fator de vantagem, ou fator de expectativa.

No cômputo de "Ro", leva-se em conta o ruído acrescentado pelo circuito e o ruído ambiente no lado receptor e no lado emissor. Segundo a Tabela 2/G.107 da ITU-T Recommendation G.107, o valor padrão de "Ro"é 94,77.

No que diz respeito ao fator degenerativo simultâneo "Is", dentre as perdas mais ou menos simultâneas ao sinal de voz, estão a queda em qualidade devido a uma conexão de volume demasiadamente alto, perdas causadas pela interferência da própria voz do locutor ao microfone sobre o fone de ouvido do mesmo handset e a distorção de quantização causada pelos codecs de PCM e ADPCM (as perdas causadas pelos codecs de baixa taxa de bits são computadas separadamente na parcela "Ie"do fator "R"). Seu valor padrão é 1,43[54].

"Id", o fator degenerativo de atraso, resulta de transmissões degenerativas resultantes de atrasos na rede, compreende a soma de perdas devido ao eco no transmissor e no receptor e as perdas relacionadas ao atraso absoluto da voz.

O valor de "Ie"depende do tipo de codificação a baixa taxa de bits e da perda de pacotes no canal de comunicação ou nos *buffers de jitter*. Assim, sua caracterização depende de resultados de testes com cada tecnologia em particular (VoIP, wireless, entre outras). Alguns valores provisórios de "Ie"estão expostos na Tabela 2a da recomendação ITU-T G.108 para planejamento de redes, supondo que não há perdas de pacotes. A tabela 2b/G.108 ITU-T oferece valores provisórios de Ie sob condições de até 16% de perda de pacotes para os codecs G.729 e G.723.1.

O fator "A"de expectativa é utilizado para definir o grau de tolerância que um usuário médio espera pelo uso de uma tecnologia, ou seja, por exemplo, se utilizadores estão cientes de que estão se comunicando com uma localidade de difícil alcance através de comunicações

satélites de múltplos saltos, serão mais tolerantes com as degenerações devido a longos atrasos. Para a telefonia fixa e tecnologia VoIP o seu valor é zero, $A = 10$ para telefonia celular e $A = 20$ para comunicação via satélite [22].

Destes fatores, os que mais afetam o desempenho da comunicação VoIP são o "Id"e o "Ie". Existem na literatura alguns trabalhos que propõem métodos para efetuar o levantamento desses fatores [23].

Assim que as transmissões degenerativas numa rede IP forem medidas, o E-Model poderá ser usado para calcular o fator de avaliação e poderá então ser transformado para o modelo MOS (*Mean Opinion Score*) de acordo com as seguintes equações:

$$Para : R < 0, MOS = 1 \tag{6.4}$$

$$Para : 0 < R < 100, MOS = 1 + 0.035R + 7R \times (R - 60) \times (100 - R) \times 10e^{-}6 \tag{6.5}$$

$$Para : R > 100, MOS = 4.5 \tag{6.6}$$

Valor de R	MOS	Satisfação do usuário
90	4,34	Usuários muito satisfeitos
80	4,03	Usuários satisfeitos
70	3,60	Alguns usuários insatisfeitos
60	3,10	Muitos usuários insatisfeitos
50	2,58	Quase nenhum usuário satisfeito

Tabela 6.6: Tabela comparativa do MOS com o modelo E

Considerando os valores padrões e que a comunicação ocorrerá utilizando a tecnologia VOIP, temos:

$$R = Ro - Is - Id - Ie + A \tag{6.7}$$

$$Ro = 15 - 1.5 * (SLR + No) \tag{6.8}$$

Onde SLR é a taxa de ruído de envio, e No é a adição de potência devido às diferentes fontes de ruído. O valor default da SLR é +8 dB [54], enquanto o valor de No pode ser calculado utilizando as expressões definidas na Recomendação G.107[54]. O resultado é: $Ro = 94.77$.

$$Is = Iolr + Ist + Iq \tag{6.9}$$

$Iolr$ representa o decréscimo de qualidade causado pelos valores muito pequenos de OLR (soma das taxas de ruídos de envio e recebimento), Ist é a perturbação causada pelo "sidetone"não-ótimo, e Iq é a perturbação devido a distorção na quantização. Os valores podem ser calculados utilizando as expressões apresentadas na Recomendação G.107 [54]. O resultado é: $Is = 1.43$, onde:

$$R = 94,77 - 1,43 - Id - Ie + 0 \tag{6.10}$$

$$R = 93,34 - Id - Ie \tag{6.11}$$

Cálculo de "Id"

"Id"representa o fator de atraso ou latência da rede, Onde "d"representa o atraso total na rede em milissegundos.

Para $d < 177.3$

$$Id = 0.024d \tag{6.12}$$

Para $d > 177.3$

$$Id = 0.024 + 0.11(d - 177.3) \tag{6.13}$$

O valor de "d"é dado pela equação:

$$d = dc + dn + dj \tag{6.14}$$

Onde "dc"representa o atraso relativo ao codec, "dn"o atraso dos meios de transmissão da rede e "dj"o atraso do *jitter buffer*. dc (g.711) = 0ms ; dc (g.729) = 30ms e dc (g.723.1) = 80ms.

Apesar de ser um método computacionalmente simples (todos os elementos que influenciam o cálculo do fator R encontram-se tabelados na ITU-T (Recomendação G.113[66])), o modelo E foi desenvolvido com o objetivo de prover uma estimativa da qualidade de voz durante o planejamento de redes de comunicações. Para medições constantes da qualidade da fala numa rede já existente, grupos de pesquisa têm estudado métodos de incorporar fatores que aprimorem o cômputo de "R"[22].

Os parâmetros que influenciam no cálculo do fator "R"não levam em consideração a variação no atraso, o *Jitter*, sendo que o mesmo influencia decisivamente na qualidade da voz.

Recomendação P.564

A Recomendação P.564[61] do ITU-T apresenta testes de conformidade para análise de métodos avaliação da qualidade da voz em redes voz sobre IP. É um esforço para definir normas de funcionamento para esses modelos e apresentar uma metodologia para medir e comparar objetivamente a precisão de seus resultados. A P.564 estabelece critérios mínimos para a avaliação da qualidade da voz para métodos que usam dados objetivos para avaliar o impacto dos parâmetros de uma rede IP na avaliação qualidade da audição em único sentido. Originalmente desenvolvida para aplicações de telefonia de banda estreita (3,1

kHz), a recomendação foi estendida para incluir a banda larga de telefonia (7 kHz), em Novembro de 2007.

Os modelos que estão de acordo com a recomendação P.564 têm as seguintes características:

- Produz uma escala comparativa com o método Mean Opinion Scores (MOS) no modo de audição de qualidade ACR, na faixa de 1 a 5, onde 1 representa "inaceitável"e 5 "excelente."

- Avalia a qualidade da voz sem se referir a real informação contida em um pacote de voz, ou seja, independente do conteúdo do discurso do fluxo RTP analisado.

- Considera o impacto do codec utilizado na voz, mas não considera o volume da voz, nível de ruído de fundo, atraso, nível de retorno de aúdio, eco, ou outras deficiências que podem interferir na qualidade de uma chamada.

- Pode ser implantado em equipamentos em pontos terminais da rede, com agentes embutidos, em pontos centrais da rede, ou com uma combinação de ambos.

A precisão de cada modelo de avaliação é determinada pela comparação do desempenho com o P.862.1 PESQ, algoritmo de referência, utilizando no teste pré vetores criados com um conjunto de quatro amostras de 8 segundos, que são arquivos de discursos que estão incluídos como anexos na recomendação P.564. A conformidade com a recomendação p.564 contribui para garantir níveis mínimos de coerência e precisão na avaliação da qualidade dos modelos, define uma metodologia objetiva para medir a sua exatidão e, assim, permite avaliar a precisão de vários modelos.

Um objetivo específico da recomendação P.564 é a redução dos erros de "falsos positivos"e "falsos negativos". Isto ocorre quando determinado discurso após avaliação, ap-

resenta pontuação de qualidade que são muito altas ou baixas, respectivamente. Isto é particularmente importante quando se fala de avaliação da qualidade de modelos que são usados para monitorar a conformidade com acordos de níveis de serviço (SLAs), uma situação em que resultados imprecisos podem ter um impacto financeiro direto, tanto sobre o prestador do serviço como sobre o cliente.

Apresentamos na tabela 6.7 um resumo dos métodos de avaliação de qualidade da voz citados neste trabalho.

Método	Tipo	Norma	Ano	Órgão	Eficiência	Dific. de medição	Observação
MOS	Subjetivo	P.800/830	1996	ITU-T	Máxima	Máxima	Referência para todos os métodos
PSQM	Objetivo, intrusivo	P.861	1996/99	ITU-T	média	grande	Primeiro método objetivo
MNB	Objetivo, intrusivo	Apêndice II P.861	1998	ITU-T	média	grande	Baseado no PSQM
PAMS	Objetivo, intrusivo	—-	1998	Psytechnics	Alta	Grande	Pontuação de qualidade auditiva e esforço de compreensão
PESQ	Objetivo, intrusivo	P.862	2001	ITU-T	Alta	Grande	Combina os méritos do PSQM e do PAMS
SQUAD	Objetivo, intrusivo	—-	2007	SwissQual	Alta	Grande	Baseado no PESQ, Adiciona um detector de degradação
INMD	Objetivo, não intrusivo	P.561	1996	ITU-T	Média	Baixa	Não considera os parâmetros combinados
CCI	Objetivo, não intrusivo	P.562	2000	ITU-T	Média	Baixa	Evolução do INMD
NINA	Objetivo, não intrusivo	—-	2001	SwissQual	Acima da Média	Baixa	Cria um sinal de referência e compara com o degradado.
P.563	Objetivo, não intrusivo	P.563	2004	ITU-T	Acima da Média	Baixa	Trabalho conjunto da SwissQual, Psytechnics e Opticom
E MODEL	Objetivo, Baseado em Parâmetros	G.107	2000	ITU-T	Média	Baixa	Não considera o Jitter

Tabela 6.7: Tabela comparativa de métodos de medição de qualidade

Capítulo 7

Medidas experimentais de qualidade de voz

7.1 Objetivo dos testes

O objetivo deste capítulo é fazer um estudo comparativo entre os principais métodos desenvolvidos para a medida de qualidade de voz (PAMS, PESQ e Modelo E), caracterizando como cada um deles obtém um parâmetro de medida e abordando os respectivos méritos e deficiências. Além disto, verificar qual método seria mais adequado para as tecnologias utilizadas nos sistemas de telecomunicações, hoje em operação, VoIP e TDM.

Para realização dos testes foi utilizado o equipamento PERFORMER da RAD TELECOM. O Performer é uma ferramenta indicada para medições em redes de voz (VoIP, TDM e etc.). Dentre as suas principais características estão o suporte a multiprotocolos (H.323, SIP, MGCP, entre outros), medição objetiva da qualidade (PAMS e PESQ), medição *online* de *jitter* , *call quality*, possibilidade de geração de chamadas simultâneas. O manual descritivo do PERFORMER está contido no apêndice "B". Os testes foram realizados no Centro de Referência Tecnológica da EMBRATEL, com o apoio dos engenheiros Anderson

Cardoso de Jesus, Adriano Alves Borges da Silveira e Marcelo Gomes. De modo a definir o escopo de experimentos, foram realizados testes em duas tecnologias: o TDM tradicional e o Voz sobre IP, utilizando *gateways* de voz.

7.2 Processo de Medição

O processo de medição ocorre quando uma chamada é realizada entre os números telefônicos que foram configurados nos equipamentos diferentes. Como o equipamento de medição tem conectado em suas interfaces as duas linhas telefônicas, hipoteticamente localizadas em locais diferentes, é possível fazer uma comparação entre um arquivo de áudio que é originado da linha telefônica do número chamador ao sinal de voz recebido no outro lado da rede na linha que está sendo chamada. Em cada medição da qualidade foram realizadas 5 chamadas de 57,6 segundos e dos valores obtidos extraído o valor médio.

No Performer existem diversos arquivos de referência de voz, sendo um por idioma. Em todo o processo foi escolhido o idioma português para geração do sinal de voz a ser comparado. No caso do idioma português, a fala utilizada como referência segue o seguinte texto: "escolinhas de esporte fazem o lazer dos fins de semana"(voz de mulher),"vai interpretar um solteirão conservador"(voz de mulher) e " a diretoria técnica pretende ampliar o número de pontos de testes de propagação"(voz de homem). Os textos são diferentes de acordo com o idioma selecionado para medição no equipamento.

Em todos os casos foram obtidas medidas referentes aos métodos de medição da qualidade disponíveis no "PERFORMER", ou seja os métodos PESQ e PAMS.

7.3 Experimento na tecnologia TDM

Devido ao fato de que na tecnologia TDM a comunicação de voz utiliza a codificação G.711, com uma taxa de 64 Kbits/s por canal de voz e que não há concorrência com aplicações de dados disputando a mesma interface do dispositivo, é natural que a qualidade da voz obtida seja maior do que aquela que será experimentada com a tecnologia VoIP. Esta medição foi realizada para exprimir numericamente este contraste e tornar clara a importância da evolução dos codificadores de voz e das técnicas de qualidade de serviço para se conseguir chegar próximo do nível de qualidade obtido pela telefonia convencional, a custa de um uso muito maior de banda que a tecnolgia VoIP, onde a convergência das aplicações de voz e dados por um único canal ou circuito é uma realidade que tem como característica negativa a redução significativa na qualidade da voz, na busca de um uso mais racional da banda disponível.

7.3.1 Infra-estrutura TDM

Para simular um ambiente de telefonia convencional, foram criados dois ambientes. No primeiro é apresentado o melhor dos casos na questão da qualidade. As medidas de qualidade foram realizadas em dois ramais analógicos de uma central telefônica convencional. Esta disposição dos ramais será denominada ambiente TDM 1 e está representada graficamente na Fig 7.1. A central telefônica convencional foi simulada por um equipamento SITEST 600, da marca DIGITRO.

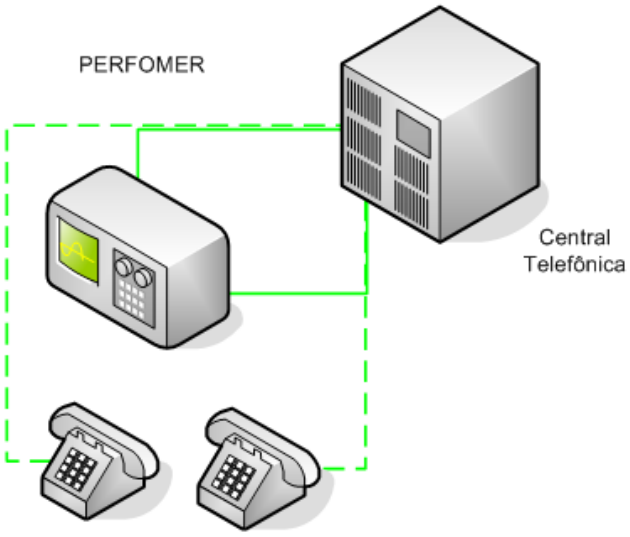


Figura 7.1: Ambiente com dois ramais de uma mesma central telefônica.

Fonte: Centro de Referência Tecnológica da EMBRATEL

Método	CODEC	Med 1	Med 2	Med 3	Med 4	Med 5	Med M	RTD (ms)
PESQ	G.711	4,07	4,08	4,06	4,07	4,07	4,07	9,0
PAMS	G.711	4,11/4,44	4,11/4,44	4,11/4,44	4,11/4,44	4,11/4,44	4,11/4,44	9,0

Tabela 7.1: Tabela de medidas TDM com dois ramais de uma mesma central.

Onde:

Med n - Representa uma medida particular;

Med M - Representa a média aritmética das medidas.

RTD (*Round Trip Delay*) - Representa o tempo total de ida e volta do pacote, em milissegundos, incluindo o processamento nos equipamentos terminais.

Para o cálculo do valor de qualidade correspondente no Modelo E, onde $R = 93,34 - Id - Ie + A$, foram utilizados os seguintes parâmetros: $A = 0$, para a tecnolgia TDM e $Ie = 0$ para o CODEC G.711. Os valores do fator "R"e "MOS E"apresentados na

tabela 7.2 foram obtidos através uso do calculador disponibilizado pelo ITU-T no endereço www.itu-t.org/e-model.

CODEC	PESQ	PAMS	MOD E
G.711	4,07	4,11	4,41

Tabela 7.2: Tabela de valores médios dos métodos na tecnologia TDM caso 1.

Na tabela 7.2 observa-se que os valores de qualidade apresentados são elevados e a voz ouvida na comunicacão foi clara e de excelente qualidade. De acordo com a recomendação G.107[54], os valores medidos estão em um patamar que indica que os usuários ficariam satisfeitos em um diálogo entre os aparelhos telefônicos utilizados. Esperava-se neste caso um valor maior, pois trata-se da melhor situação possível, dois ramais de uma mesma central, sem tráfego concorrente e utilizando o codificador G.711, porém, como já foi citado anteriormente, trata-se de um teste com um simulador de central e não uma central telefônica propriamente dita. Os valores obtidos pelo algoritimo PAMS se mostraram sempre superiores aos apresentados pelo método PESQ. Outro fato relevante é que quando utiliza-se o método PAMS verifica-se que são apresentados 2 valores, o primeiro indica a qualidade auditiva, que também é representado no método PESQ, e o segundo o esforço de compreensão, somente mostrado no PAMS. No que se refere ao valor obtido através do cálculo do modelo E, apresentado na tabela 7.2, verifica-se um elevado valor, muito acima daqueles obtidos pelos métodos intrusivos PESQ e PAMS.

No segundo ambiente foram conectadas as interfaces do PERFORMER a dois ramais de centrais telefônicas diferentes, interligadas por uma rede determinística TDM (interface E1 com 30 canais). As centrais telefônicas convencionais foram simuladas por dois equipamentos SITEST 600, da marca DIGITRO. Esta topologia esta representada na Fig 7.2.

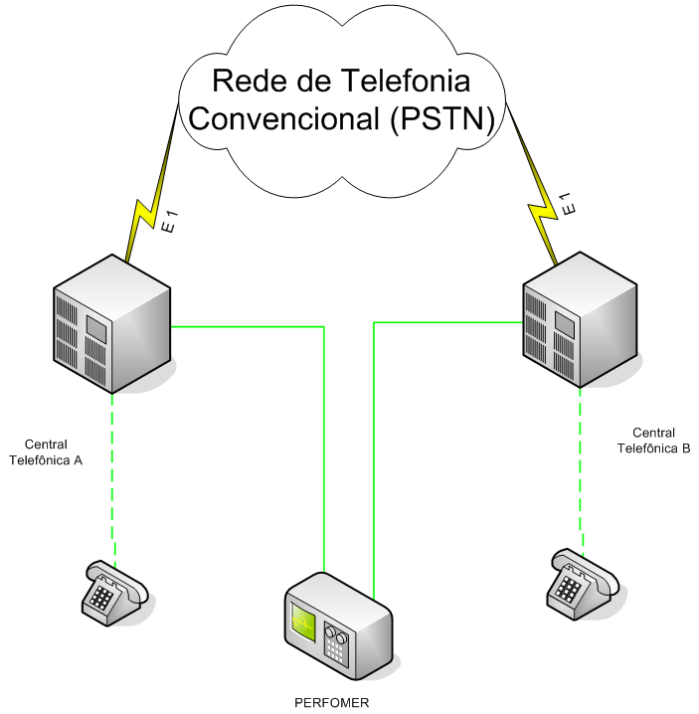


Figura 7.2: Ambiente com dois ramais em centrais telefônicas diferentes.

Fonte: Centro de Referência Tecnológica da EMBRATEL

Método	CODEC	Med 1	Med 2	Med 3	Med 4	Med 5	Med M	RTD (ms)
PESQ	G.711	3,89	3,90	3,88	3,98	3,91	3,91	13,0
PAM	G.711	3,96/4,33	3,96/4,33	3,95/4,32	3,96/4,33	3,96/4,33	3,96/4,33	13,0

Tabela 7.3: Tabela de medidas TDM entre 02 centrais telefônicas.

Neste caso, observa-se ainda na tabela 7.3 um elevado valor de qualidade nas medições

realizadas e uma pequena redução na clareza da voz em relação ao ensaio anterior. A diminuição na qualidade é justificada, pois neste caso a voz passa por um enlace E1 (30 canais de 64 Kbits/s) que interliga as duas centrais além do processamento interno de cada central a que estão conectados os telefones. Este padrão de qualidade é muito próximo do que experimentamos quando fazemos uma ligação local (dentro da mesma cidade) pelos sistemas de telefonia convencional.

CODEC	PESQ	PAMS	MOD E
G.711	3,91	3,96	4,40

Tabela 7.4: Tabela de valores médios dos métodos na tecnologia TDM caso 2.

O valor verificado no modelo E, tabela 7.4, continuou muito acima dos demais métodos, com uma pequena redução na medida (0,01).

7.4 Experimento na tecnologia *VoIP*

Para melhor interpretar os parâmetros envolvidos na medição da qualidade da voz na tecnologia VoIP, o processo de medição foi dividido em 03 casos distintos, de forma que gradativamente pudesse ser analisado um número maior de parâmetros atuando simultaneamente. A idéia principal é, em cada medição de qualidade, de determinado método, alterar o codificador de voz, partindo de um de melhor qualidade e acompanhar a variação dos valores objetivos de qualidade apresentados pelo método em relação à percepção do usuário. Em todas as medidas foram alternados os codificadores de voz entre os padrões G.711, G.729 e G.723, de maior uso no mercado de telecomunicações, de modo a identificar o impacto que os *CODECS* causam na qualidade da voz.

7.4.1 Infra-estrutura *VoIP*

Foi criado o seguinte cenário: Foram instalados dois roteadores do fabricante CISCO, modelos 2811 e 2801, com portas FXS que simulam as funções de centrais telefônicas. Os referidos roteadores foram configurados em redes diferentes e interligados ao *backbone* do Centro de Referência Tecnológica da EMBRATEL através de interfaces seriais com taxa de transferência de 1984 Kbits/s. Para simular o tráfego concorrente das aplicações de dados com os pacotes de voz em uma rede de comunicação convencional, foram instalados dois geradores de tráfego FST-2802 do fabricante ACTERNA. Nas interfaces dos roteadores onde seriam conectados os telefones, foi ligado o equipamento de medição PERFORMER. Para estabelecimento das chamadas foi utilizado o protocolo de sinalização SIP e como servidor de usuários VoIP, o servidor SIP da EMBRATEL. Em nenhum dos ambientes montados foi verificada perda de pacotes nos enlaces de comunicações.

Onde:

SBC (*session border controller*) - Dispositivo usado para controle de sinalização SIP.

SoftX (*Softswitch*) - servidor de usuários SIP.

AS2905 - Indicativo da vrf (*Virtual Routing and Forwarding*), instância de roteamento usada nos testes.

GSR12 - Roteador de grande porte do backbone do CRT.

ROT01 - Roteador de grande porte do backbone do CRT.

DATACOM - Conversor fracional TDM.

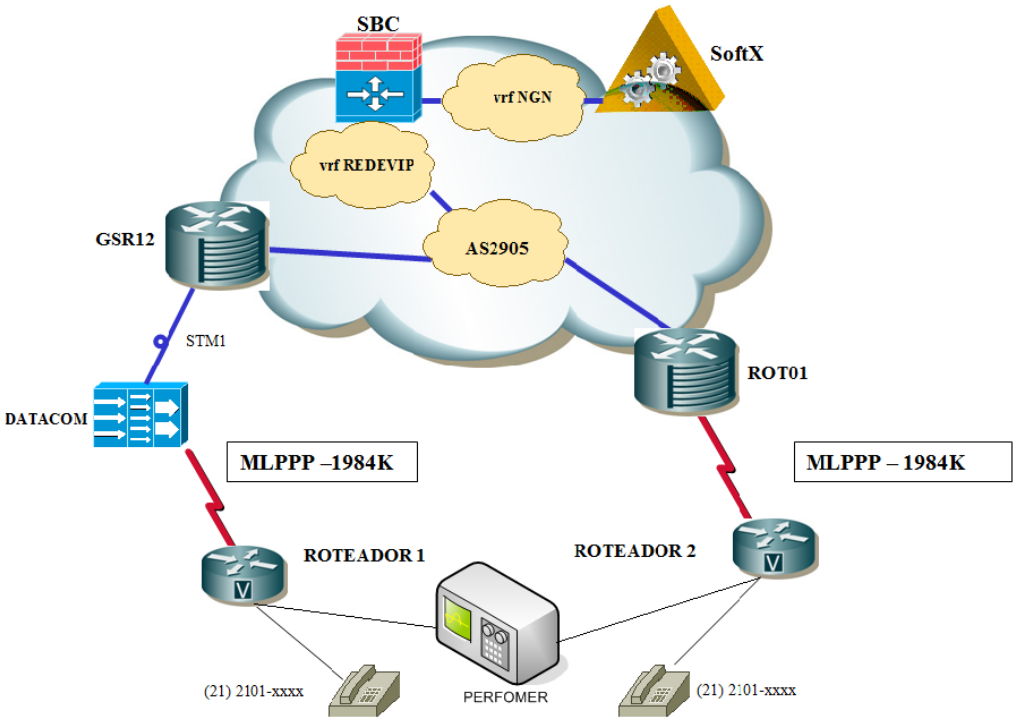


Figura 7.3: Cenário montado para as medições na tecnologia VoIP.

Fonte: Centro de Referência Tecnológica da EMBRATEL

Caso 1

No primeiro caso foi configurada uma rede sem nenhum tráfego concorrente e sem nenhuma preparação para utilizar a tecnologia VoIP, ou seja, não foi aplicada nenhuma técnica de qualidade de serviço que priorizasse os pacotes de voz ou fosse reservada uma banda específica para aplicações de voz. Trata-se de um cenário pouco real, mas que retrata bem o nível de qualidade que cada codificador de voz impõem ao sistema. Para as características citadas, as seguinte medidas foram obtidas.

Método	CODEC	Med 1	Med 2	Med 3	Med 4	Med 5	Med M	RTD (ms)
PESQ	G.723	3,63	3,67	3,66	3,63	3,67	3,65	263,13
PESQ	G.729	3,65	3,63	3,65	3,64	3,65	3,65	174,13
PESQ	G.711	4,06	4,06	4,03	4,07	4,07	4,06	156,38
PAMS	G.723	4,17/4,39	4,15/4,38	4,15/4,38	4,13/4,36	4,14/4,37	4,15/4,37	251,13
PAMS	G.729	4,04/4,29	3,95/4,22	4,01/4,26	3,95/4,22	3,99/4,26	3,99/4,25	172,13
PAMS	G.711	4,12/4,43	4,14/4,45	4,11/4,37	4,13/4,45	4,11/4,43	4,12/4,43	155,38

Tabela 7.5: Tabela de medidas de VOIP, sem QOS e sem tráfego concorrente.

Para o cálculo do valor de qualidade correspondente no Modelo E foram utilizados os seguintes parâmetros: $A = 0$, para a tecnologia VoIP e $Ie = 0$ para o CODEC G.711, $Ie = 10$ para o CODEC G.729 e $Ie = 15$ para o CODEC G.723[66]. Ie representa a perda associada ao CODEC. O codificador G.723 utilizado é o que ocupa a banda de 6,3 Kbits/s em cada chamada.

Método	CODEC	Ie	RTD (ms)	Fator R	MOS E
MOD E	G.723	15	263,13	75	3,82
MOD E	G.729	10	174,13	80,9	4,06
MOD E	G.711	00	156,38	91,1	4,37

Tabela 7.6: Medidas de qualidade, de acordo com o Modelo E.

Os valores do fator "R"e "MOS E"apresentados na tabela 7.6 foram obtidos através uso do calculador disponibilizado pelo ITU-T no endereço www.itu-t.org/e-model.

Este primeiro caso com o uso da tecnolgia VoIP, trata-se de um ambiente irreal, pois na prática sempre haverá tráfego concorrente de dados. O ensaio possui aplicabilidade para avaliação dos codificadores. Pelos valores obtidos, observa-se a melhor qualidade da

comunicação quando da utilização do CODEC G.711. No que se refere aos CODECS G.729 e G.723, a diferença nas medidas foi pouco significativa entre os dois codificadores.

Considera-se que na faixa do MOS entre 4,01 e 4,30 a qualidade da voz é ótima sem nenhuma objeção. No intervalo entre 3,60 e 4,00; a qualidade da voz é boa, com poucos usuários insatisfeitos. Na faixa entre 3,10 e 3,59 a qualidade é razoável, já existindo uma quantidade significativa de reclamações, abaixo destes valores fica muito difícil a comunicação[54].

CODEC	PESQ	PAMS	MOD E	MOS
G.711	4,06	4,12	4,37	4,3
G.729	3,65	3,99	4,06	3,92
G.723	3,65	4,15	3,82	3,90

Tabela 7.7: Tabela comparativa de valores médios dos métodos no caso 1.

De acordo com a tabela 7.5, observou-se que o método PESQ subestimou um pouco a qualidade da voz, no que diz respeito ao codificadores G.729 e G.723, porém, representou bem a percepção de igualdade entre os mesmos. O método PAMS não apresentou resultados coerentes em relação ao CODEC G.723, ou seja, o valor numérico apresentado não condizia com a realidade, isto é, a percepção auditiva não foi de ótima qualidade. O modelo E não conseguiu mostrar a semelhança no desempenho dos CODECS G.729 e G.723 e superestimou a qualidade do codificador G.711. Em resumo, o método PESQ representou melhor numericamente a boa qualidade da voz apresentada pelos codificadores G.729 e G.723 e o ótimo desempenho com o CODEC G.711. A coluna "MOS" mostra o valor máximo esperado para os CODECs[58].

Caso 2

No segundo caso foi simulado um elevado tráfego de dados TCP entre as redes, nos sentidos de transmissão e recepção, a uma taxa de 1980 Kbits/s, o que ocupou totalmente o enlace de dados entre os roteadores. O referido fluxo foi estabelecido com pacotes de 1024 bytes, tamanho muito superior a um típico pacote de voz de 60 bytes, fator importante para simular a variação no atraso causada pela serialização de pacotes com tamanhos muito diferentes. Como não foi aplicada nenhuma técnica de qualidade de serviço, ou seja, não foi estabelecida nem prioridade, nem reserva de banda para os pacotes de voz, os valores medidos tiveram um queda acentuada em seus valores, fato que reflete a baixa qualidade obtida no entendimento da fala neste experimento.

Método	CODEC	Med 1	Med 2	Med 3	Med 4	Med 5	Med M	RTD (ms)
PESQ	G.723	3,17	3,28	3,17	3,26	3,07	3,19	293,62
PESQ	G.729	3,17	3,28	3,29	3,38	3,43	3,31	201,75
PESQ	G.711	2,76	2,3	2,39	2,22	2,08	2,35	184,57
PAM	G.723	3,81/4,06	3,85/4,05	3,73/3,94	3,63/3,93	3,47/3,68	3,69/3,93	293,02
PAM	G.729	2,87/3,34	2,98/3,46	3,03/3,47	3,30/3,71	3,03/3,47	3,04/3,49	201,15
PAM	G.711	1,0/1,33	1,23/2,02	1,00/1,61	1,60/2,1	1,15/1,54	1,20/1,17	183,90

Tabela 7.8: Tabela de medidas de VOIP, sem QOS e com tráfego concorrente.

Na tabela 7.8 observa-se que os CODECS que ocupam menor banda (G.729 e G.723) tiveram um resultado melhor do que o G.711, devido a dificuldade de usar uma quantidade de banda para efetivar uma comunicação intelegível. O método PAMS superestimou os valores do codificador G.723. De acordo com a percepção auditiva, o CODEC G.729 teve um melhor desempenho que o G.723, conforme indicou o método PESQ.

A realidade apresentada neste caso assemelha-se muito com a INTERNET, onde existe elevada concorrência de tráfego, mas não existe qualidade de serviço.

CODEC	PESQ	PAMS	MOD E
G.711	2,35	1,20	2,23
G.729	3,31	3,04	2,22
G.723	3,19	3,69	2,17

Tabela 7.9: Tabela comparativa de valores médios dos métodos no caso 2.

Neste caso como houve grande tráfego concorrente, ocorreram perdas de pacotes em média de 1%, nas interfaces seriais dos roteadores, fato que influiu decisivamente na qualidade da voz. De acordo com a tabela 7.9, observa-se que neste casos onde o ambiente é adverso para o trafego dos pacotes de voz, o Modelo E, método baseado em parâmetros, apresentou valores irreais, pois não conseguiu diferenciar a qualidade entre os CODECs. A voz com o codificador G.711 ficou muito pior que a dos codecs G.729 e G.723. O método PAMS mostrou-se inconsistente na avaliação de qualidade do codificador G.723, pois, apresentou o valor de "3,69"que indica uma qualidade boa para a comunicação medida, fato que não retratou a realidade, porque a voz se apresentou, em alguns momentos, intercortada e de difícil entendimento. O método PESQ mostrou-se coerente, pois o valor apresentado nos codificadores G.723 e G.729 está de acordo com a percepção de qualidade razoável, com alguns usuários insatisfeitos. Dentre os codificadores testados, o G.729 mostrou-se mais adequado para ambientes típicos de INTERNET, apesar de utilizar uma largura de banda maior que o G.723. Este fato pode ser explicado por uma característica básica do codificador G.729, o tamanho de sua janela de amostragem. O G.729 possui janela de amostragem de 10 ms[30], já o G.723 tem uma janela de 30 ms[31]; como na INTERNET

o ambiente é sujeito a grandes variações de atraso oriundas de tráfego concorrente; quanto menor a janela de amostragem, menor a chance de naquele intervalo ocorrerem oscilações.

Caso 3

No terceiro caso foram simuladas as características de uma rede de comunicação de dados corporativa, onde existe tráfego concorrente, porém são aplicadas técnicas de qualidade de serviço que proporcionam os pacotes de voz: prioridade e reserva de banda suficientes para uma boa qualidade da voz, quando transmitida pela tecnologia VoIP.

Método	CODEC	Med 1	Med 2	Med 3	Med 4	Med 5	Med M	RTD (ms)
PESQ	G.723	3,40	3,60	3,62	3,59	3,65	3,57	293,62
PESQ	G.729	3,58	3,46	3,38	3,39	3,34	3,43	201,75
PESQ	G.711	3,80	3,91	3,94	3,74	3,85	3,85	184,57
PAM	G.723	4,17/4,31	4,15/4,39	4,06/4,27	3,55/3,73	4,14/4,37	4,01/4,23	293,02
PAM	G.729	3,43/3,79	3,68/4,00	3,76/4,07	3,52/3,80	3,64/3,92	3,61/3,92	201,15
PAM	G.711	3,97/4,00	4,08/4,32	3,90/3,89	3,97/3,99	3,93/4,23	3,97/4,09	183,90

Tabela 7.10: Tabela de medidas de VOIP, com QOS e com tráfego concorrente.

Foi utilizada a técnica de Qualidade de Serviço chamada *Low Latency Queue*, onde foi criada uma classe para voz e outra para dados. As classes representam as filas na interface do roteador, sendo que a fila específica para voz tem prioridade sobre o tráfego de dados. Foi reservada para classe de voz 200 Kbits/s, possibilitando sempre comunicação livre independente do congestionamento simulado.

O primeiro fato que merece comentário sobre este experimento é a constatação de que o modelo E não se aplica para redes onde é utilizada qualidade de serviço. Como o referido método utiliza parâmetros globais da rede, o valor que dimensiona a qualidade

CODEC	PESQ	PAMS	MOD E
G.711	3,85	3,97	2,23
G.729	3,43	3,61	2,22
G.723	3,57	4,01	2,17

Tabela 7.11: Tabela comparativa de valores médios dos métodos no caso 3.

não interpreta de maneira diferente um pacote de voz, que tem tratamento separado e priorizado em relação a um pacote de dados comum. Problemas como latência e perda de pacotes afetam os parâmetros de cálculo da qualidade de forma indiscriminada, ou seja, o pacote de voz, que possui qualidade de serviço, não é medido de forma separada para determinação do fator R.

Outra observação importante é que o codificador G.723 comportou-se melhor em um ambiente de tráfego concorrente com QOS, do que quando não há priorização e reserva de banda. Com o ambiente controlado, quanto maior a janela de amostragem, melhor a representação do sinal original, ou seja, os 30 ms do G.723 permitem um melhor desempenho que os 10 ms do G.729. Este comportamento é muito interessante para as corporações, pois além de utilizarem menos banda poderão ter maior qualidade na comunicação de voz.

O codec G.729 teve desempenho oposto, apresenta melhor qualidade dentre os demais quando existe tráfego concorrente e não há qualidade de serviço, tal qual apresentado no caso 2.

O método PAMS novamente superestimou a qualidade do codificador G.723, pois, apesar de apresentar bom desempenho, não chegou a superar o CODEC G.711. O algorítimo PESQ foi coerente nos valores indicados, com o G.711 chegando próximo aos valores de qualidade da telefonia convencional.

As figuras 7.4, 7.5 e 7.6 mostram os valores de qualidade obtidos nas diversas medições e separados por codificadores. Verifica-se nestes gráficos que o modelo E, através do cômputo da influência de cada elemento de perdas no sistema de comunicação, fornece ainda uma avaliação limitada da qualidade de voz, apresentando valores muito diferentes dos métodos intrusivos PESQ e PAMS.

Onde:

IP-C1 = Medição realizada com a tecnologia VoIP no caso 1.

IP-C2 = Medição realizada com a tecnologia VoIP no caso 2.

IP-C3 = Medição realizada com a tecnologia VoIP no caso 3.

TDM-C1 = Medição realizada com a tecnologia TDM no caso 1.

TDM-C2 = Medição realizada com a tecnologia TDM no caso 2.

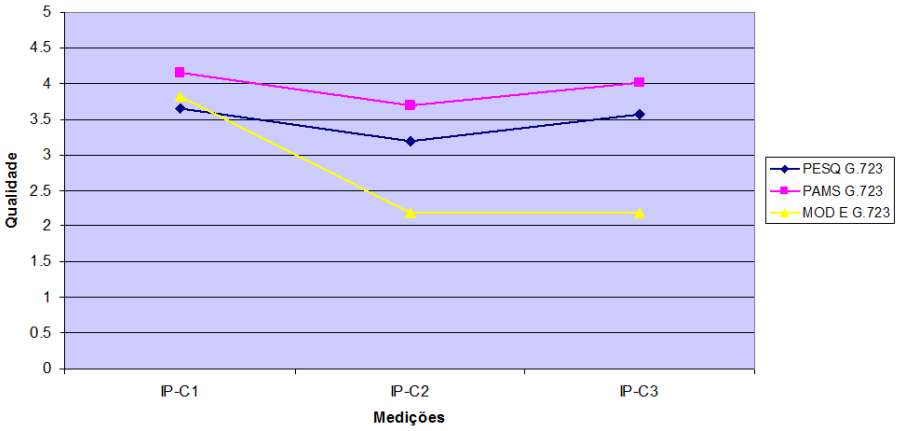


Figura 7.4: Medições de qualidade com o codec G.723

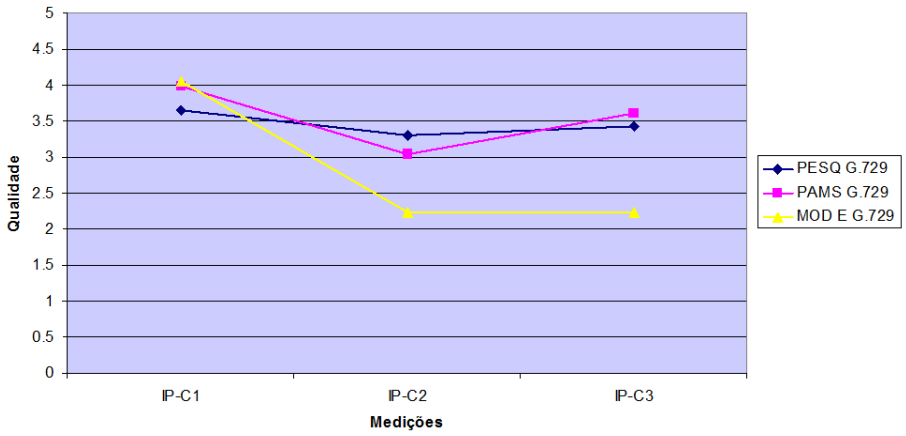


Figura 7.5: Medições de qualidade com o codec G.729

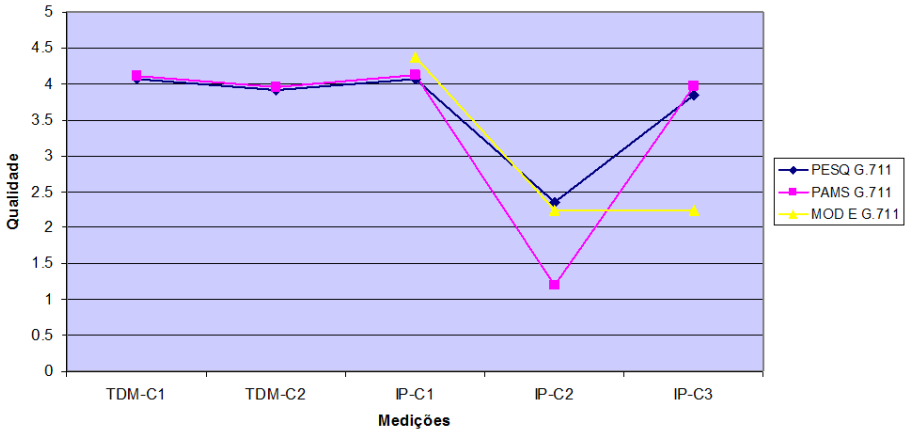


Figura 7.6: Medições de qualidade com o codec G.711

Capítulo 8

Conclusões e trabalhos futuros

8.1 Conclusões

Neste trabalho foram vistos vários fatores que impactam consideravelmente na qualidade final da voz para o usuário dos sistemas de telecomunicações, destes, destacam-se o codec de voz utilizado, os atrasos causados durante a transmissão do sinal, até este chegar ao receptor, e as perdas de pacotes, tendo como efeitos negativos o corte da voz, o eco e o tempo de espera entre falas.

Vimos que as características não determinísticas de redes baseadas na comutação de pacotes têm um impacto negativo na qualidade de voz percebida quando existe concorrência com aplicações de dados. Desenvolver métodos que meçam esse impacto é importante para projetar, implementar, manter e expandir redes VoIP. Nesse sentido, têm surgido iniciativas de estabelecer-se métodos confiáveis para avaliação da qualidade de voz.

Observa-se ainda, que apesar dos grandes avanços, os codificadores e as melhores técnicas de qualidade de serviço não conseguem ainda superar a qualidade de uma chamada realizada pela telefonia convencional. Nos casos onde existe grande concorrência de tráfego e com banda disponível limitada, as reclamações dos usuários tem fundamento. Na IN-

TERNET, conforme relatado no caso 2, como trata-se de um ambiente sem controle de prioridade e reserva de banda, não se pode garantir ou esperar desempenho próximo ao que estamos acostumados na telefonia convencional. Mesmo nas redes corporativas, com QoS, quando se utilizam CODECS que demandam pequena largura de banda, a qualidade fica boa, porém, inferior a da telefonia convencional.

Em relação aos métodos existentes de medição de qualidade da voz, podem ser observadas características próprias sobre os grupos de métodos apresentados, em relação ao custo-benefício da utilização destes.

Os métodos subjetivos são uma referência, pois retornam a qualidade da chamada que retrata a real percepção do usuário, porém são de difícil utilização, visto que tomam muito tempo e não é possível utilizá-los em uma chamada em tempo-real. Servem apenas como base para outros métodos objetivos. Neste tipo de método existem muitas condições de controle necessárias para que as medidas sejam confiáveis: o volume dos locais onde se vai gerar e ouvir a fala devem ter valores determinados, as dimensões das salas devem ser possuir proporcionlidade, de modo a reduzir ao mínimo os efeitos de ondas estacionárias, o tempo de reverberação nas salas precisa ser menor que 500ms, os ouvintes não podem ser pessoas envolvidas em trabalhos relacionados a performance de sistemas de comunicações, os ouvintes não podem ter participado em nenhum teste subjetivo nos últimos seis meses e não ter participado de testes de opinião de audição no último ano, o número de ouvintes deve ser de no mínimo 06 pessoas, sendo o ideal maior que 12[4]. Apesar das dificuldade de reprodutibilidade e de consumo de tempo, representam a referência para os métodos objetivos, pois são uma amostra da opinião do usuário final sobre a qualidade da voz ouvida.

Os métodos objetivos intrusivos baseados em sinal, apesar de serem os métodos objetivos mais precisos na predição da qualidade de uma chamada de voz, são onerosos pois necessitam de equipamentos de testes complexos, visto que interferem no tráfego da voz transmitida durante a conexão, resultando em um baixo custo/benefício para as empresas que oferecem serviços de voz por uma rede de comutação de pacotes, e para seus usuários. Na prática este tipo de método serve para testes em laborátórios, pois para sua utilização é necessário ter acesso às duas pontas de um sistema de comunicação a ser medido. Em uma rede operacional seria economicamente inviável manter equipamentos desta monta em cada localidade, para monitorar a qualidade das chamadas de voz.

Entre os métodos objetivos intrusivos de avaliação da qualidade de voz, destacam-se os perceptuais, que procuram modelar matematicamente como o homem capta a fala e a avalia sua qualidade. Métodos tais como o PAMS e o PESQ, utilizados nos ensaios descritos nesta dissertação, têm grande aceitação e oferecem uma boa avaliação da qualidade da voz ao longo de sistemas que apresentam problemas de perdas de pacotes e atraso variável. O método PESQ mostrou-se mais adequado para a gama de situações possíveis de uma rede, ou seja, rede com pouco tráfego, com muito tráfego concorrente, sejam elas com ou sem aplicação de técnicas de qualidade de serviço.

Inicialmente desenvolvido para prever a qualidade da voz na fase de projeto da rede, o modelo E tem sido cada vez mais estudado para prover uma medida objetiva em redes em pleno funcionamento, devido à sua simplicidade computacional. Até onde se obteve conhecimento sobre esta forma de medição, ainda não é confiável para redes onde existe grande congestionamento ou concorrência significativa entre aplicações de voz e dados. Existe uma nova versão do modelo E, chamada *Extended E Model* que apresenta melhores

resultados na avaliação da qualidade da voz, com valores mais próximos dos métodos subjetivos e objetivos intrusivos[23][75].

De acordo com os testes realizados, o método PESQ mostrou-se superior ao PAMS e ao modelo E, pois os valores apresentados são mais coerentes com a percepção do ouvido humano, nas diversas situações simuladas. Este modelo é sem dúvida uma excelente ferramenta para aferir os diversos sistemas de telecomunicações com que convivemos diariamente e uma forma eficaz de avaliar os serviços de voz contratados e projetados por empresas e instituições.

8.2 Trabalhos futuros

Como idéia para trabalhos futuros, sugere-se a realização de testes com métodos não-intrusivos baseados em sinal, pois o mercado sinaliza como sendo a tendência de utilização por parte das operadoras de telecomunicações. Os métodos não-intrusivos baseados em sinal, assim como os baseados em parâmetros, possuem um melhor custo/benefício, pois não interferem no tráfego de dados, são baseados em algoritmos e podem ser executados em tempo-real sobre uma chamada de voz, tendo a possibilidade de se complementarem, como é o caso do INMD e CCI com o E-Model, retornando uma predição menos precisa do que os outros métodos intrusivos, porém, as vezes satisfatória.

A grande dificuldade é que estas novas formas de medição ainda são técnicas proprietárias em desenvolvimento por grandes empresas do mercado, como por exemplo, a Telchemy que possui o VQMon - *End Point and Stream Analyzers*[74], produto baseado na combinação de métodos não-intrusivos baseados em sinal e por parâmetros. Seria muito interessante o desenvolvimento de uma aplicação aberta, capaz de medir a qualidade através

de um método não-intrusivo baseado em sinal.

Outro assunto que merece trabalhos de pesquisa é o aprimoramento ou ampliação dos parâmetros do modelo E, de modo que o mesmo possa ser capaz de medir a real qualidade da voz de uma rede em produção, com concorrência entre voz e dados.

Um tópico de grande importância, que merece uma intensa pesquisa, é a avaliação da qualidade nas redes de telefonia celular, com seus codificadores próprios e referência de qualidade muito inferior à telefonia convencional.

Bibliografia

[1] Titze, I. R., "Principles of Voice Production"Prentice Hall, USA, 1994.

[2] Mitra,S. K., "Digital Signal Processing - A Computer-Based Approach", The McGraw-Hill Companies, Inc., 1998.

[3] Anderson, John , "Methods for Measuring Perceptual Speech Quality", Agilent Technologies White Paper, USA, 2001.

[4] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", Genève, 1996.

[5] ITU-T Recommendation P.830, "Subjective performance assessment of telephoneband and wideband digital codecs", Genève, 1996.

[6] ITU-T Recommendation P.861, "Objective quality measurement of telephone-band speech codecs", Genève, 1996.

[7] ITU-T Recommendation 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", Genève, 2001.

[8] Gouveia, L. M., "A Fonoaudiologia e o Canto", Universidade Cruzeiro do Sul, http://www.geocities.com/Vienna/9177/fonocanto.html, 2005.

[9] Schroder, G., Sherif, M. H. " The Road to G.729: ITU 8-kb/s Speech Coding Algorithm with Wireline Quality", IEEE Communication Mag. , 1997.

[10] Salami, R. et al, "ITU-T G.729 Annex A: Reduced Complexity 8 kb/s CS-ACELP Codec for Digital Simultaneous Voice and Data", IEEE Communication Mag. , setembro de 1997.

[11] Ramires, M. A., "Codificador Preditivo de Voz por Análise Mediante Síntese", Dissertação de Mestrado, USP, São Paulo, 1992.

[12] Ramires, M. A., "Codificadores preditivos de voz". IX Encontro Nacional da ANPOLL, Caxambu, 1994. Anais, v.2, João Pessoa, 1995.

[13] Martins, L.G., Flores,R., Ghissoni, S., Freitas, J.P., Martins, J.B., Ceretta, R., "Comparação entre os CODECS de compressão de voz iLBC e G.729 considerando atraso e pontuação MOS". UFSM - Universidade Federal de Santa Maria (RS), 2006.

[14] Lucero, J. C., Koenig, L. L., "On the relation between the phonation threshold lung pressure and the oscillation frequency of the vocal folds", Journal of the Acoustical Society of America, 2007.

[15] Quartieri, T. F., "Discrete-Time Speech Signal Processing". Prentice-Hall, 1 edition, 2001.

[16] Vieira, M. N., "Automated Measures of Disphonias and the Phonatory Effects of Asymmetries in the Posterior Larynx", Thesis for the degree of Ph.D., University of Edinburgh, UK, 1997.

[17] Frederico, J. V. C. F., "Novas Contribuições a Verificação Automática de Locutor para Fins Forenses", Dissertação de Mestrado, IME Rio de Janeiro, 2008.

[18] Macedo, J. F.,"Reconhecimento Automático de Identidade Vocal Utilizando Modelagem Híbrida: Paramétrica e Estatística", Tese de doutorado, Campina Grande - PB(UFCG), 2004.

[19] Hersent, O., Gurle, D., Petit, J. P. "Telefonia IP", São Paulo, Addison Wesley, 2002.

[20] Hall, T. A., "Objective Speech Quality Measures for Internet Telephony. In Voice over IP (VoIP) Technology", Petros Mouchtaris, Editor, Proceedings of SPIE, 2001.

[21] Voran, S. D., "Objective Estimation of Perceived Speech Quality Using Measuring Normalizing Blocks", NTIA Report, U.S. Department of Commerce. USA, 1998.

[22] Psytechnics Group,"PAMS: Measuring speech quality over networks as the customers hear it", Psytechnics White Paper, United Kingdom, 2001.

[23] Clark, A., "Comparison of Extended E Model with PSQM and PAMS", Committee T1, Standards ProjectT 1A1-17, USA, 2001.

[24] Miras, D., "A Survey on Network QoS Needs of Advanced Internet Applications", Internet2 - QoS Working Group, 2002.

[25] Barbedo, J. G. A., "Avaliação objetiva de qualidade de codecs de voz na faixa de telefonia", Dissertação de Mestrado, UNICAMP São Paulo, 2001.

[26] Haykin, S., Communication Systems, McMaster University, USA, John Wiley and Sons, 2001.

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

[28] ITU-T Recommendation G.726,"General Aspects of Digital Transmission Systems Terminal Equipments - 40, 32, 24, 16 Kbits/s Adaptive Differential Pulse Code Modulation (ADPCM) ", 1990.

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

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

[31] ITU-T Recommendation G.723.1, "General Aspects of Digital Transmission Systems - Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 Kbits/s", 1996.

[32] Diniz, P. S. R., Da Silva, E. A. B., Netto, S. L., Digital Signal Processing: System Analysis and Design, Cambridge, UK, Cambridge, 2002.

[33] Deller, J. R., Proakis, J. G., Hansen, J. H. L., Discrete Time Process of Speech Signals, New York, NY, USA, MacMillan, 1993.

[34] Maia, R. S., Codificação CELP e Análise Espectral de Voz, Tese de M.Sc., PEECOPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2000.

[35] Kurose, J. F. R., Keith, W.; Rede de Computadores e a Internet: uma nova abordagem; Tradução Arlete Simille Marques; 1. ed. - São Paulo: Addison Wesley, 2003.

[36] Alves, V., "Telefonia IP", Mestrado em Redes e Serviços de Comunicação Sistema Multimídia. Universidade do Porto - Faculdade de Engenharia.2002.

[37] Costa, S., "Qualidade de Serviço em Transmissão de Voz via Internet: QoS em VoIP na rede de dados da UFAM", Programa Institucional de Bolsas de Iniciação Científica(PIBIC). 2003.

[38] Costa, J. C. A., "Implementação e Gerência de uma Arquitetura de Voz sobre IP", Dissertação de mestrado, UFRJ, Rio de Janeiro, 2003.

[39] ITU-T Recommendation P.562, "Analysis and interpretation of INMD voice-service measurements", 2000.

[40] Florivaldo, M. F.,"A acustica musical em palavras e sons"., Editora Ateliê, 2003.

[41] ITU-T Recommendation P.561, "In-service non-intrusive measurement device - Voice service measurements", 2002.

[42] Fant, G., "Acoustic Theory of Speech Production". The Hague: Mouton and Co. 1960.

[43] Werner, H., "Quality of Service in IP Telephony: An End to End Perspective", Tese (Mestrado),Department Of Signaler Och System, Chalmers University Of Technology, Göteborg, 2004.

[44] PSYTECHNICS, "CCI: Getting the Message Loud and Clear, Measuring the Clarity os Speech Over Networks", White Paper, disponível em www.psytechnics.com/site/sections/resources/whitepapers.php, 2008.

[45] Yang, W., "Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality"Measure Based On Audible Distortion And Cognition Model. PhD Dissertation. Temple University, Philadelphia, USA, 1999.

[46] Dai, R., "A Technical White Paper on Sage's PSQM test", Sage Instruments Inc, 2000.

[47] Atkinson, D. J. ,"Proposed Annex A to Recommendation P.861", ITU-T Study Group 12 Contribution 24 (COM 12-24-E), International Telecommunication Union, Geneva, Switzerland, December 1997.

[48] ITU-T Recommendation P.563: "Single Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications", 2004.

[49] Wittmann, A., "Assuring VoIP Quality: Not There Yet", Network Magazine, 2005.

[50] Morrissey, P., "How To Measure Call Quality", Network Computing, 2005.

[51] TIPHON 4, ETSI, "Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 4, End-to-end Quality of Service in TIPHON Systems, Part 2: Definition of Speech Quality of Service (QoS) Classes", 2003.

[52] Martins, J., "Qualidade de Serviço (QoS) em Redes IP: Princípios Básicos, Parâmetros e Mecanismos", Apostila do Prof. Dr. Joberto Martins utilizada no curso de Pós-graduação em Tecnolgia da Informação da UNISANTA - Santos - SP, 1999.

[53] ITU-T Recommendation Y.1540, "Internet protocol data communication service IP packet transfer and availability performance parameters", 2002.

[54] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning", 2000.

[55] ITU-T Recommendation G.108, "Application of the E-model: A planning guide", 2003.

[56] Silva, J.G.M., "Aplicações VoIP Utilizando o Teleporto da Rede Metropolitana da Prefeitura Municipal de Manaus", Dissertação de Mestrado, Universidade Federal de Pernambuco, Recife, 2004.

[57] ITU-T Recommendation G.114, "One-way transmission time", 2003.

[58] Nascimento, M. S., "Medidas de Qualidade de Voz Em Redes IP", Dissertação de Mestrado, Universidade Federal do Paraná-UFPR, Curitiba, 2006.

[59] Jori, L., "Voice over IP in networked virtual environments", Thesis of Master Degree, School voor Kennistechnologie, Dutch, 2007.

[60] Al-akaidi,P, Urwin, P., Blackledge, J., "Speech coders for mobile communications", Faculty of Computing Sciences Engineering, De Montfort University, Leicester, UK, 2005.

[61] ITU-T Recommendation P.564, "Conformance testing for narrowband voice over IP transmission quality assessment models", 2006.

[62] Alencar, M. S., "Telefonia digital", 4° Edição, São Paulo, 2002.

[63] Barradas, O. C. M., "Telecomunicações, sistemas analógicos-digitais", Editora LTC , 1980.

[64] Deller, J. R., Proakis, J.G., Hansem, J.H.L., "Discrete-Time Processing of Speech Signals", MacMillan Coll Div, 1995.

[65] Fletcher, H, Munson, W.A., "Loudness, Its Definition, Measurement and Calculation", The Journal of the Acoustical Society of America, 1933.

[66] ITU-T Recommendation G.113, "Transmission Impairments - Provisional planning values for the equipment impairment factor Ie", 1998.

[67] Kuwabara, E.Y.T., Quarto, R.A.R, Santos, V.F., "Garantia de QOS para aplicações VoIP", Monografia de graduação em Engenharia de Telecomunicações, UFF, 2008.

[68] Finamore, W. A., Princípios de comunicações - Representação digital de sinais analógicos, Notas de Aula, PUC-RJ - 2003.

[69] Lamas, R.M.L.S.,"Avaliação de Codificadores de Voz em Ambiente VoIP", Dissertação de mestrado, UFRJ, 2005.

[70] Tanenbaum, A.S.; "Redes de computadores", 4° Edição, Prentice Hall, 2003.

[71] Rix, A.W., "Comparison between subjetive listening quality and P.862 PESQ score", Psytechnics white paper, United Kingdom, 2003.

[72] Request for Comments 3611, T. Friedman, R. Caceres, A. Clark, " RTP Control Protocol Extended Reports (RTCP XR)", 2003.

[73] Vianna, B.A., "Proposta e análise de um algoritmo adaptativo de ajuste de taxa de transmissão para sistemas VoIP", Dissertação de Mestrado, UFF, 2007.

[74] David, F., "Ferramentas de Monitoração Ativa e Passiva para Avaliação da Qualidade de Redes VoIP", Dissertação de Mestrado, NCE-UFRJ, 2003.

[75] TIPHON 23 (Telecommunications and Internet Protocol Harmonization Over Networks) - ETSI, "Comparison of TS101 329-5 Annex E with E Model", Sophia Antipolis, 2001.

APÊNDICE A

CONFIGURAÇÃO UTILIZADA NOS ROTEADORES NOS TESTES COM A TECNOLOGIA VoIP.

```
Current configuration : 3053 bytes
!
version 12.4
service timestamps debug datetime msec
service timestamps log datetime msec
no service password-encryption
!
hostname ROUTER_01
!
boot-start-marker
boot-end-marker
!
card type e1 0 2
!
no aaa new-model
no network-clock-participate wic 2
dot11 syslog
!
ip cef
!
no ip domain lookup
multilink bundle-name authenticated
!
voice-card 0
 no dspfarm
!
voice call send-alert
!
voice service voip
 sip
  bind control source-interface Multilink1
  bind media source-interface Multilink1
  rel1xx disable
  no call service stop
!
voice class codec 1
 codec preference 1 g723r63
```

```
 codec preference 2 g729br8
 codec preference 3 g711alaw
!
voice translation-rule 1
 rule 1 /^##/ //
!
voice translation-profile Remove#
 translate called 1
!
archive
 log config
  hidekeys

controller E1 0/2/0
 framing NO-CRC4
 channel-group 0 timeslots 1-31
!
interface Multilink1
 ip address 192.168.125.2 255.255.255.252
 ppp multilink
 ppp multilink group 1
!
interface GigabitEthernet0/0
 ip address 10.10.1.1 255.255.255.0
 duplex auto
 speed auto
!
interface GigabitEthernet0/1
 ip address 10.10.20.200 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/0/0
 no ip address
 shutdown
 clock rate 2000000
!
interface Serial0/0/1
 no ip address
 shutdown
 clock rate 2000000
!
interface Serial0/2/0:0
 no ip address
```

```
 encapsulation ppp
 ppp multilink
 ppp multilink group 1
!
ip forward-protocol nd
ip route 0.0.0.0 0.0.0.0 Multilink1
!
ip http server
no ip http secure-server
!
control-plane
!
voice-port 0/1/0
 cptone AR
 timeouts initial 15
 timeouts call-disconnect 3
 timeouts ringing 90
 timing hookflash-in 1000 50
 station-id number 2121011594
 caller-id enable
!
voice-port 0/1/1
!
dial-peer voice 10 pots
 preference 2
 destination-pattern 2121011594
 progress_ind alert strip 8
 port 0/1/0
!
dial-peer voice 30 voip
 destination-pattern ...T
 voice-class codec 1
 session protocol sipv2
 session target sip-server
 session transport udp
 incoming called-number .
 dtmf-relay rtp-nte
 fax protocol t38 ls-redundancy 0 hs-redundancy 0 fallback none
 ip qos dscp cs3 signaling
!
dial-peer voice 40 voip
 translation-profile outgoing Remove#
 max-conn 1
 destination-pattern ##T
```

```
 modem passthrough nse codec g711alaw
 session protocol sipv2
 session target sip-server
 session transport udp
 dtmf-relay rtp-nte
 playout-delay nominal 200
 playout-delay mode fixed no-timestamps
 codec g711alaw
 ip qos dscp cs3 signaling
 no vad
!
dial-peer voice 50 voip
 preference 1
 destination-pattern 2121011594
 voice-class codec 1
 session protocol sipv2
 session target sip-server
 session transport udp
 dtmf-relay rtp-nte
 fax protocol t38 ls-redundancy 0 hs-redundancy 0 fallback none
 ip qos dscp cs3 signaling
!
dial-peer terminator F
sip-ua
 retry invite 3
 retry register 3
 registrar ipv4:172.18.100.110 expires 3600
 sip-server ipv4:172.18.100.110
!
line con 0
 logging synchronous
line aux 0
line vty 0 4
 login
!
scheduler allocate 20000 1000
!
end
```

1. ESCOPO

Definir os passos necessários para a utilização do equipamento PERFORMER, do fabricante RADCOM, destinado à medição de parâmetros de qualidade de voz em redes de telecomunicações.

2. DOCUMENTOS DE REFERÊNCIA

Performer User Guide version 7.50
QPro User Guide version 7.00
SIP Simulator Family version 7.50
323 Sim User Guide version 6.69
Capture User Guide version 7.50

3. CONDIÇÕES AMBIENTAIS

"*NÃO APLICÁVEL"*

4. INSUMOS

"*NÃO APLICÁVEL"*

5. PREPARAÇÃO E CUIDADOS

"*NÃO APLICÁVEL"*

6. PROCEDIMENTOS

6.1 IDENTIFICAÇÃO DAS CARACTERÍSTICAS DO EQUIPAMENTO

Figura 1 – Visão geral do PERFORMER

6.1.1 INTERFACES DE REDE

MANAG – Interface Ethernet 10/100 padrão para conexão a uma rede IP.
SIP SIM - Interface Ethernet 10/100 utilizada quando a funcionalidade de simulação SIP for ativada, fica posicionada na parte frontal do equipamento.

6.1.2 INTERFACES PARA CONEXÕES DE PERIFÉRICOS

VÍDEO – DB-09
TECLADO/MOUSE - DIM
USB(duas interfaces)

6.1.3 BOTÕES LIGA DESLIGA

Na parte frontal, do lado direito, existem 03 botões para esta finalidade:

Power Switch – Liga e desliga a alimentação do equipamento.
Reset – Reinicia o sistema operacional do equipamento.
Software Shutdown – Desliga todos os *softwares* do equipamento. Este botão é ideal para desligar de forma rápida o equipamento e não perder tempo, fechando os vários programas que estão abertos.

6.1.4 VERSÃO DO SOFTWARE

RADCOM PERFORMER 7.50.12

6.1.5 – INTERFACES DE TESTE

As interfaces de teste estão localizas nos LIM (Line Interface Module). Cada LIM é apresentado no software como uma FEP (Front End Processor), tecnologia do fabricante no qual o processamento das funcionalidades de cada LIM ocorre quase que exclusivamente na própria placa e utiliza poucos recursos da máquina. Cada FEP do Performer utiliza o chip GEAR, trata-se de um chip ASIC totalmente customizado e desenhado pela RADCOM. O equipamento da EMBRATEL possui os seguintes LIMs :

6.1.5.1 LIM ANALÓGICO :

Este módulo possui 08 portas analógicas agrupadas duas a duas no padrão RJ-14. Para utilizá-las, é imprescindível o conector ‘SPLITTER” que vem como acessório.

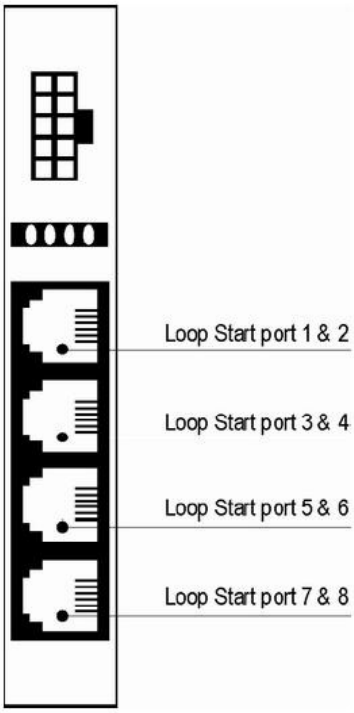


Figura 2 – LIM analógico

Conecta-se então o lado RJ-14 do conector ‘SPLITTER” na placa de interfaces analógicas e nas duas portas de saída do referido conector, faz-se a conexão com às linhas telefônicas que serão utilizadas no teste.

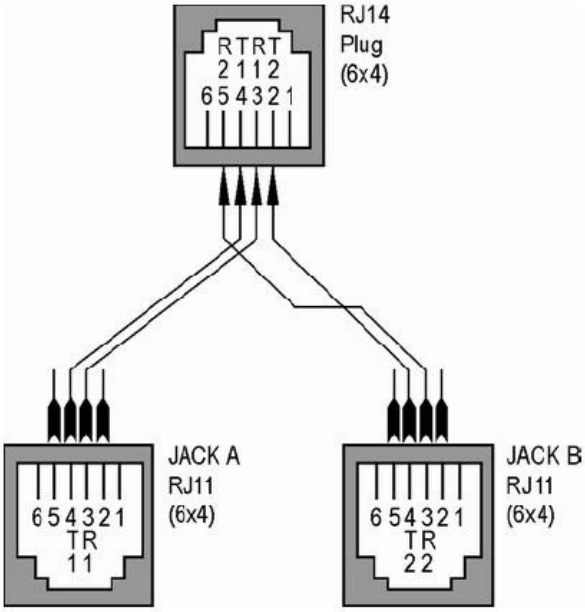


Figura 3 – Conector “SPLITTER”

6.1.6 LIM E1/T1

Possui 04 interfaces que podem ser configuradas como E1 ou T1.

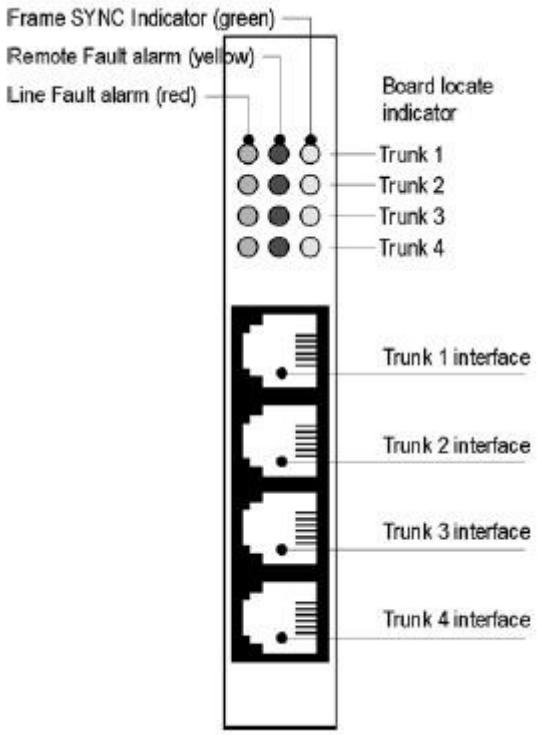


Figura 4 – LIM E1/T1

Line Fault Alarm – O alarme de falha de linha aparece em vermelho e indica uma perda de quadro(frame), perda de sinal ou taxa de bit de erro.

Remote Fault Alarm – O alarme de falha remota aparece em amarelo e indica uma perda remota quadros ou perda remota de sinalização multi-quadro.

Frame SYNC Indicator – O indicador de sincronismo de quadros aparece em verde e indica o correto sincronismo para o tronco, que significa, alinhamento perfeito de todos os quadros. Este LED quando está apagado pode significar as seguintes condições:

- Todos os outros alarmes.
- Perda de quadro.
- Perda de sinalização multi-quadro.
- Erros de CRC.

6.1.7 LIM ETHERNET

Possui 04 portas ethernet 10/100. As quatro interfaces são configuradas duas a duas, sendo que 02 podem operar de forma bidirecional e as 04 operam de forma unidirecional.

Figura 5 – LIM ETHERNET

As portas ethernet possuem os seguintes modos de funcionamento:

Switch mirror port – Utilizado quando operando a aplicação "CAPTURE" e o equipamento está conectado a uma porta que foi configurada como espelhamento de outra.
Pass Through - Utilizado quando operando a aplicação "CAPTURE" e o equipamento está conectado em série com os dispositivos geradores e receptores dos pacotes.
Traffic Generator – Utilizado quando operando a aplicação "TRAFFIC GENERATOR".
MediaStress - Utilizado quando operando a aplicação "MEDIA STRESS AS".

Todas as portas utilizam o conector Tipo RJ-45 UTP e suportam os modos de transmissão Half duplex, Full duplex e Auto-negotiation.

6.2 FORMAS DE CONEXÃO AO EQUIPAMENTO

6.2.1 Conexão direta

Neste caso, o equipamento não necessita estar conectado a uma rede IP, conectam-se os periféricos (monitor, teclado e mouse) diretamente ao dispositivo. Para iniciar o software acessa-se a aplicação "Performer Console".

6.2.2 Remotamente através da funcionalidade terminal server

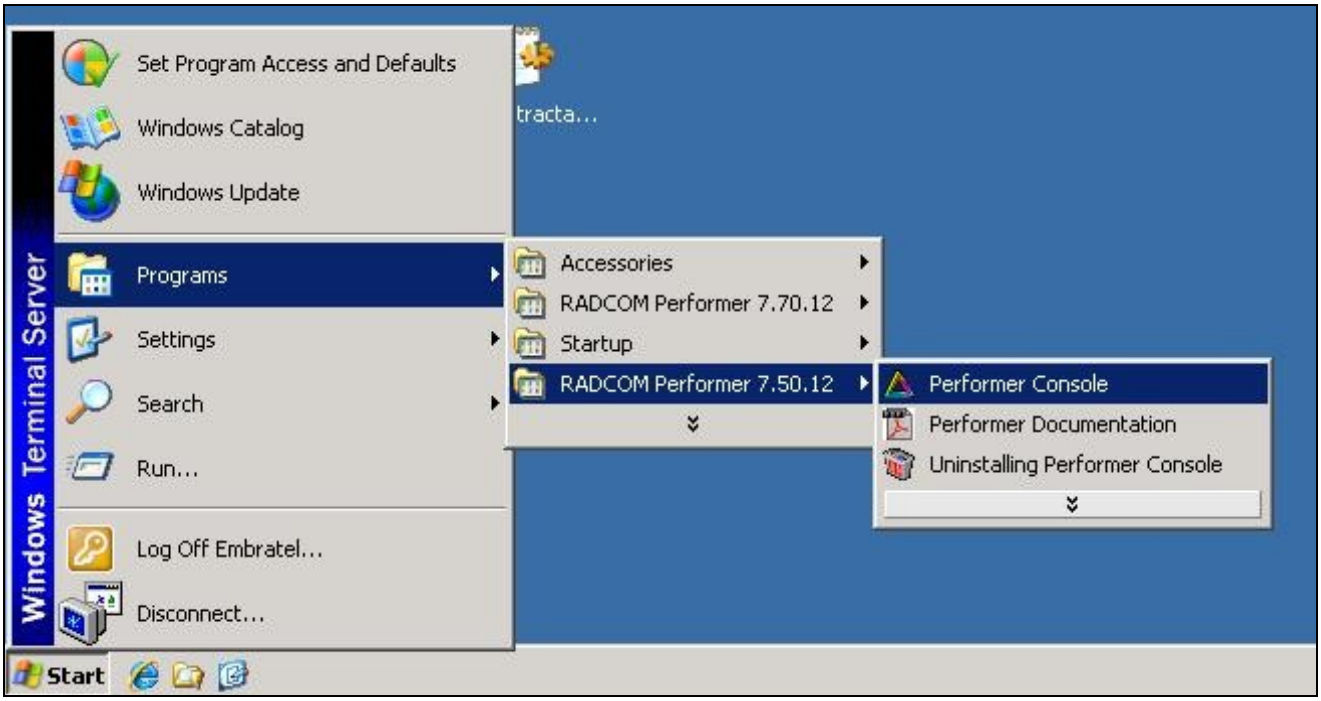
Qualquer computador conectado a uma rede IP utilizando a função de conexão remota do WINDOWS pode acessar a máquina. Para iniciar o software acessa-se a aplicação “Performer Console”.

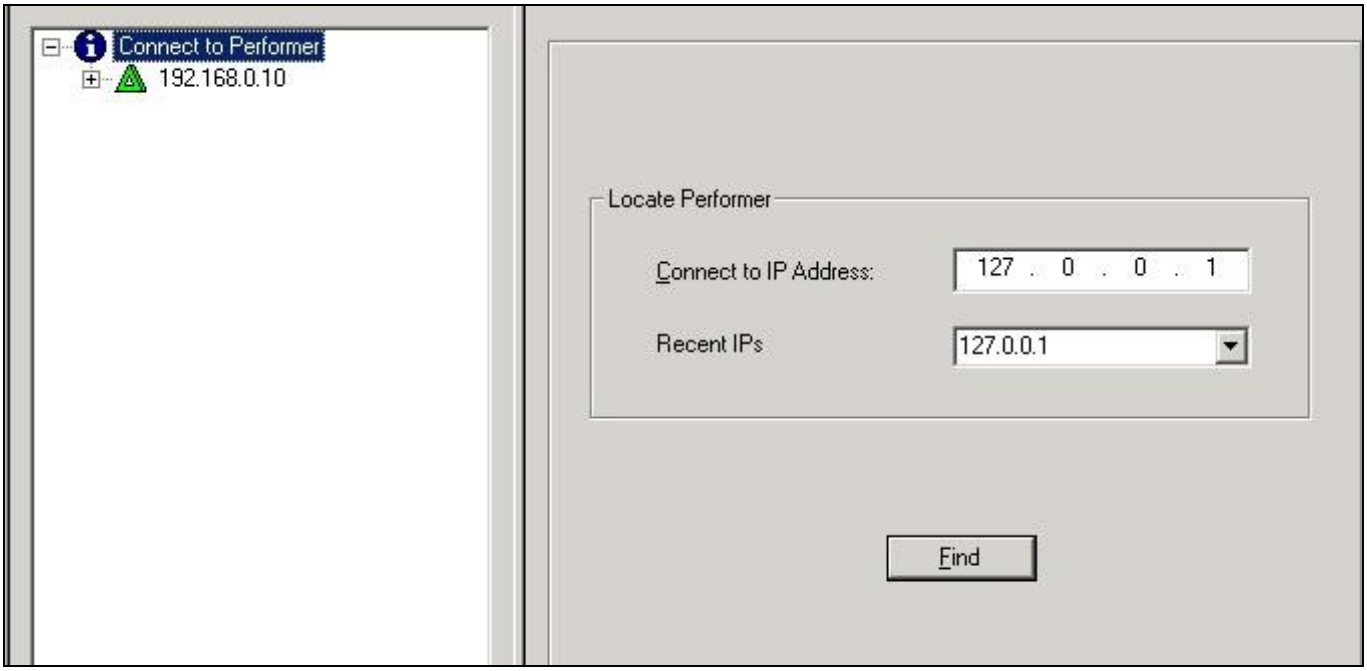
6.2.3 Remotamente através do software Performer Console

Utiliza-se um notebook, conectado a uma rede IP, com o software **Performer Console** instalado.

6.3 OPERAÇÃO DO SOFTWARE

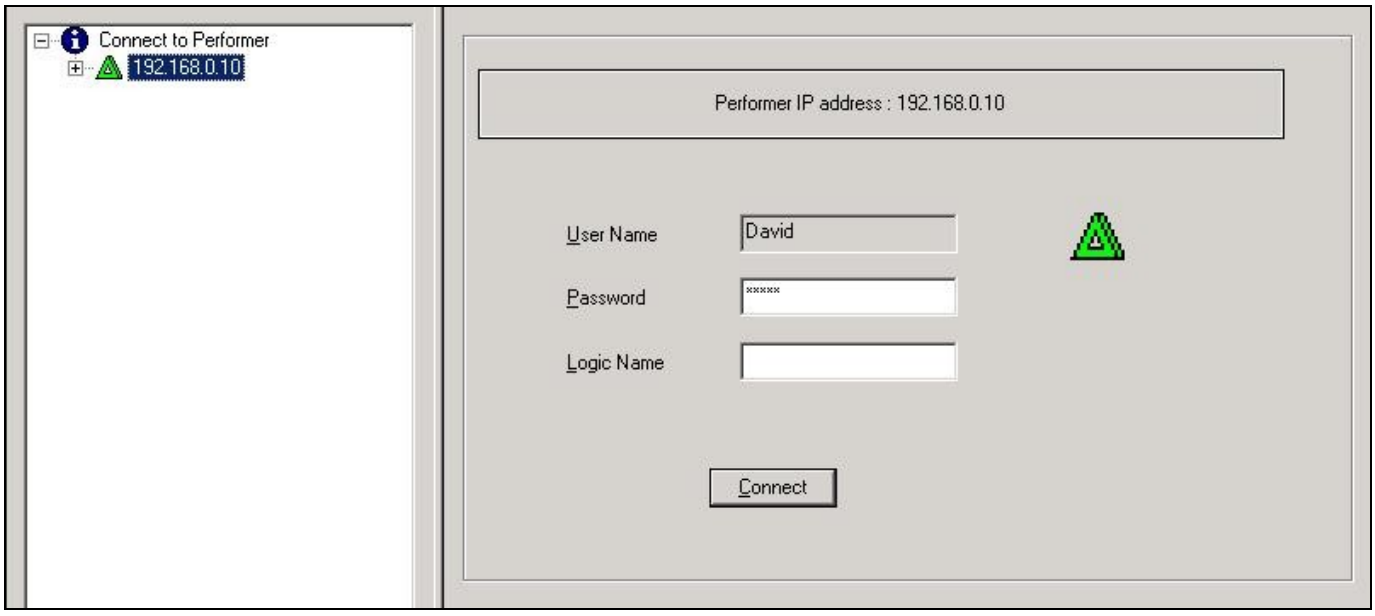
6.3.1 Conexão ao sistema





Selecione o endereço IP do Performer em "Recent Ips"e click em Find. O triângulo ao lado IP do performer deve estar verde.

Click duas vezes no endereço do Performer ao lado do tringulo VERDE e depois em "Connect"



CORES	INFORMAÇÃO
	Performer está disponível, pronto para conexão.
	Performer conectado.
	Performer ocupado, algum usuário já está conectado. Coloque o cursor do mouse neste ícone e nome do usuário estará visível.
	O software do servidor não está de acordo com o software da console. Não é possível trabalhar no equipamento até que o programa seja atualizado.
	Performer desconectado.

6.3.2 – GUIAS DE TECNOLOGIAS

As guias ou abas na lateral esquerda da tela inicial apresentam as tecnologias que podem ser testadas ou medidas. São disponibilizadas as tecnolgias VOIP, UMTS, GPRS, CDMA e as funcionalidades Favorites, General e Channel APP. Nesta Instrução será abordada a tecnologia VOIP.

6.3.2.1 – TECNOLOGIA VOIP

Dentro de cada tecnologia o software apresenta em uma caixa vertical os ícones dos serviços disponíveis para teste:

MediaPro, MEGASIP, 323Sim, Qpro, Capture, Physical Layer, Analysis e MediaStress SA.

MediaPro : O MediaPro captura chamadas *on line* e transmissões de fax e as amostra (ambas, abertas ou fechadas) de forma acurada, medindo os aspectos de mídia e sinalização nos protocolos H.323,

MGCP and SIP, tornando possível visualizar informações sobre uma simples chamada ou canal ou visualizar detalhes de todas as chamadas como um todo. O MediaPro provê todos os aspectos de análise que são vitais para avaliação de alto tráfego de chamadas.
O MediaPro combina medidas objetivas e subjetivas de qualidade da voz. As medidas objetivas são baseadas nos métodos PESQ (Perceptual Evaluation of Speech Quality) e PAMS (Perceptual Analysis Measurement System). As medidas subjetivas envolvem as gravações e reproduções do próprio MediaPro.

MEGASIP : O MEGASIP é uma aplicação que apresenta funções de simulador de chamadas SIP. O MegaSIP inclui dois componentes: MegaSIP estresse de sinalização e MegaSIP estress de mídia. O MegaSIP estresse de mídia está disponível para 04 diferentes níveis com 150, 4.500, 20.000 ou 50.000 canais RTP de saída.

323Sim : O 323Sim é um gerador de chamadas capaz de emular as funcionalidades de um terminal H.323. Várias simulações podem ser implementadas, aumentando o volume de chamadas para milhares de chamadas simultâneas. Cada 323Sim pode gerar até 2.000 chamadas simultaneamente a uma taxa que pode chegar a 80.000 chamadas por hora.
As seguintes configurações do 323Sim estão atualmente disponíveis:
• **323Sim Starter**: até 75 chamadas.
• **323Sim Pro**: até 500 chamadas (100 chamadas RTP).
• **323Sim Platinum**: até 2000 chamadas (200 chamadas RTP).
O 323Sim conecta telefones VoIP, gateways e PBXs. Ele gera chamadas com ou sem gatekeeper, abrindo canais lógicos e transmitindo fluxos RTP (Realtime Transport Protocol). Ele inicia e responde chamadas e habilita desenvolvedores e provedores de serviço a estabelecer padrões, fazer testes de carga e verificar o funcionamento do protocolo em implementações de equipamentos VoIP.

Qpro : O Qpro é uma ferramenta objetiva de medida da qualidade da voz que provê a ultima solução na predição da qualidade da fala sobre redes de múltiplas tecnologias, na mesma escala dos testes subjetivos. Transforma o sinal da fala representado no tempo, na freqüência e na amplitude, e faz a comparação do sinal degradado com o original. O QPro é baseado no modelo da audição humana e pode predizer a qualidade da voz na rede. Pode também enviar e receber transmissões de fax habilitando usuários para estressar sua rede sobre a capacidade de atender chamadas de fax. O QPro testa muitos cenários de forma rápida e com baixo custo, proporcionando rápidas e acuradas medidas. Oferece, ainda, rápido retorno a mudanças no desenho de equipamentos e objetivamente escalona diferentes projetos de soluções.

Principais funcionalidades:
• Provê uma avaliação objetiva da qualidade da voz (na unidade MOS), usando os algoritmos PESQ ou PAMS.
• Provê aprimoramento de medidas para casos de degradação localizada da qualidade.
• Gera chamadas bidirecionais e transmissões de fax, medindo parâmetros de sinalização e voz .
• Mede o atraso fim-a-fim da rede.
• Prove medidas de perda de retorno e tempo de atraso do eco.
• Permite testar a integridade do DTMF nas chamadas.
• Mede o tempo de estabelecimento de chamadas.
• Implementa *stress* de chamadas para similar o ambiente da rede utilizando tráfico estatístico.
• Gera e mede chamadas para aparelhos de telefonia celular (no modo analógico).

Capture : A aplicação Capture é utilizada para capturar, filtrar, analisar e decodificar dados em uma grande variedade de redes, incluindo ATM, E1/T1/J1, IMA, e redes Ethernet 10/100/1000. A aplicação **Capture** facilita a captura passiva de dados, assim como o fluxo entre duas entidades de comunicação. Ele provê transparência, sem mudanças nos dados passivos. Os dados são capturados da linha sob teste, de acordo com o definido e customizado pelos filtros e armazenado na memória RAM da unidade Performer. Após o processo parar, existem vários filtros "offline" que permitem analisar o conteúdo capturado.

Physical Layer : O Physical Layer é uma aplicação usada para examinar todos os aspectos da camada física em uma rede LAN, Gigabit Ethernet, Ethernet 10/100/1000, ATM, POS, quadros em linhas E1/T1/J1, Multi-portas OC-3, Canais OC-3, DS3/E3 e ATM sobre linhas E1/T1/J1 .
A aplicação Physical Layer pode ser configurada para monitorar sua linha e reportar eventos significativos de acordo com níveis previamente definidos pelo usuário.
O processo calcula se cada nível foi ultrapassado durante um determinado período de tempo. Os resultados são apresentados de duas formas, pelo valor e por uma série códigos de cores em LEDs. Os ciclos e períodos de tempos podem ser definidos para avaliar problemas na linha. Os dados podem ser mostrados em segundos, um período definido de tempo e no tempo total do processo. Os dados coletados podem ser impressos e exportados para planilhas do EXCEL.

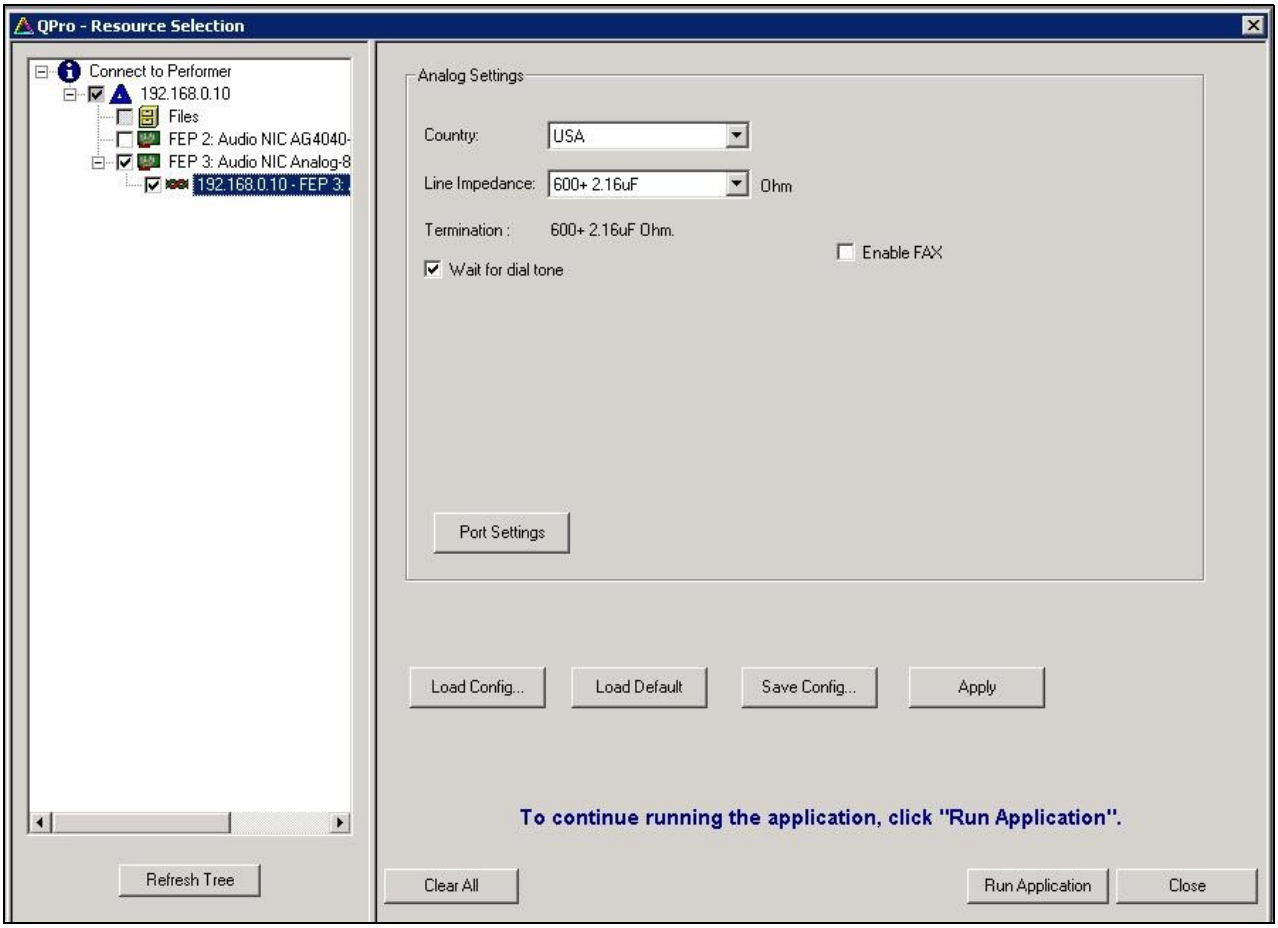
Nesta instrução será abordada a aplicação QPro que deve ser selecionada dentro da tecnologia VOIP.

6.3.3 – QPRO - SELEÇÃO DA INTERFACE DE TESTE

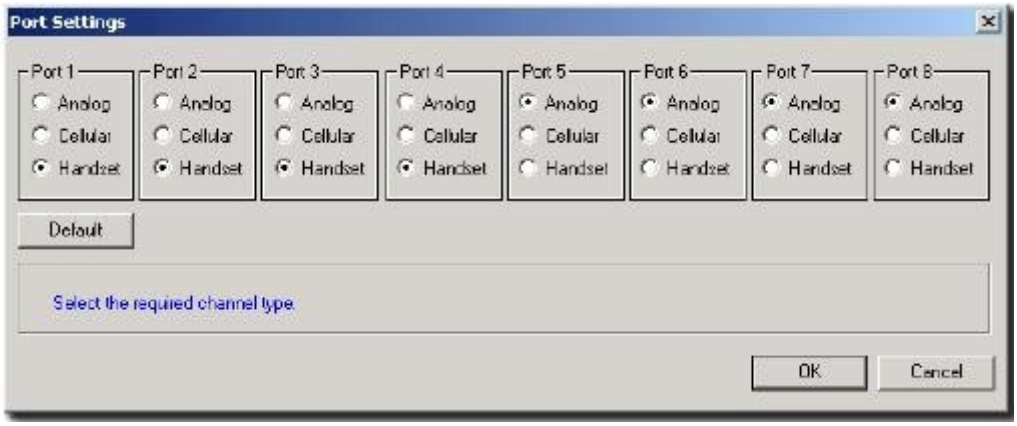
Para utilizar o QPro é necessário selecionar qual ou quais interfaces serão utilizadas no teste marcando a opção da FEP.

6.3.3.1 - FEP ANALÓGICA (Áudio NIC Analog 8)

Marque a caixa no lado esquerdo da descrição da FEP, caso seja necessário, altere os parâmetros na tela “Analog Settings”, em seguida click em “RUN APLICATION”

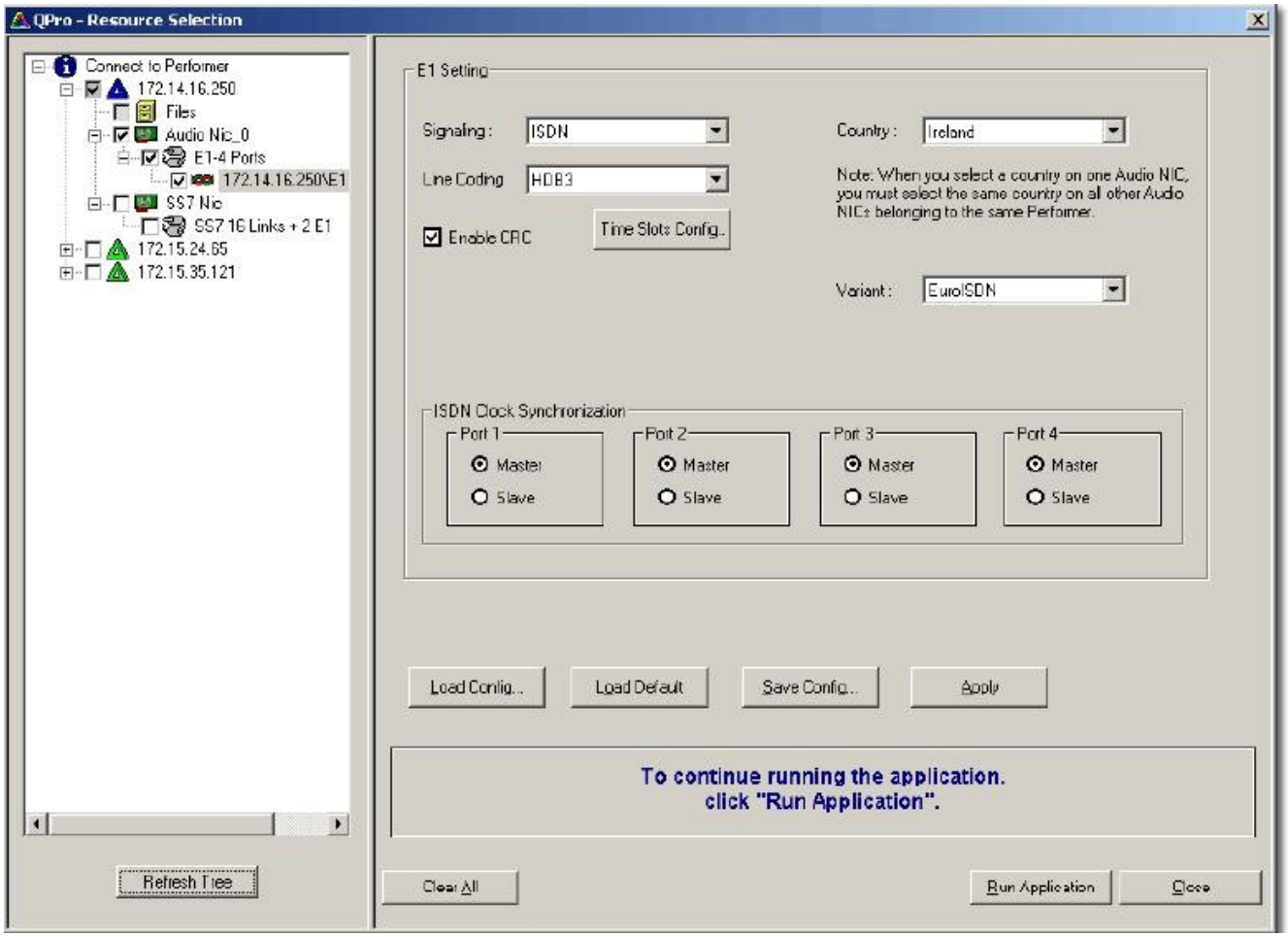


Clicando em “Port Settings” pode-se configurar os tipos dispositivos que serão conectados nas interfaces.



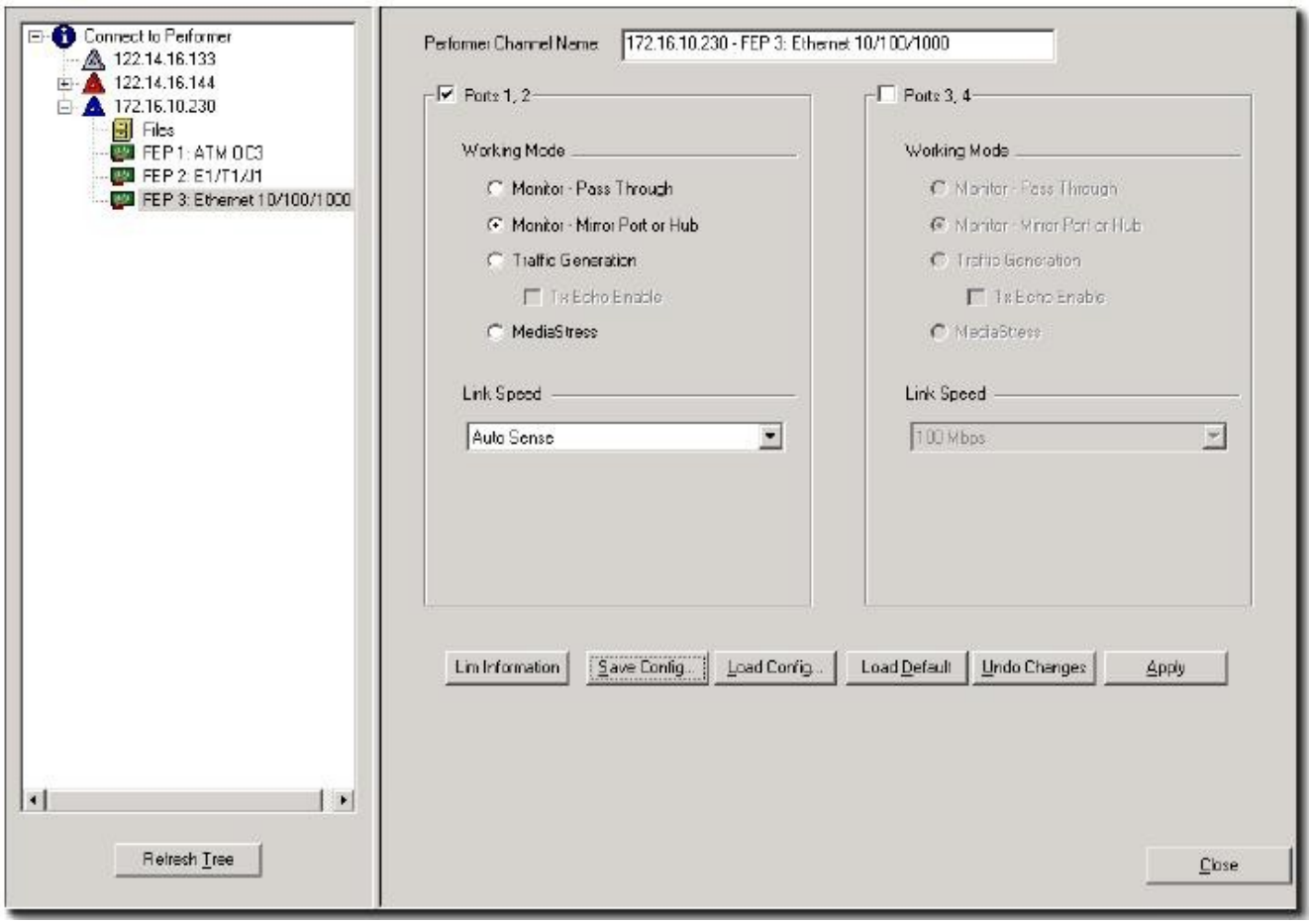
6.3.3.2 FEP E1/T1

Marque a caixa no lado esquerdo da descrição da FEP, caso seja necessário, altere os parâmetros na tela “E1 Settings”, em seguida click em “RUN APLICATION”



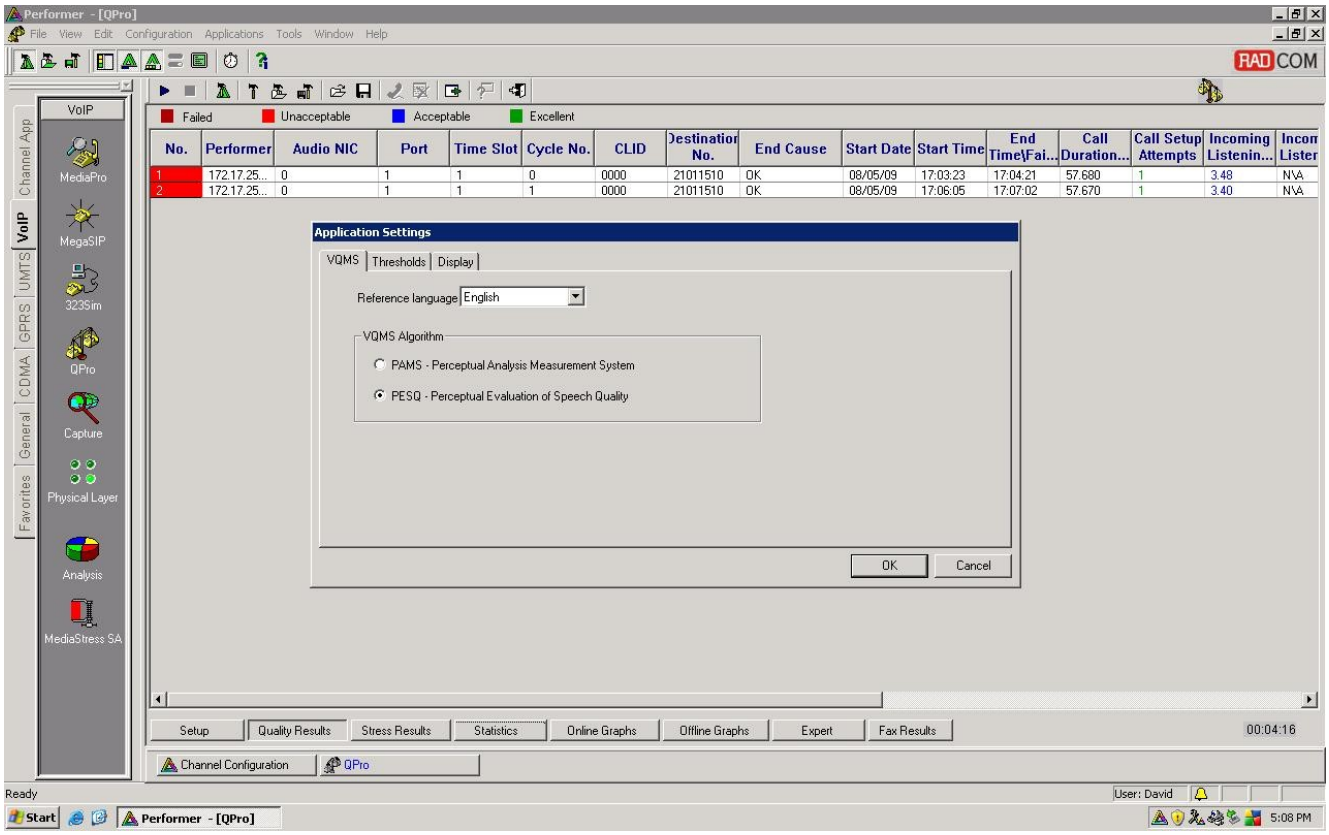
6.3.3.3 - FEP ETHERNET

Selecione a FEP correspondente, marque as duplas de interfaces que serão utilizadas, selecione o modo de trabalho e taxa da interface.



6.3.4 - SELEÇÃO DO MÉTODO DE MEDIÇÃO DA QUALIDADE E IDIOMA.

Para selecioná-los siga até a tela APLICATIONS SETTINGS. O ícone representativo desta tela tem o símbolo de um martelo.



6.3.4.1 TIPOS DE VOZES POR PAÍS

Na guia VQMS, em “Reference Language”, seleciona-se o idioma dos arquivos de voz que serão utilizados na medição. Por padrão, o idioma inglês vem selecionado.
No diretório D:\radcon\performer server\configure\qpro\qpro reference encontram-se os arquivos de voz por país .

No caso do idioma português, a fala segue o seguinte texto:
“ Escolinhas de esporte fazem o lazer dos fins de semana” (voz de mulher).
“ Vai interpretar um solteirão conservador” (voz de homem).
“ A diretoria técnica pretende ampliar o número de pontos de teste de propagação”(voz de homem).

Obs: os textos são diferentes por idioma.

6.3.4.2 MÉTODO DE MEDIÇÃO DA QUALIDADE

Na mesma tela encontramos “VQMS Algoritim”, onde seleciona-se o método que será utilizado na medição de qualidade da voz. Pode-se selecionar os métodos PESQ e PAMS. Por padrão, o método PESQ vem selecionado.

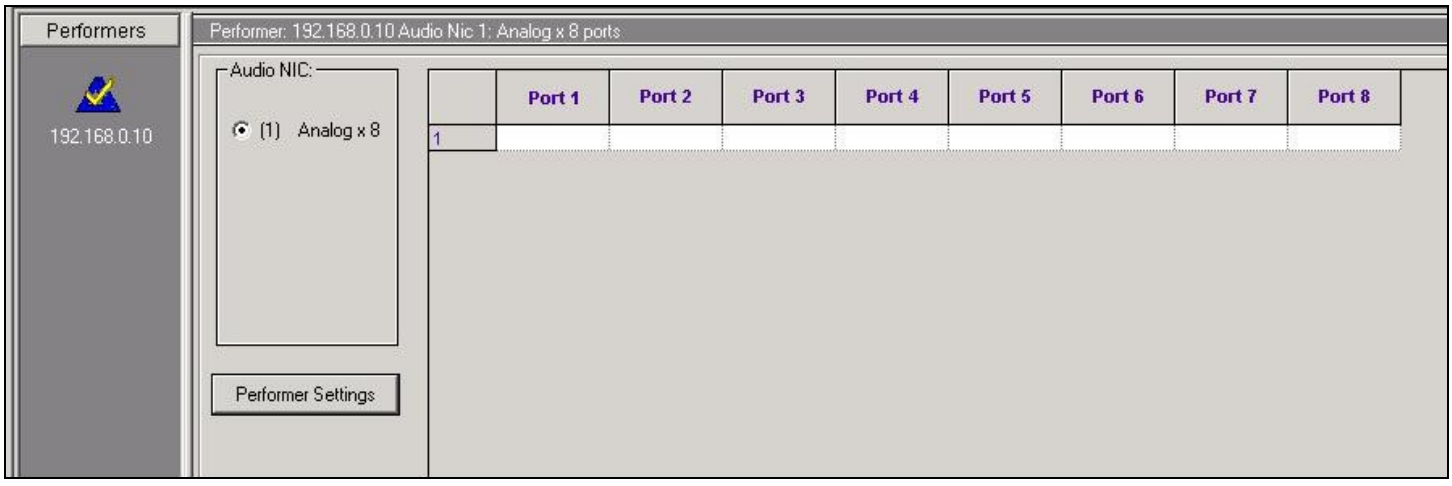
6.3.5 - SETUP - CONFIGURAÇÃO DAS CHAMADAS

Para configurar chamadas clica-se em  , onde pode-se estabelecer os números: chamado e chamador.

	Inicia as chamadas que foram configuradas.
	Para as chamadas.
	Retorna para a tela de configuração do canal.
	Permite que sejam definidos os ajustes da aplicação.
	Carrega arquivos de configuração.
	Salva arquivos de configuração.
	Abre arquivos.
	Salva arquivos.

6.3.5.1 SETUP DA FEP ANALÓGICA

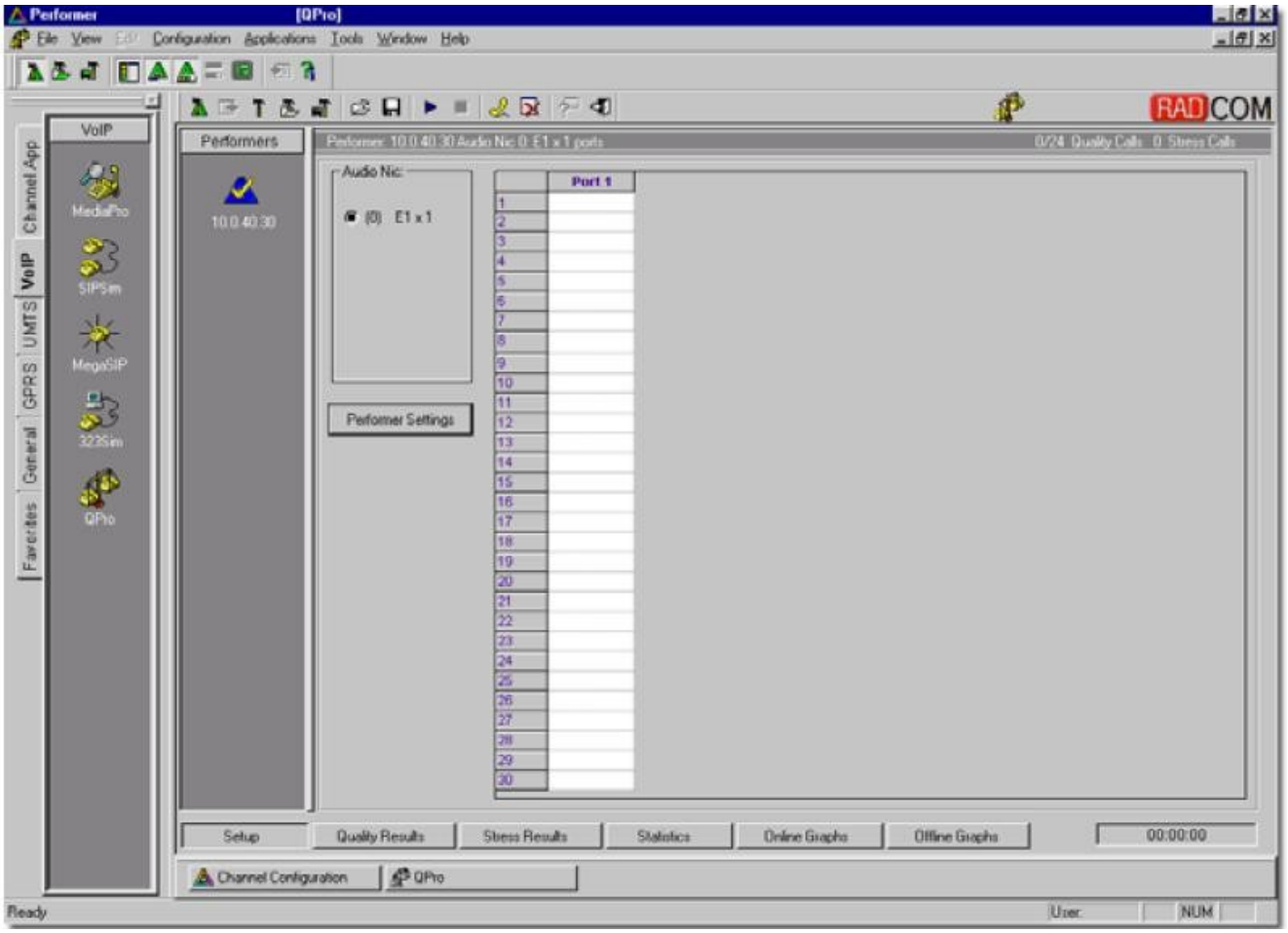
Selecione o ícone que possibilita a configuração do canal  e faça as configurações dos números das chamadas.



Após a aplicação ser posta em operação é apresentada à tela SETUP do QPRO, onde as linhas analógicas utilizadas são configuradas. Por exemplo, em "port 1" insere-se o número telefônico da linha que receberá a chamada. A linha que originará a chamada não precisa ser configurada. As portas são utilizadas duas a duas, ou seja, porta 1 - 2, 3 - 4, 5 - 6 e 7-8.

6.3.5.2 - SETUP DA FEP E1/T1

No caso de uma FEP E1, Selecione o ícone possibilita a configuração do canal  e faça as configurações dos números das chamadas em cada uma das 30 posições disponíveis.



Neste caso é apresentada a tela de uma FEP com somente uma E1. Em cada um dos 30 canais disponíveis podem ser configuradas chamadas independentes.

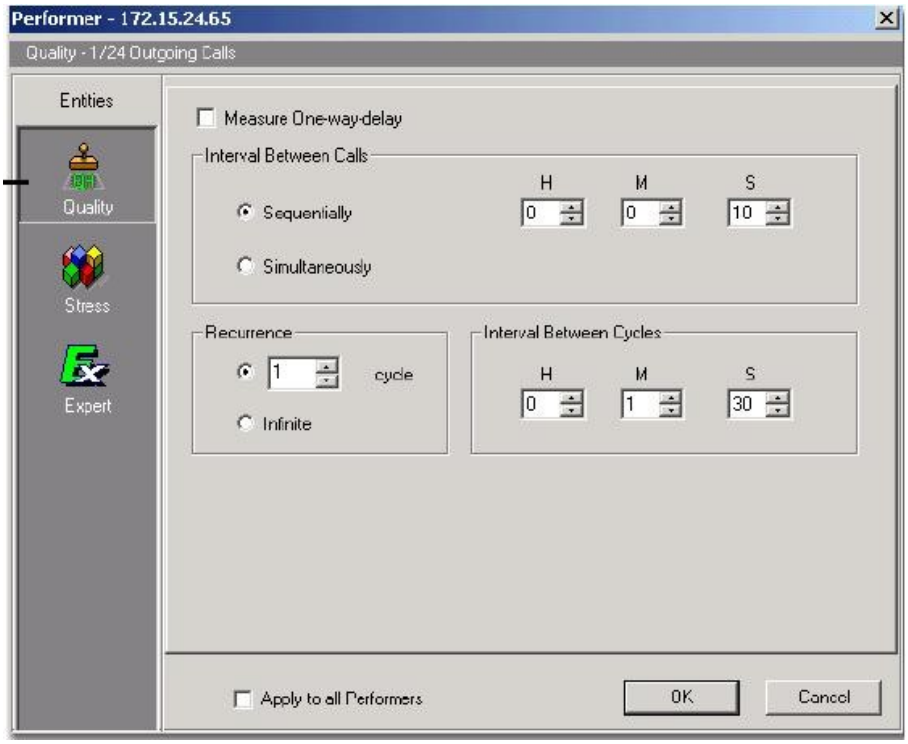
6.3.5.3 – PERFORMER SETINGS

Em ambos os casos, FEP analógica e E1, clicando em "Performer Settings", apresentam-se 03 opções: Quality, Stress e Expert.

6.3.5.3.1 – QUALITY

Configura-se a quantidade e periodicidade das chamadas:

- **Measure one-way delay** – Caso seja marcado mede somente o delay em um sentido.
- **Interval betweem calls** – Determina-se o tempo de intervalo entre chamadas.
Sequentially – Para chamadas sequenciais.
Simultaneously – Para chamadas simultâneas.
- **Recurrence** – Número de repetições das chamadas
- **Interval betweem cycles** – Tempo de intervalo entre os ciclos.

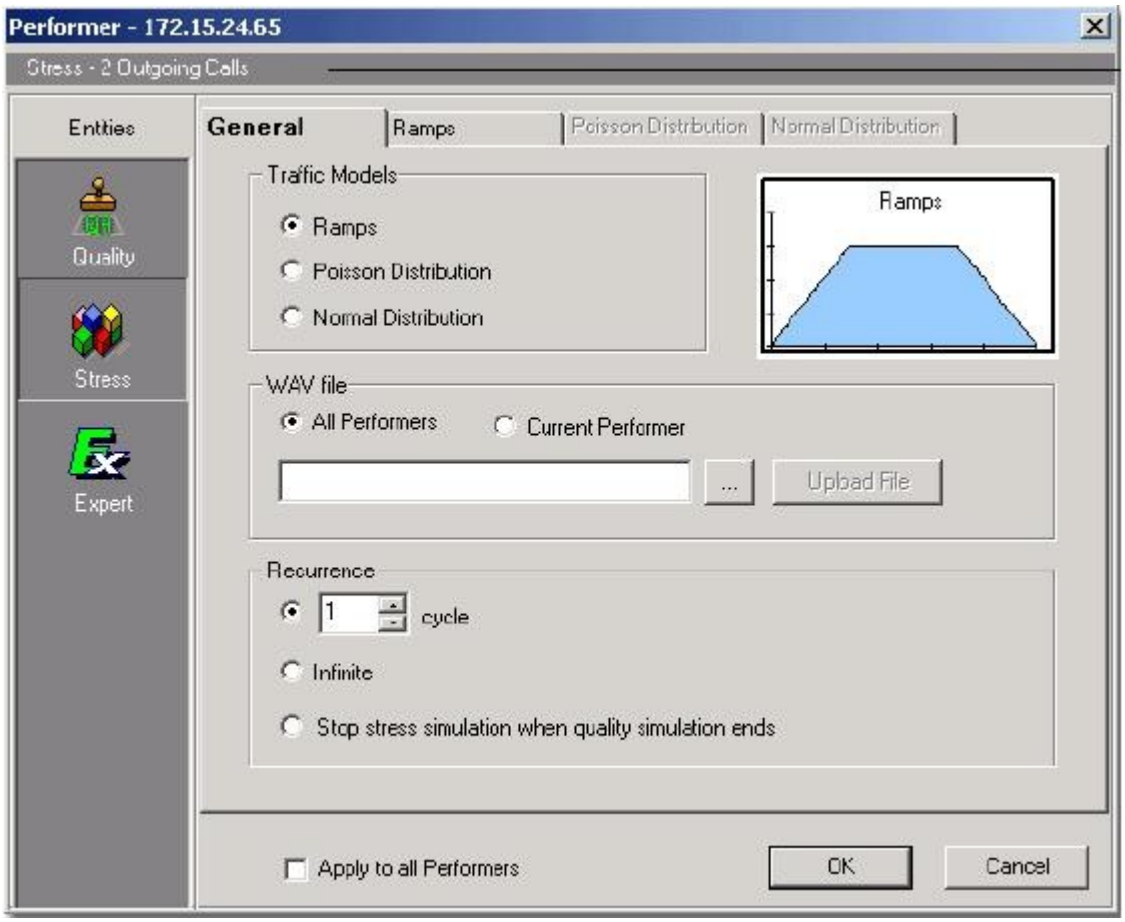


6.3.5.3.2 – STRESS

Funcionalidade que permite a geração de grande número de chamadas de diversas formas de sequenciamento e distribuição.

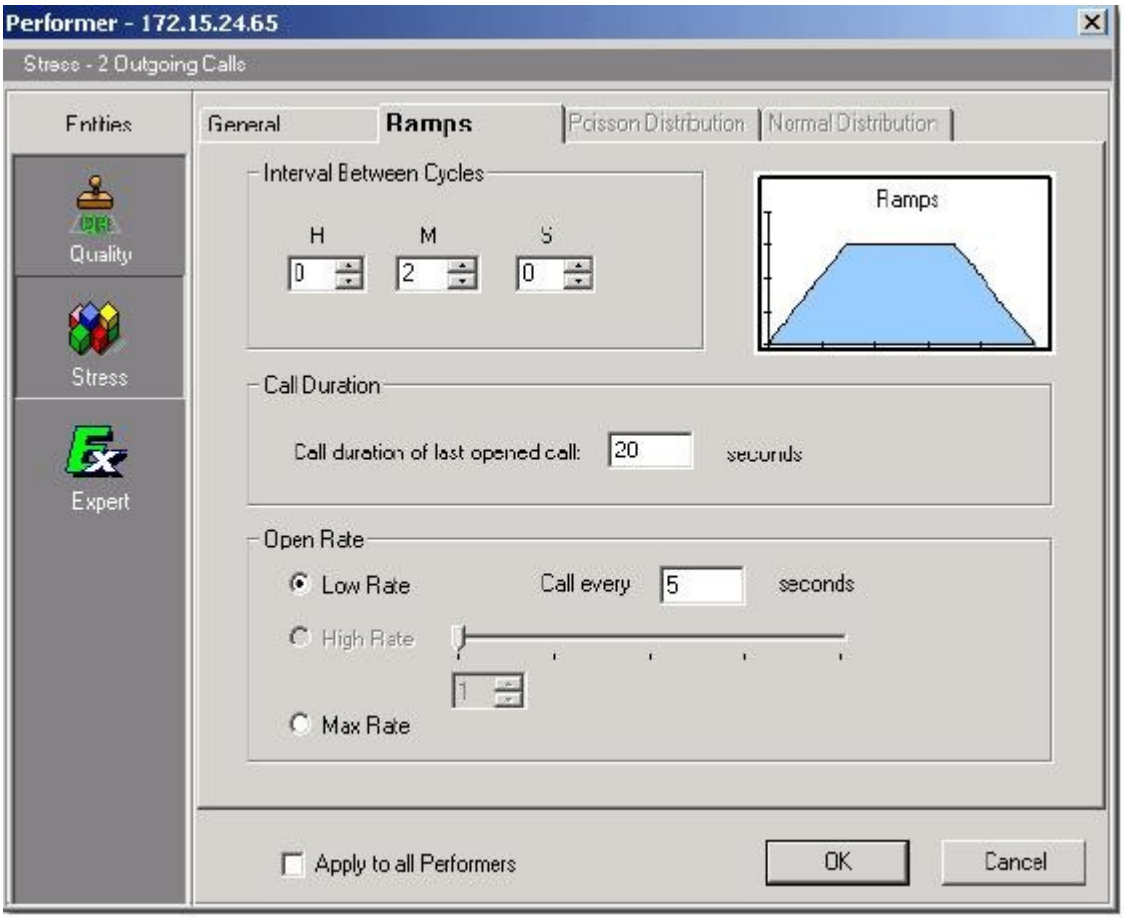
A guia "General" apresenta as seguintes opções:
Trafic Models – permite definir o modelo de distribuição de tráfego;
WAV file – Permite atualizar o arquivo de voz utilizado na simulação de Stress;
Recurrence – Com as possibilidades: Número de ciclos, Infinito e Para a simulação de Stress quando a simulação de qualidade termina.

Modelos de Distribuição de Tráfego – A escolha do modelo determinará a distribuição estatística da seqüência de chamadas que gerarão o "Stress" na rede.



6.3.5.3.2.1 – STRESS/RAMP

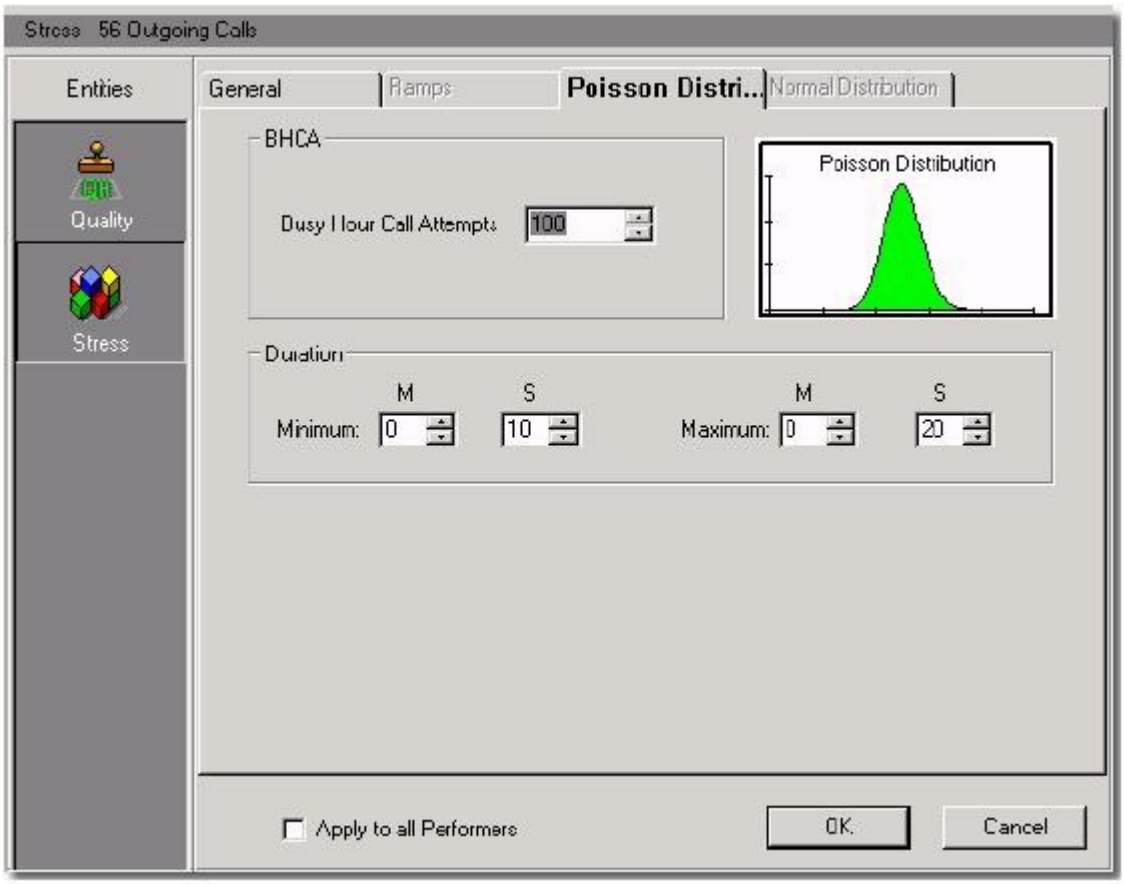
É o tipo de distribuição estatística padrão para o *performer*, ou seja, em uma seqüência de muitas chamadas contínuas o formato da distribuição é em forma de rampa. Na mesma tela pode-se também ajustar os parâmetros de periodicidade e duração das chamadas.



6.3.5.3.2.2 – STRESS / POISSON DISTRIBUTION

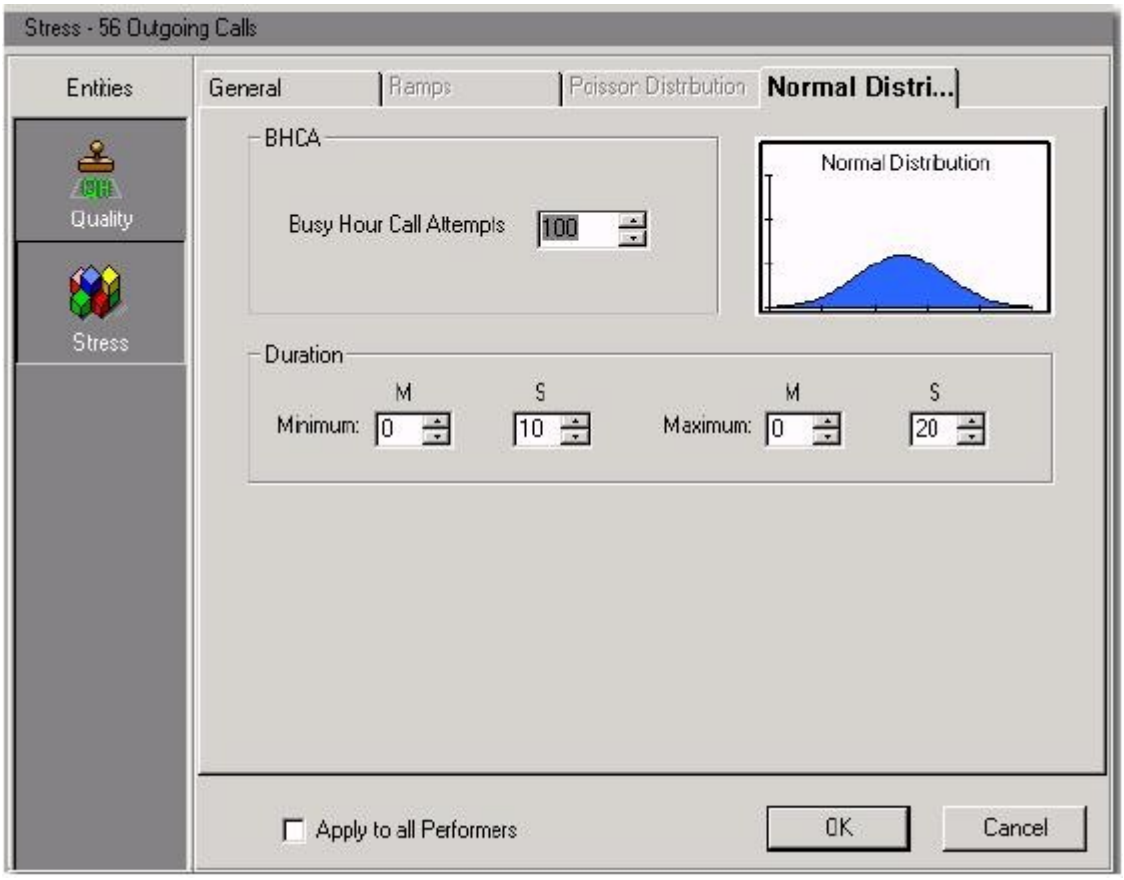
Utilizando este tipo de distribuição é necessário definir o BHCA. A carga de uma central telefônica é medida em BHCA (Busy Hour Call Attemps), ou seja, o número de tentativas de chamada na HMM (Hora de Maior Movimento). A Hora de Maior Movimento é o intervalo contínuo de 3600 segundos dentro de um período de 24 horas, onde se apresenta a maior intensidade de tráfego. Uma tentativa de chamada pode ser bem ou mal sucedida, mas nos 2 (dois) casos parte da capacidade de processamento é utilizada para realizar a tentativa.
Uma chamada bem sucedida dura em média de 120 a 180 segundos, já a mal sucedida dura alguns segundos de processamento. A média total é baixa e da ordem de 40 segundos.
A expressão que relaciona o tráfego na HMM com o BHCA é obtida da equação:

Tráfego na HMM = (BHCA / 3600) * Duração média de todas as chamadas (em segundos).



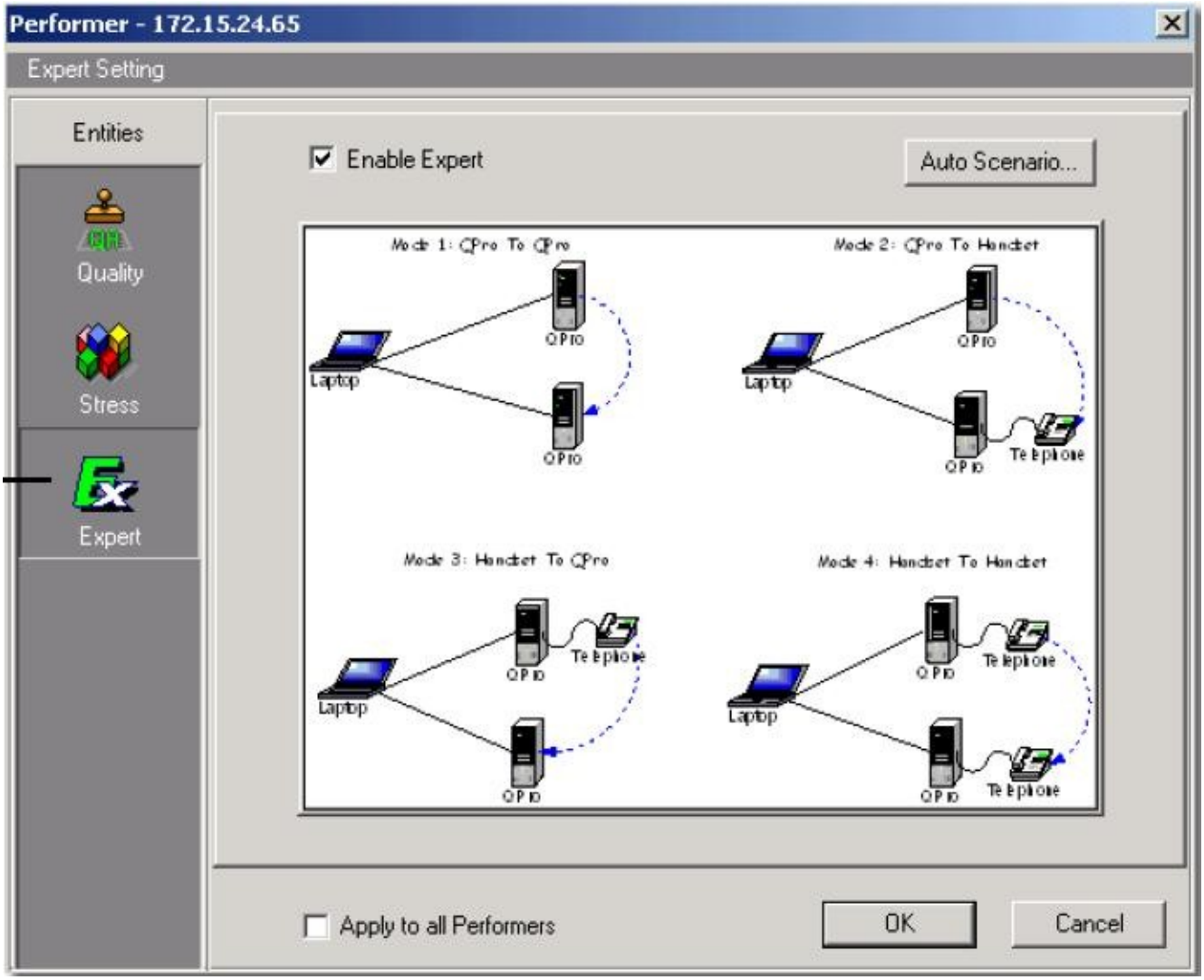
6.3.5.3.2.3– STRESS / NORMAL DISTRIBUTION

Neste tipo de distribuição também define-se o BHCA, ou seja a carga da central e têm-se as chamadas distribuídas de forma mais suave no intervalo de tempo.



6.3.5.3.3 - EXPERT

A funcionalidade “EXPERT” é uma ferramenta onde um ambiente de medição pode ser simulado em várias possibilidades. Os parâmetros a serem utilizados na simulação são configurados através da opção “ Auto Scenario”.

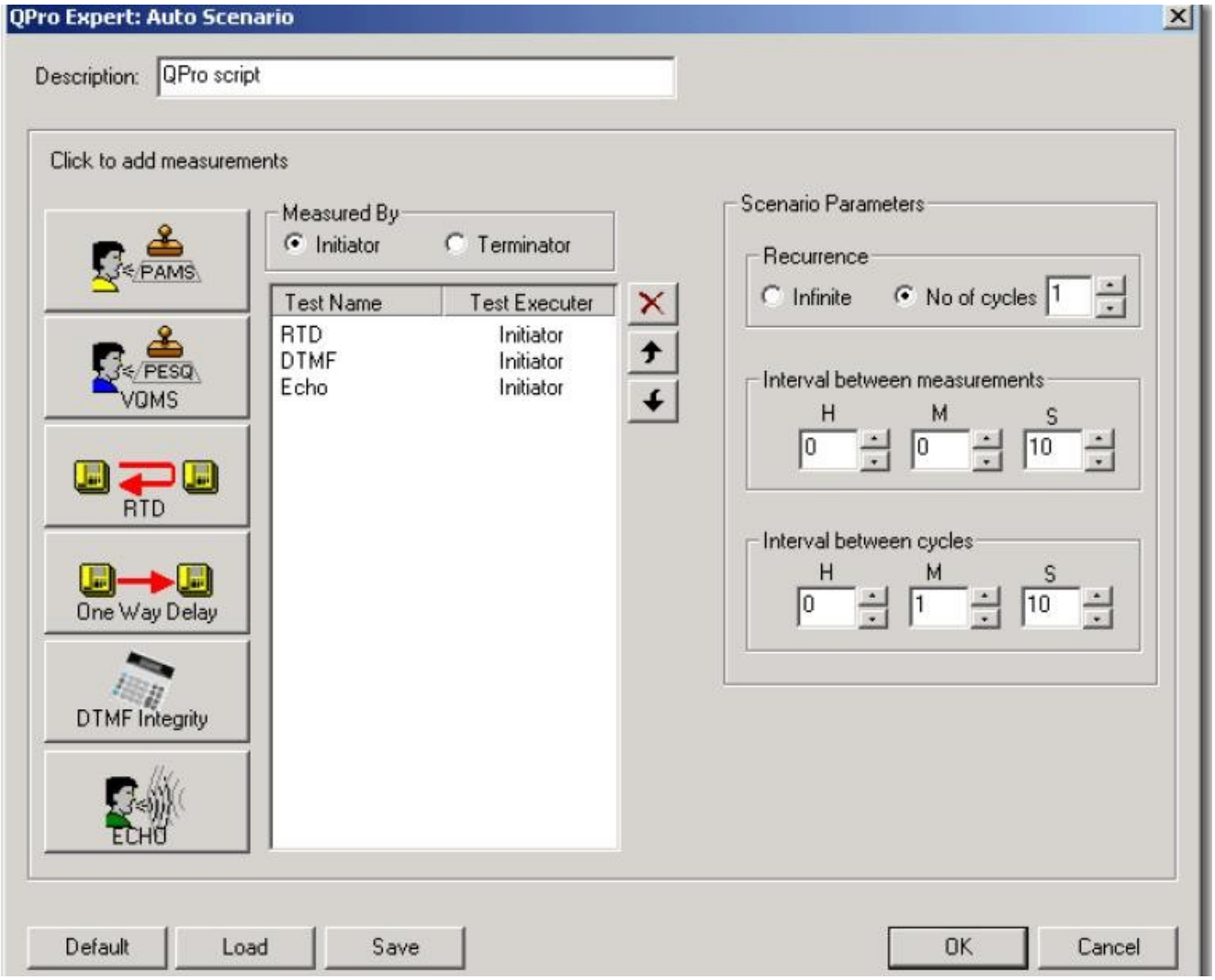


Por padrão a caixa “*Enable Expert”*está sempre selecionada.

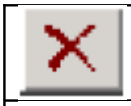
A caixa de seleção “*Apply to All Performers” quando marcada aplica as configurações para todos “PERFORMERS”presentes na rede.*

EXPERT AUTO CENÁRIO

Na opção "**Auto Scenario"** são definidos os parâmetros de um script QPro, que permite o funcionamento da simulação.



Nos botões PAMS, PESQ, round trip delay (RTD), one way delay, DTMF integrity ou echo, adiciona-se estes parâmetros ao script.
No campo "**Description"** é atribuído um nome para o script .
Na caixa de seleção **"Measured By"** Selecionam-se as formas de medição dos parâmetros:
• *Initiator*: O servidor Performer .
• *Terminator: O canal* selecionado.

	Deleta o parâmetro selecionado.
	Muda a ordem dos parâmetros no *script*, movendo o parâmetro selecionado para cima na lista.
	Muda a ordem dos parâmetros no *script*, movendo o parâmetro selecionado para baixo na lista.

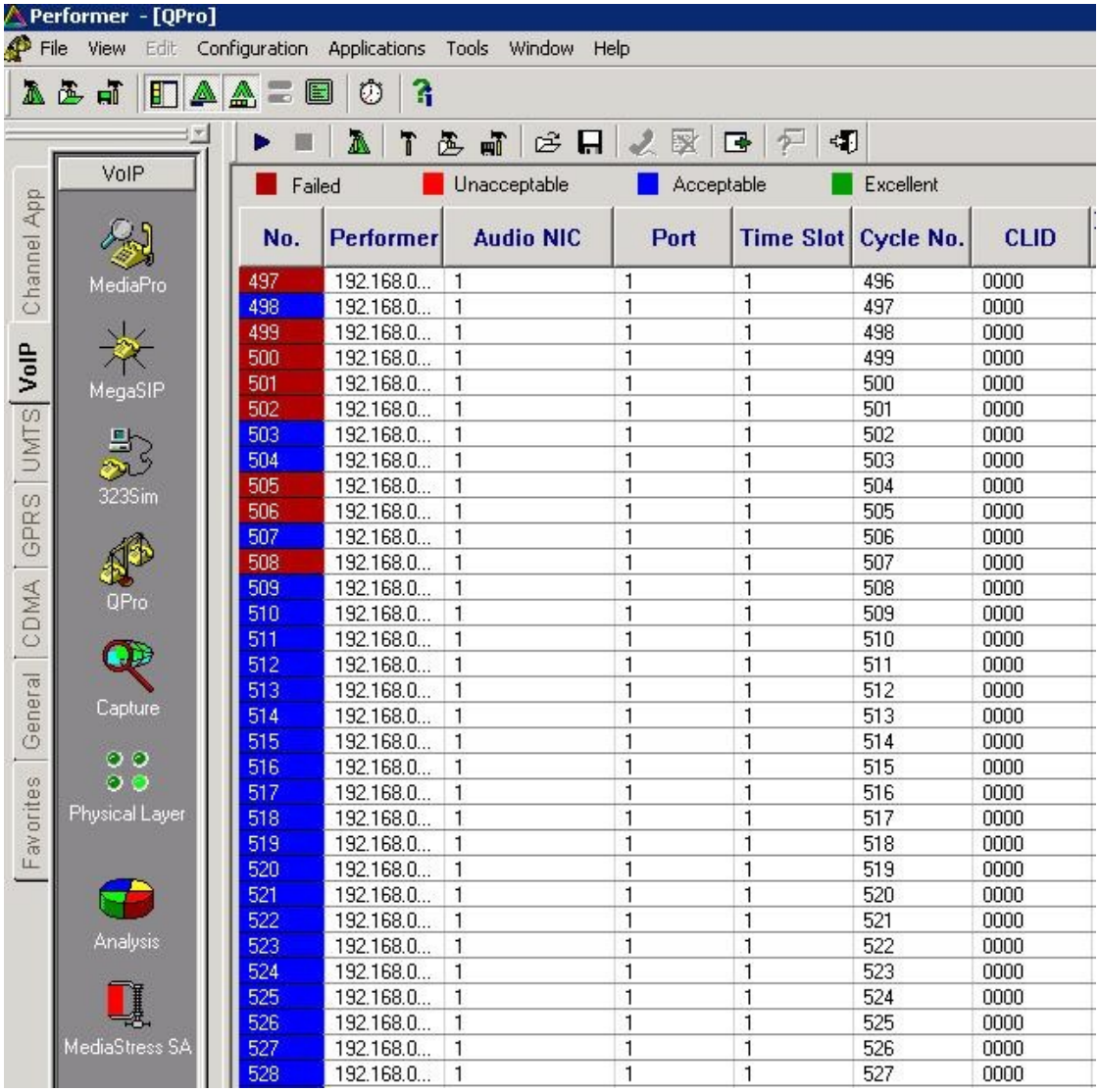
6.3.6 MEDIDAS DE QUALIDADE DAS CHAMADAS

A aplicação QPro tem as seguintes opções para visualização das medidas:

Tendo sido realizados os ajustes e configurações necessárias, inicia-se a aplicação Qpro clicando no botão iniciar.

6.3.6.1 QUALITY RESULTS

Nesta tela observam-se os valores dos valores medidos durante os testes. Cada linha da tabela representa uma chamada realizada. Nas colunas, têm-se os parâmetros de qualidade obtidos em cada chamada.



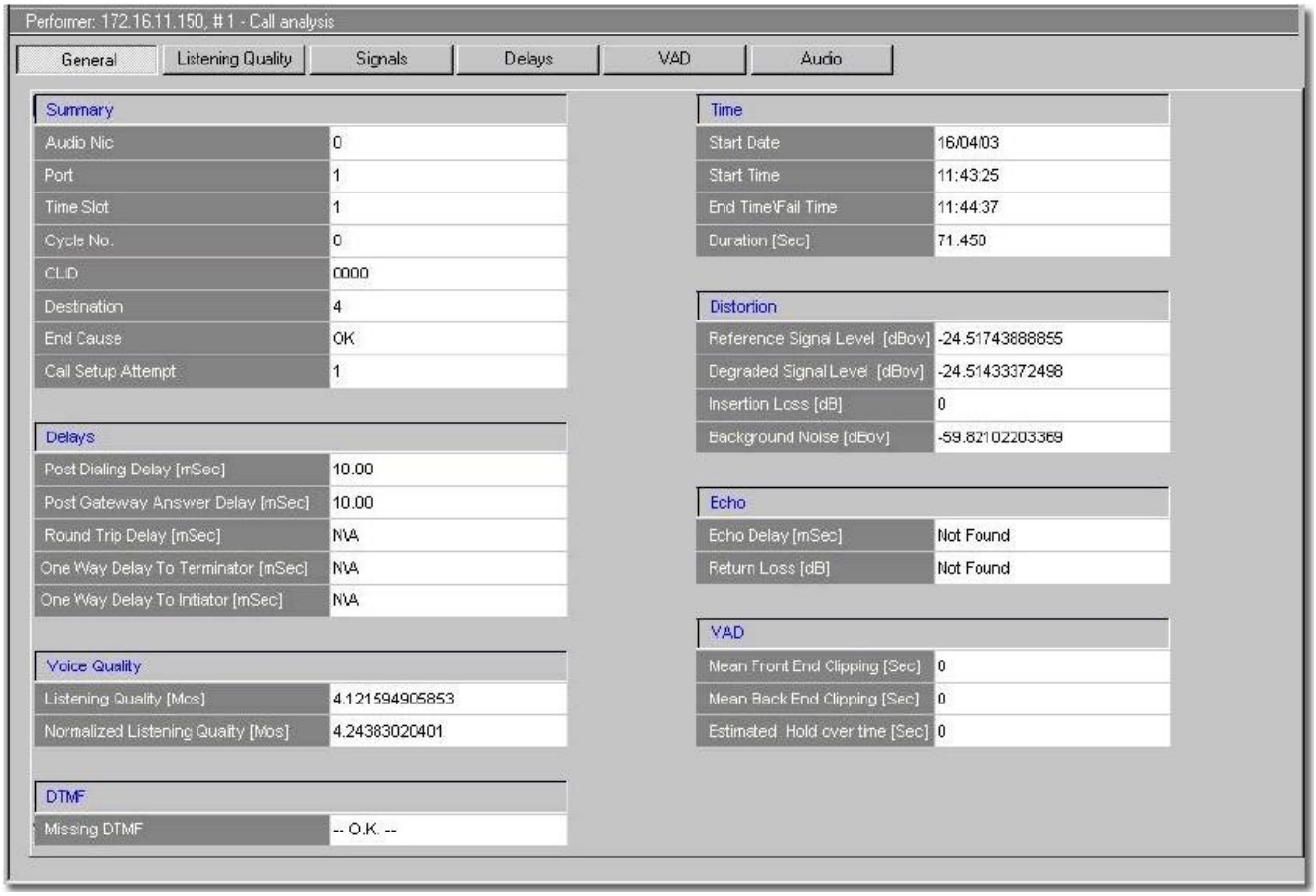
No.	Performer	Audio NIC	Port	Time Slot	Cycle No.	CLID
497	192.168.0...	1	1	1	496	0000
498	192.168.0...	1	1	1	497	0000
499	192.168.0...	1	1	1	498	0000
500	192.168.0...	1	1	1	499	0000
501	192.168.0...	1	1	1	500	0000
502	192.168.0...	1	1	1	501	0000
503	192.168.0...	1	1	1	502	0000
504	192.168.0...	1	1	1	503	0000
505	192.168.0...	1	1	1	504	0000
506	192.168.0...	1	1	1	505	0000
507	192.168.0...	1	1	1	506	0000
508	192.168.0...	1	1	1	507	0000
509	192.168.0...	1	1	1	508	0000
510	192.168.0...	1	1	1	509	0000
511	192.168.0...	1	1	1	510	0000
512	192.168.0...	1	1	1	511	0000
513	192.168.0...	1	1	1	512	0000
514	192.168.0...	1	1	1	513	0000
515	192.168.0...	1	1	1	514	0000
516	192.168.0...	1	1	1	515	0000
517	192.168.0...	1	1	1	516	0000
518	192.168.0...	1	1	1	517	0000
519	192.168.0...	1	1	1	518	0000
520	192.168.0...	1	1	1	519	0000
521	192.168.0...	1	1	1	520	0000
522	192.168.0...	1	1	1	521	0000
523	192.168.0...	1	1	1	522	0000
524	192.168.0...	1	1	1	523	0000
525	192.168.0...	1	1	1	524	0000
526	192.168.0...	1	1	1	525	0000
527	192.168.0...	1	1	1	526	0000
528	192.168.0...	1	1	1	527	0000

Lista de parâmetros medidos:

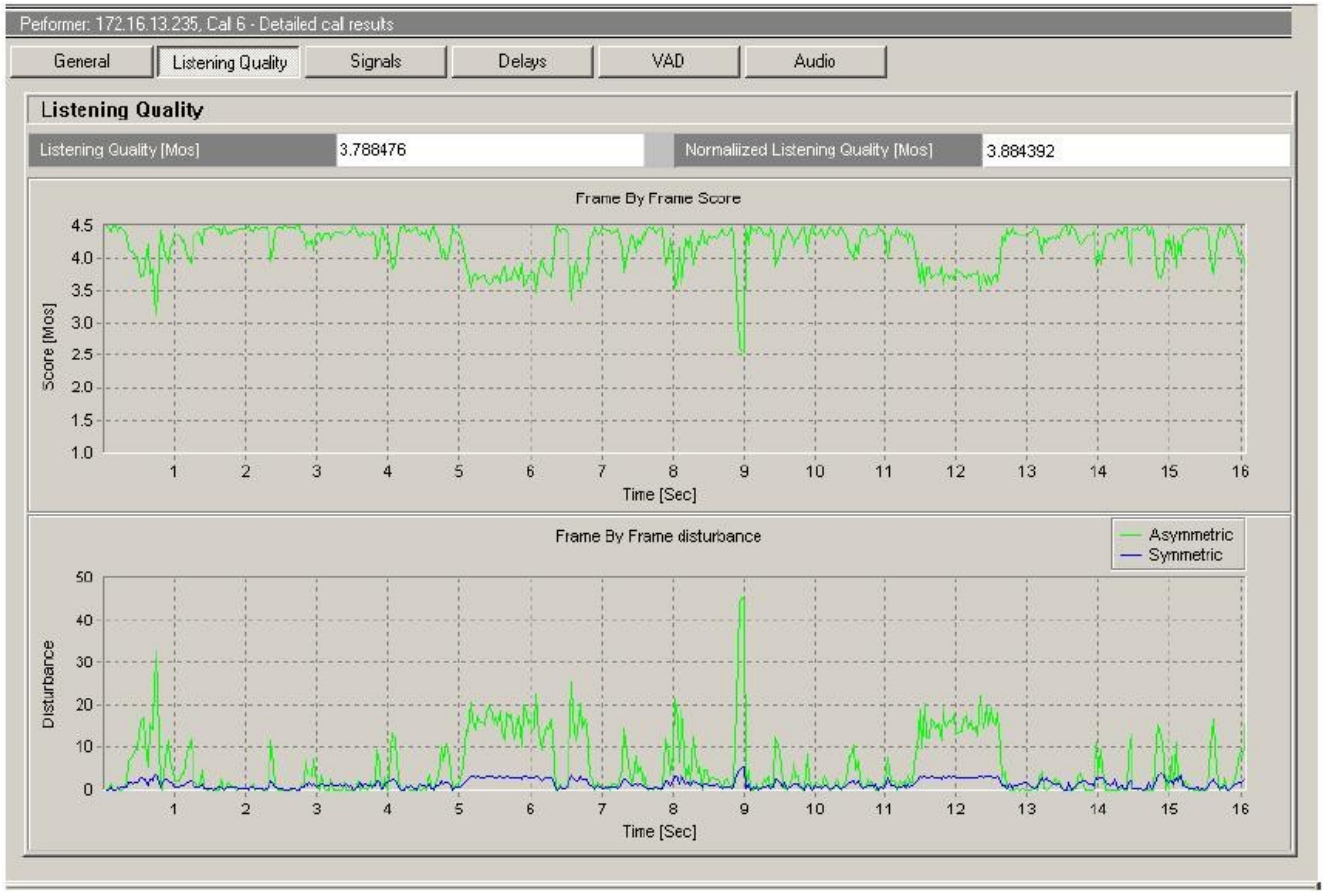
Audio NIC – Número da placa .
Port - Identifica a porta que originou a chamada.
Time Slot – No caso de utilizar interfaces E1 ou T1, identifica em qual "Time Slot" foi originada a chamada.
Cycle No. - Mostra o número da chamada, quando se realiza várias chamadas ciclicamente ou em série.
CLID – Identificação da linha que originou a chamada.
Destination Number – Número da linha a ser chamada.
End Cause – Quando as chamadas são bem sucedidas aparece neste campo "OK", caso ocorra uma falha retorna um código numérico que indica o motivo do problema.
Start Date – A data em que a chamada foi realizada.
Start Time – A hora em que a chamada foi realizada.
End Time/Fail Time – A hora em que chamada foi encerrada.
Call Duration – Duração da chamada (em segundos).
Call Setup Attempts – Número de tentativas até a execução da chamada.
Incoming Listening Effort [somente para PAMS] – O algoritmo PAMS provê este parâmetro que mede o esforço de audição na escala MOS, onde o valor *5* representa nenhum esforço (Completo relaxamento) e o valor *1* representa um tremendo esforço (sem compreensão).
Listening Quality [MOS] – Medida de qualidade do algoritmo PESQ, que é calculado de acordo com o ITU-T P.862. O escore do PESQ varia na escala de –0.5 a 4.5, também em muitos casos entre 1 e 4.5. O escore do PESQ correlaciona-se com a qualidade subjetiva.
Background Noise [Sone; dBov] - Parâmetro provido pelo algoritmo VQMS que indica o ruído de fundo do sinal degrado (o sinal que passou através da rede) na escala Sone/dBov. (No PAMS is medido em Sone; no PESQ é medido em dBov.)
Insertion Loss [dB] - Parâmetro provido pelo algoritmo VQMS que determina o efeito da potência do sinal na rede. A perda de inserção apresenta-se na escala de dB. Quando o sinal entra na rede e sai da rede com a mesma potência, a perda de inserção é de 0 dB.
Incoming Signal Level [dBov] - Parâmetro provido pelo algoritmo VQMS que determina o nível de audio do sinal degradado (sinal que passou através da rede) na escala dBov.
Post Dialing Delay (PDD) [mSec] – Intervalo de tempo entre a discagem ter sido completada e a recepção de um tom apropriado, um anuncio gravado ou a chamada abortada sem o tom (ITU-T E.600).
Post Gateway Answer Delay [mSec] - Intervalo de tempo entre um circuito internacional e a recepção da resposta de supervisão.
Round Trip Delay [mSec] – Tempo gasto pelo sinal, no interior da rede, para seguir do ramal originário QPro para o destino QPro e voltar.
One Way Delay – Tempo gasto pela voz para seguir da fonte ao destino, em milisegundos. Você pode somente medir o atraso em único sentido quando a fonte e o destino estão no mesmo Performer.
NOTA – Se você habilita o "one way delay", as seguintes limitações são aplicadas. Você não pode configurar *stress* de chamadas enquanto o "one way delay" está sendo usado, e chamadas de qualidade precisam ser geradas seqüencialmente e não simultaneamente .

Echo Delay (mSec) – O intervalo de tempo entre a transmissão do sinal de saída e a recepção do sinal refletido. Se o atraso do eco for menor que 4 ms, ou a perda de retorno for maior que 40 dB, o QPro envia a mensagem "*Not Found"*.
Return Loss [dB] – A taxa de potência de sinal refletido (eco).
DTMF Integrity – A integridade da rede quando passam dígitos DTMF. O QPro envia um série de DTMFs e confirma se eles passaram corretamente.

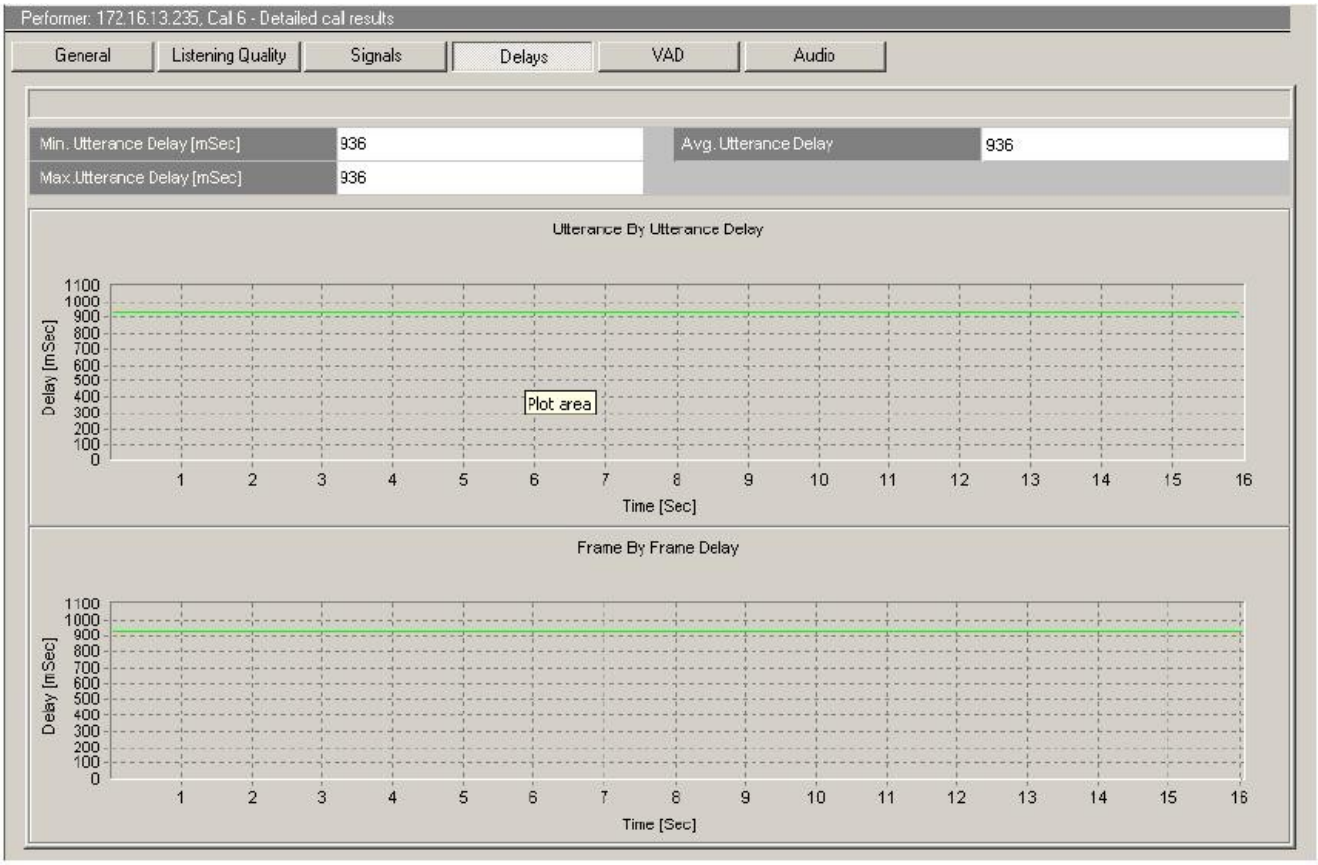
Clicando duas vezes em uma das linhas das chamadas, apresenta-se uma tela onde os parâmetros medidos são agrupados de forma mais organizada.



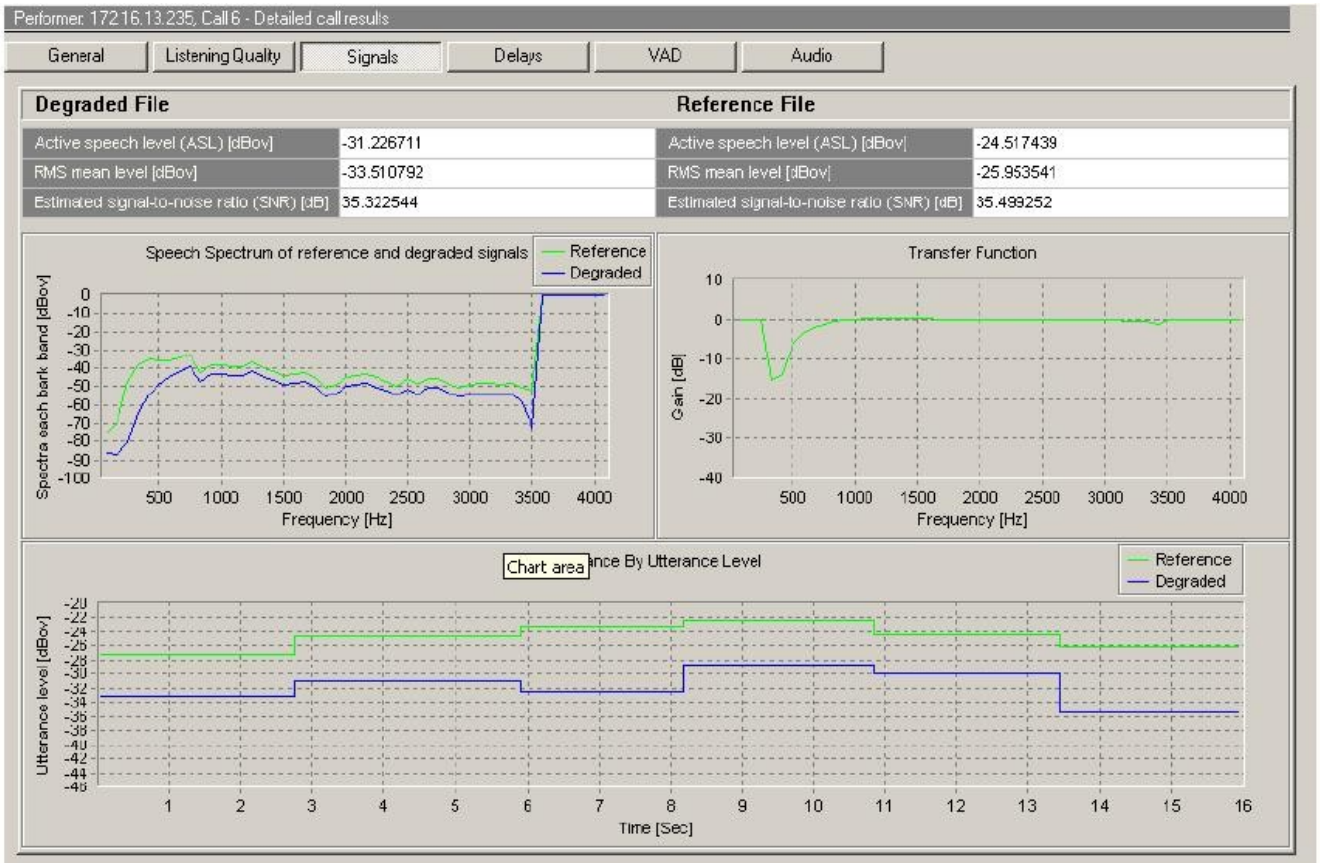
Opção Listening Quality – Mostra a evolução da medida de qualidade da voz ao longo da sequência de chamadas**.**



Opção Delays – Mostra graficamente a evolução dos atrasos ao longo da sequência de chamadas

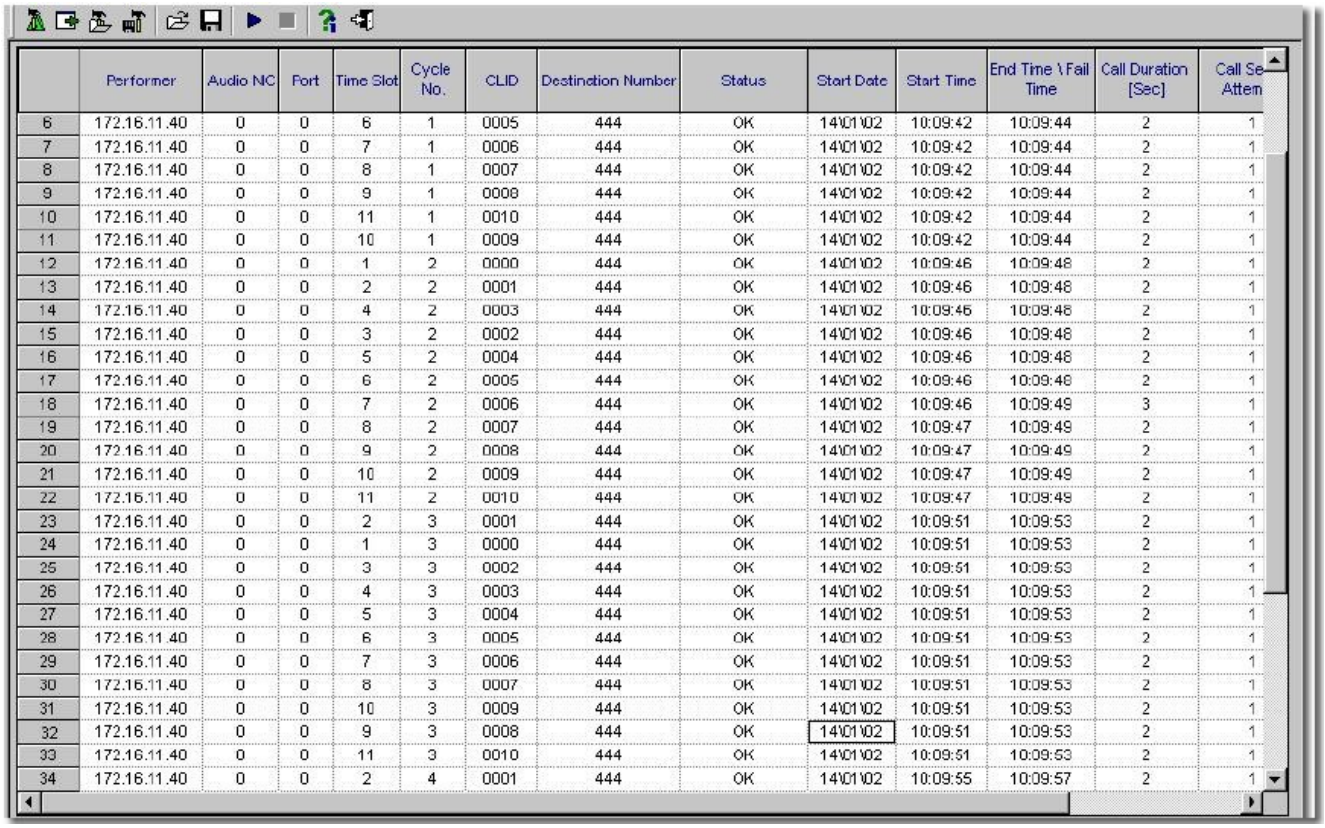


Opção Signals – Mostra graficamente a evolução do sinal de voz no domínio do tempo e da freqüência.



6.3.6.2 STRESS RESULTS

Mostra a sequência de chamadas e seus respectivos parâmetros medidos.

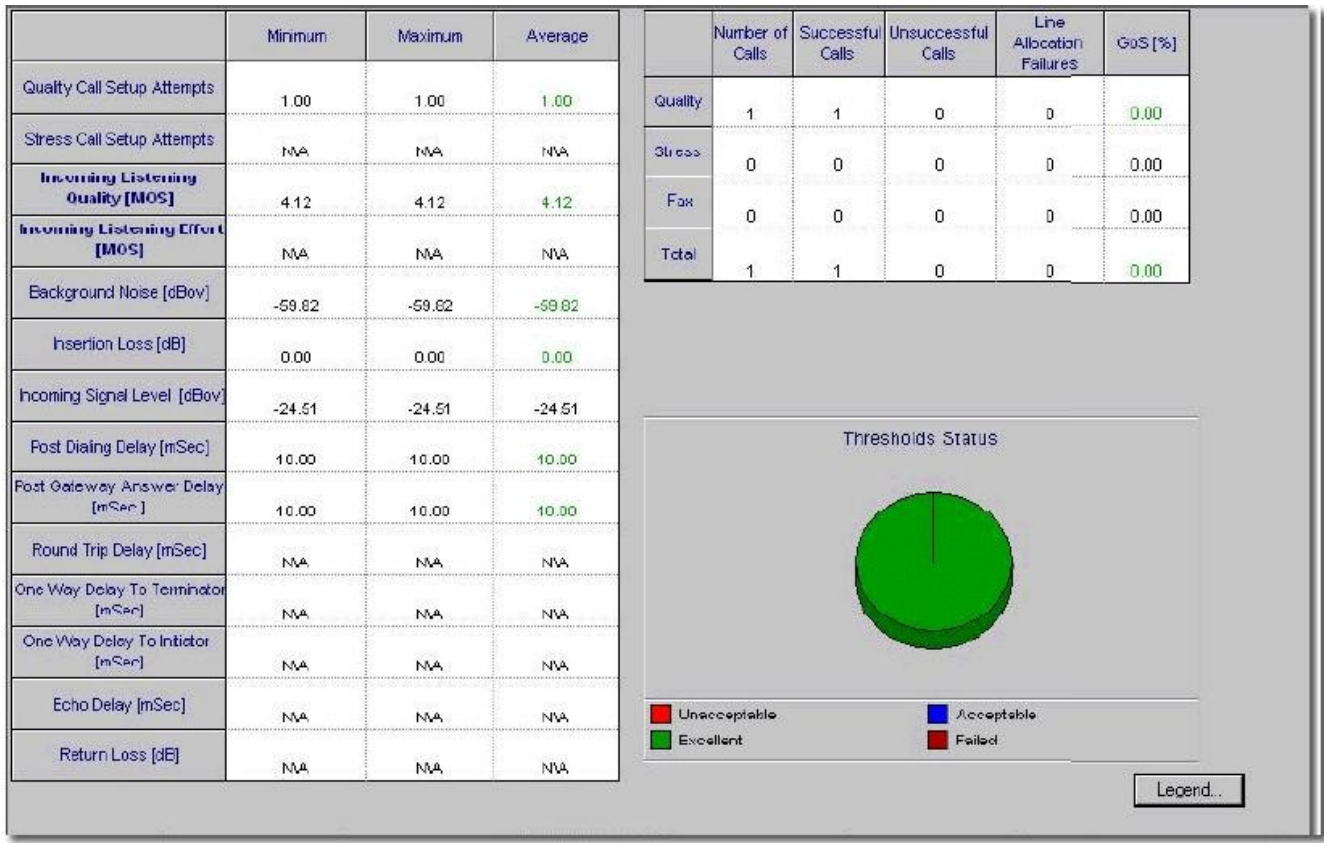


	Performer	Audio NIC	Port	Time Slot	Cycle No.	CLID	Destination Number	Status	Start Date	Start Time	End Time \ Fail Time	Call Duration [Sec]	Call Se Attem
6	172.16.11.40	0	0	6	1	0005	444	OK	14\01\02	10:09:42	10:09:44	2	1
7	172.16.11.40	0	0	7	1	0006	444	OK	14\01\02	10:09:42	10:09:44	2	1
8	172.16.11.40	0	0	8	1	0007	444	OK	14\01\02	10:09:42	10:09:44	2	1
9	172.16.11.40	0	0	9	1	0008	444	OK	14\01\02	10:09:42	10:09:44	2	1
10	172.16.11.40	0	0	11	1	0010	444	OK	14\01\02	10:09:42	10:09:44	2	1
11	172.16.11.40	0	0	10	1	0009	444	OK	14\01\02	10:09:42	10:09:44	2	1
12	172.16.11.40	0	0	1	2	0000	444	OK	14\01\02	10:09:46	10:09:48	2	1
13	172.16.11.40	0	0	2	2	0001	444	OK	14\01\02	10:09:46	10:09:48	2	1
14	172.16.11.40	0	0	4	2	0003	444	OK	14\01\02	10:09:46	10:09:48	2	1
15	172.16.11.40	0	0	3	2	0002	444	OK	14\01\02	10:09:46	10:09:48	2	1
16	172.16.11.40	0	0	5	2	0004	444	OK	14\01\02	10:09:46	10:09:48	2	1
17	172.16.11.40	0	0	6	2	0005	444	OK	14\01\02	10:09:46	10:09:48	2	1
18	172.16.11.40	0	0	7	2	0006	444	OK	14\01\02	10:09:46	10:09:49	3	1
19	172.16.11.40	0	0	8	2	0007	444	OK	14\01\02	10:09:47	10:09:49	2	1
20	172.16.11.40	0	0	9	2	0008	444	OK	14\01\02	10:09:47	10:09:49	2	1
21	172.16.11.40	0	0	10	2	0009	444	OK	14\01\02	10:09:47	10:09:49	2	1
22	172.16.11.40	0	0	11	2	0010	444	OK	14\01\02	10:09:47	10:09:49	2	1
23	172.16.11.40	0	0	2	3	0001	444	OK	14\01\02	10:09:51	10:09:53	2	1
24	172.16.11.40	0	0	1	3	0000	444	OK	14\01\02	10:09:51	10:09:53	2	1
25	172.16.11.40	0	0	3	3	0002	444	OK	14\01\02	10:09:51	10:09:53	2	1
26	172.16.11.40	0	0	4	3	0003	444	OK	14\01\02	10:09:51	10:09:53	2	1
27	172.16.11.40	0	0	5	3	0004	444	OK	14\01\02	10:09:51	10:09:53	2	1
28	172.16.11.40	0	0	6	3	0005	444	OK	14\01\02	10:09:51	10:09:53	2	1
29	172.16.11.40	0	0	7	3	0006	444	OK	14\01\02	10:09:51	10:09:53	2	1
30	172.16.11.40	0	0	8	3	0007	444	OK	14\01\02	10:09:51	10:09:53	2	1
31	172.16.11.40	0	0	10	3	0009	444	OK	14\01\02	10:09:51	10:09:53	2	1
32	172.16.11.40	0	0	9	3	0008	444	OK	14\01\02	10:09:51	10:09:53	2	1
33	172.16.11.40	0	0	11	3	0010	444	OK	14\01\02	10:09:51	10:09:53	2	1
34	172.16.11.40	0	0	2	4	0001	444	OK	14\01\02	10:09:55	10:09:57	2	1

6.3.6.3 STATISTICS

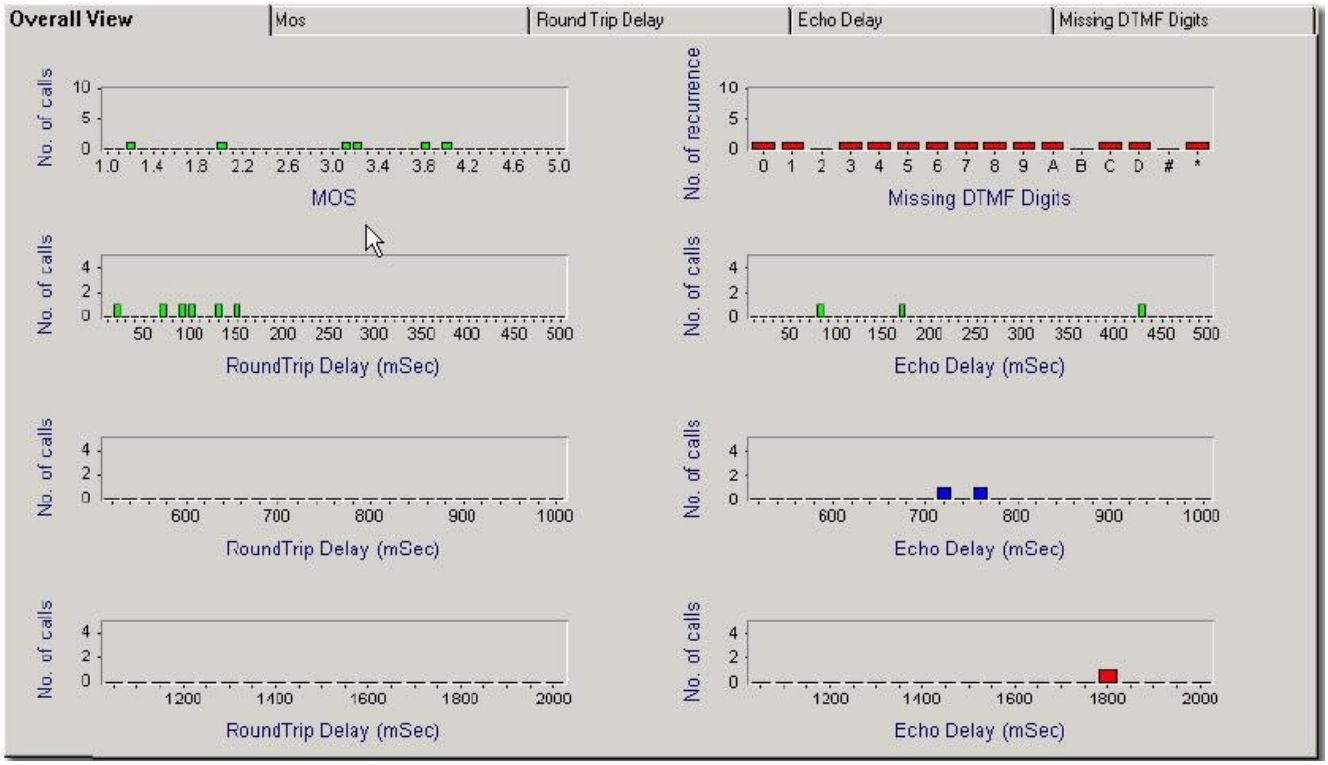
A Tela “estatística” prove um resumo de toda medição realizada, além de mostrar alguns parâmetros de forma diferenciada :

Number of Calls – Número total de chamadas.
Successful Calls – Número de chamadas com sucesso.
Unsuccessful Call – Número de chamadas sem sucesso.
Line Alocation Failures – Número de chamadas que falharam, porque todas as linhas do QPro estavam ocupadas.
GoS Grade of service – Representa a probabilidade de uma chamada ter sido bloqueada ou colocada em filas por algum período de tempo, devido os recursos do sistema terem sido ultrapassados durante a hora de maior ocupação do dia.



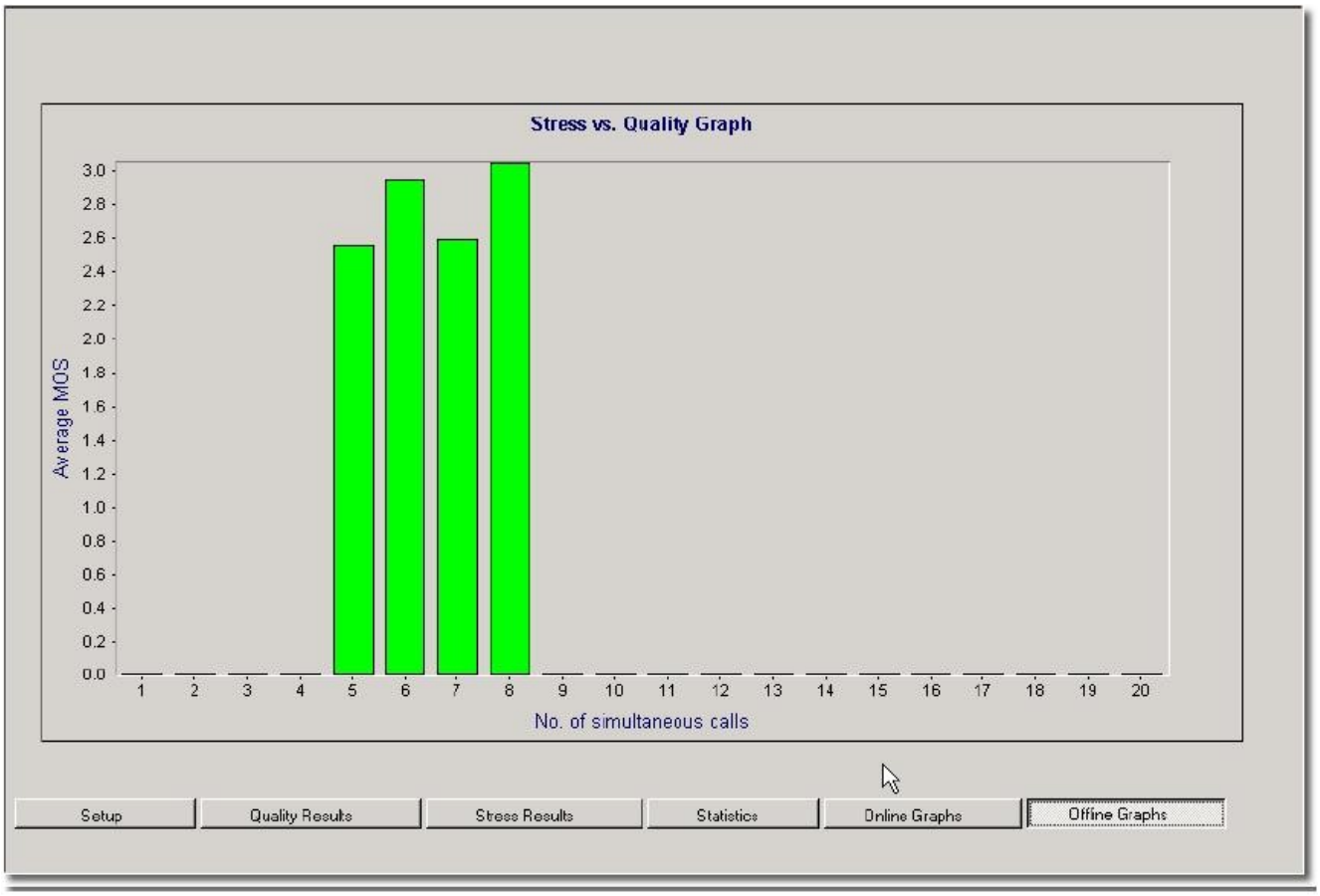
6.3.6.4 ONLINE GRAPHS

A tela “online graphs” mostra os parâmetros medidos de forma gráfica e separados de forma mais organizada. Na opção “Overall View” mostra um resumo gráfico das principais medidas.



6.3.6.5 OFFLINE GRAPHS

A tela “Offline graphs” mostra a média do valor de qualidade MOS para o máximo número de chamadas abertas, proporcionando estatísticas para “stress” de chamadas versus qualidade das chamadas. O eixo “Y”mostra a médio do MOS e o eixo “X”o número de chamadas simultâneas.



7. REGISTROS DA QUALIDADE

“NÃO APLICÁVEL”.

8. ANEXOS

“NÃO APLICÁVEL”.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo