



Universidade Federal do Rio de Janeiro

DISSERTAÇÃO DE MESTRADO

André de Abrantes Davim Pereira e
Souza

Avaliação e Extensão do Modelo E
para Monitoramento de Qualidade
em Tempo Real



Instituto de Matemática



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

ANDRÉ DE ABRANTES DAVIM PEREIRA E
SOUZA

**Avaliação e Extensão do Modelo E
para Monitoramento de Qualidade em
Tempo Real**

Prof. Dr. Paulo H. de Aguiar Rodrigues
Orientador

Rio de Janeiro, Março de 2011

Ficha Catalográfica

de Abrantes Davim Pereira e Souza, André

Avaliação e Extensão do Modelo E para Monitoramento de Qualidade em Tempo Real / André de Abrantes Davim Pereira e Souza. – Rio de Janeiro: UFRJ IM, 2011.

125 f.: il.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro. Programa de Pós-Graduação em Informática, Rio de Janeiro, BR–RJ, 2011.

Orientador: Paulo H. de Aguiar Rodrigues.

I. de Aguiar Rodrigues, Paulo H.. II. Título.

Avaliação e Extensão do Modelo E para Monitoramento de Qualidade em Tempo Real

André de Abrantes Davim Pereira e Souza

Dissertação de Mestrado submetida ao Corpo Docente do Departamento de Ciência da Computação do Instituto de Matemática e Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro como parte dos requisitos necessários para obtenção do título de Mestre em Informática.

Aprovado por:

Prof. Dr. Paulo H. de Aguiar Rodrigues (Orientador)

Profa. Dr. Luci Pirmez

Profa. Dr. Morganna Carmem Diniz

Prof. Dr. Célio Vinicius Neves de Albuquerque

Rio de Janeiro, Março de 2011

*Dedico esta dissertação àqueles cujos trabalhos me deram base para desenvolvê-la.
Mais um pequeno tijolo é posto na grande parede do conhecimento humano,
esperando ser útil a outrem no futuro.*

AGRADECIMENTOS

Dedico meus sinceros agradecimentos para:

- o professor Paulo Aguiar pela orientação acadêmica durante o mestrado e a graduação e pelas tantas oportunidades de pesquisa dos últimos anos;
- todos os amigos que fiz durante tantos anos no LabVoIP e sempre deram apoio a este e outros trabalhos;
- todos os professores e funcionários do PPGI e do DCC pelo excelente curso de mestrado e pela infraestrutura da qual pude desfrutar;
- a minha tão amada Carolina pelo companheirismo e apoio durante todo o andamento deste mestrado;
- a minha família, fisicamente presente ou não, e em especial a minha mãe por terem me amado, orientado e dado condições de chegar até aqui.

“Ciência da computação tem tanto a ver com o computador como a Astronomia com o telescópio, a Biologia com o microscópio, ou a Química com os tubos de ensaio. A Ciência não estuda ferramentas, mas o que fazemos e o que descobrimos com elas.”
citação atribuída a Edsger Dijkstra

RESUMO

Com a Internet sendo usada como um meio multimídia convergente, é essencial o monitoramento em tempo real da qualidade dos serviços VoIP. Diversos métodos de medição de qualidade têm sido criados, entre os quais se destacam o PESQ, um modelo preciso mas que não funciona em tempo real, e o Modelo E, que foi criado para a fase de planejamento de redes e posteriormente adaptado para avaliação em tempo real. O Modelo E, porém, ao ser usado para um propósito diferente daquele para o qual foi criado, não necessariamente gera medições confiáveis, precisando passar por testes para aprová-lo ou não para esta nova função.

Com o objetivo de pôr a prova a precisão do Modelo E, suas medições de qualidade foram comparadas às do PESQ em uma série de diferentes cenários de configurações de codificação de áudio e sob a ação de uma série de fatores de degradação de qualidade presentes nas redes IP, como perdas de pacotes, uniformes ou em rajada, e *jitter*. Centenas de milhares de chamadas foram feitas usando uma aplicação completa de emulação de chamadas. Os resultados da aplicação dos testes de conformidade da P.564 mostraram que o Modelo E não foi devidamente adaptado para funcionar com diferentes configurações de codificação de voz.

Mais uma vez foi usado o emulador para realizar um amplo estudo da qualidade de chamadas de voz conforme medido pelo PESQ. A partir dos resultados das chamadas emuladas, métodos matemáticos foram usados para modelar o comportamento da qualidade. Usando tais modelos, o Modelo E teve sua fórmula e seus parâmetros adaptados para melhorar sua precisão, além de suportar uma variedade maior de configurações de codificação.

Depois da adaptação, o Modelo E passou em quase todos os testes ao qual foi submetido, sendo aprovado conforme os critérios contidos na recomendação P.564. De tal forma, o Modelo E passa a ser um método aprovado e aconselhável para a avaliação de qualidade de chamadas de voz em tempo real.

Palavras-chave: Redes de Computadores, VoIP, Qualidade, QoS, QoE.

Evaluation and Extension of the E-Model to Real Time Quality Monitoring

ABSTRACT

With the Internet being used as a convergent multimedia environment, it is essential to monitor the quality of VoIP services in real time. Several methods for measuring quality have been created, among which stand out PESQ, an accurate model that does not work in real time, and the E-Model, which was created for the network planning stage and later adapted for real time evaluation. However, to use the E-Model for a purpose other than the one for which it was created will not necessarily generate reliable measurements, needing to go through tests to approve it or not for this new role.

In order to put to test the accuracy of the E-Model, its quality measurements were compared to those of PESQ in a number of different encoding scenarios under the action of a number of impairment factors present in IP networks such as packet loss, both bursty and random, and jitter. Hundreds of thousands of calls were made using a full featured call emulation application. The results of running the compliance tests of recommendation P.564 showed that the E-Model was not adequately adapted to work with different configurations of speech coding.

The emulator was used again to perform a comprehensive study of the quality of voice calls as measured by PESQ. From the results of these calls, mathematical methods were used to model the behavior of quality. Using this study, the E-Model E had its formula and its parameters adjusted to improve its accuracy, supporting a wider range of encoding settings.

After the upgrade, the E-Model has been approved in almost all tests to which it was submitted, and approved according to the criteria contained in recommendation P.564. As such, the E-Model is now an approved and recommended method for evaluating the quality of voice calls in real time.

Keywords: Computer Networks, VoIP, Quality, QoS, QoE.

LISTA DE FIGURAS

Figura 2.1: Modelo de Gilbert para perda em rajada	32
Figura 3.1: Pontos de interesse calculados para o <i>codec</i> G.711 μ -law e <i>ptime</i> de 20 ms	47
Figura 4.1: Cadeia de Markov de modelo de perdas em rajada de tamanho > 1	57
Figura 4.2: Diagrama de componentes do emulador de chamadas	61
Figura 4.3: I_{max} da perda e do <i>jitter</i> para diferentes <i>ptimes</i>	67
Figura 4.4: Distribuição dos desvios de previsão do Modelo E original e Estendido quando se usa G.711 sem PLC e sem VAD	71
Figura 4.5: Erro quadrático médio de todas as emulações nos pontos de interesse sem VAD e PLC para o Modelo E original e o Estendido com <i>codec</i> G.711 μ -law	72
Figura 5.1: Qualidade de diferentes <i>ptimes</i> para G.711 sem VAD e sem PLC em função de perdas aleatórias	78
Figura 5.2: Efeitos sobre a curva do MOS medido pelo Modelo E conforme se varia Bpl_1 e Bpl_2	85
Figura 5.3: Comparação das curvas do Modelo E em função de perda aleatória usando uma ou duas variáveis para representar a robustez do <i>codec</i>	87
Figura 5.4: Superfície que representa o MOS em função da perda e do <i>ptime</i> em chamadas G.711 sem VAD e sem PLC	90
Figura 5.5: Comparação da curva de I_d original do Modelo E à simplificação de Clark e à uma nova simplificação proposta	96
Figura 5.6: Qualidade medida pelo PESQ e pelo Modelo E Estendido para testar o efeito de memória recente	101

LISTA DE TABELAS

Tabela 2.1: Valores para o fator A (ITU-T, 2009)	30
Tabela 2.2: Valores de I_e e Bpl para uma série de diferentes <i>codecs</i> (ITU-T, 2003a)	34
Tabela 2.3: Relação entre o fator R e a satisfação dos usuários (ITU-T, 2009)	36
Tabela 2.4: Dados usados como entrada pela VQuality	38
Tabela 3.1: Parâmetros de entrada permitidos pela P.564 para uso de um modelo aderente à recomendação	45
Tabela 3.2: Faixas de classificação do desvio de previsão e resultados necessários para aderência de um modelo à P.564	50
Tabela 4.1: Resultado dos testes de conformidade para o G.711 sem VAD nem PLC	68
Tabela 4.2: Resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas aleatórias e <i>jitter</i>	74
Tabela 4.3: Resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas em rajada	74
Tabela 4.4: Resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas aleatórias e <i>jitter</i>	74
Tabela 4.5: Resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas em rajada	74
Tabela 5.1: Cálculos de novos valores de I_e	83
Tabela 5.2: Valores de Bpl_1 e Bpl_2 calculados usando regressão	88
Tabela 5.3: Valores de Bpl_1 e Bpl_2 calculados usando regressão	92
Tabela 5.4: Valores calculados de FCV e seu EQM associado	94
Tabela 5.5: Todos os valores necessários para usar o Modelo E adaptado com os <i>codecs</i> G.711 e GSM	95
Tabela 5.6: Novos resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas aleatórias e <i>jitter</i>	98

Tabela 5.7: Novos resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas em rajada	98
Tabela 5.8: Novos resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas aleatórias e <i>jitter</i>	98
Tabela 5.9: Novos resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas em rajada	98

LISTA DE ABREVIATURAS E SIGLAS

CM	Cadeia de Markov
CDR	Call Detail Record
EQM	Erro Quadrático Médio
ETSI	European Telecommunications Standards Institute
FCV	Fator de Compensação de VAD
GoB	Good or Better
iLBC	Internet Low Bitrate Codec
IP	Internet Protocol
ITU	International Telecommunication Union
ITU-T	Telecommunication Standardization Sector of ITU
MOS	Mean Opinion Score
PCM	Pulse Code Modulation
PESQ	Perceptual Evaluation of Speech Quality
PLC	Packet Loss Concealment
PoW	Poor or Worse
POLQA	Perceptual Objective Listening Quality Assessment
QoE	Quality of Experience
QoS	Quality of Service
RNP	Rede Nacional de Ensino e Pesquisa
RTP	Real-time Transport Protocol

RTCP	RTP Control Protocol
RTCP-XR	RTP Control Protocol Extended Reports
SEAM	Single Ended Assessment Mode
SLA	Service Level Agreement
TME	Terminated Early
VAD	Voice Activity Detection
VoIP	Voz sobre IP

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Trabalhos Relacionados	18
1.2	Contribuições	19
1.3	Organização da Dissertação	21
2	MÉTODOS DE AVALIAÇÃO DE QUALIDADE DE CHAMADAS DE VOZ	23
2.1	PESQ	26
2.2	Modelo E	28
2.2.1	Conversão do Fator R para Outras Medidas	34
2.3	Modelo E Estendido	36
2.4	VQuality: Uma Implementação do Modelo E	37
2.5	P.563	40
3	PRECISÃO DOS MODELOS DE MEDIÇÃO DE QUALIDADE	41
3.1	Precisão do PESQ	42
3.2	A Recomendação P.564	44
3.2.1	Definição do Modelo Aderente	44
3.2.2	Testes de Conformidade	46
4	EMULAÇÃO DE CHAMADAS E RESULTADOS DOS TESTES DE CONFORMIDADE	52
4.1	Cálculo dos Pontos de Interesse	53
4.2	Emulação de Fatores de Degradação	53
4.2.1	Jitter	54
4.2.2	Perdas Aleatórias	55
4.2.3	Perdas em Rajada	56
4.3	VQSim: Emulador de Chamadas VoIP para Testes de Qualidade	59
4.4	Nova Versão da VQuality	63

4.5	Aplicação da P.564 ao Modelo E	65
4.6	Conformidade à P.564	68
5	ADAPTAÇÃO DO MODELO E	77
5.1	Ajustando o I_e	79
5.2	Representação da Robustez de um <i>Codec</i>	84
5.3	Novas Simplificações do Modelo E	95
5.4	Novos Testes de Precisão	97
5.5	Testes de Efeito de Memória Recente	101
6	CONCLUSÃO	104
	REFERÊNCIAS	108
	APÊNDICE A RELATÓRIOS DOS TESTES DE CONFORMIDADE .	113
A.1	Modelos originais	113
A.1.1	Perdas Aleatórias e <i>Jitter</i>	113
A.1.2	Perdas em Rajada	117
A.2	Modelos adaptados	120
A.2.1	Perdas Aleatórias e <i>Jitter</i>	120
A.2.2	Perdas em Rajada	123

1 INTRODUÇÃO

Com VoIP cada vez mais presente nos serviços de telecomunicação, o usuário final deixa de considerar esta tecnologia apenas como uma opção barata em relação à telefonia tradicional e passa a exigir uma qualidade pelo menos equivalente. E com a Internet sendo usada como um meio multimídia convergente, é essencial o monitoramento em tempo real da qualidade dos serviços VoIP. Com esse objetivo, diversos métodos de medição de qualidade têm sido criados a partir de diferentes filosofias, como analisar o sinal sonoro da chamada ou apenas dados estatísticos do funcionamento da rede e do software do cliente final. Entre eles, se destacam aqueles padronizados pelo Setor de Normatização das Telecomunicações da União Internacional de Telecomunicações (ITU-T), como o PESQ (ITU-T, 2001), o SEAM (ITU-T, 2004) e o Modelo E (ITU-T, 2009). Esses modelos, por serem recomendados por um órgão internacional respeitado, acabam sendo amplamente adotados no mercado e, por consequência, se desenvolvendo melhor com a experiência de seu uso.

A recomendação P.563 define um algoritmo chamado SEAM (*Single Ended Assessment Model*) capaz de medir a qualidade das chamadas de um sistema de telefonia em tempo real. O áudio de cada sentido de uma chamada é analisado e fatores como voz robotizada, sons não naturais como *beeps*, volume de ruídos de fundo são

computados e levados em conta ao se calcular uma nota para a chamada.

O ITU-T definiu em sua recomendação P.862 um algoritmo altamente sofisticado para a medição da qualidade de uma chamada de voz, o PESQ. O PESQ se tornou o algoritmo de medição de qualidade mais confiável e têm sido usado muitas vezes como referência de comparação para métodos menos precisos. Esse algoritmo, todavia, é inviável para uso em tempo real e para aferição de conversas privadas, pois precisa comparar o áudio original com o áudio degradado pelo sistema telefônico, invadindo a privacidade dos usuários. Ou seja, para que o PESQ analise a qualidade de uma rede de telefonia, chamadas artificiais devem ser realizadas para que um áudio original inserido na rede seja comparado à gravação feita no terminal do outro lado da chamada.

Por outro lado, o Modelo E do ITU-T (ITU-T, 2009), originalmente concebido para a fase de planejamento de redes de telecomunicação, tem sido amplamente utilizado na aferição de qualidade de chamadas em infraestrutura VoIP (LUSTOSA et al., 2005). Ao Modelo E, foi feita uma importante extensão que consegue melhorar a capacidade de medição de qualidade em uma série de cenários (CLARK, 2001), chamada Modelo E Estendido. A adoção do Modelo E se deve a sua composição estritamente matemática e baseada única e exclusivamente em dados estatísticos do fluxo de mídia sem necessidade de acesso à mídia em si. Seu uso em tempo real, porém, não foi devidamente homologado e, portanto, não pode ser considerado recomendável.

Recentemente, foi publicada a recomendação P.564 (ITU-T, 2007a), especificando critérios mínimos para modelos objetivos de medição de qualidade de voz para aplicações de telefonia que usam as pilhas dos protocolos IP, UDP e RTP. Além de definir características básicas do modelo aderente, a P.564 apresenta uma metodologia de testes de conformidade baseada na comparação das medições do modelo sob

teste com o PESQ em cenários com degradação variável. Considerando o amplo uso do Modelo E, é fundamental identificar, com o uso da P.564, os cenários para os quais ele pode ser uma opção válida e precisa para a monitoração da telefonia IP em tempo real.

1.1 Trabalhos Relacionados

A ideia de passar o Modelo E pelo crivo da P.564 não é nova. De fato, alguns anos atrás, a comunidade da ITU-T usou os testes de conformidade da P.564 para avaliar o Modelo E (ITU-T, 2007). Os testes foram feitos numa configuração limitada de cenários de codificação e os resultados não foram positivos para o modelo.

Foi testada a capacidade do Modelo E de medir a qualidade de chamadas usando os *codecs* G.711 e G.729, sempre com o mesmo *ptime* (tempo em milissegundos da voz transportada no pacote IP) de 20 ms. Não foi usado VAD (*voice activity detection* - detecção de atividade para supressão de silêncio) nem PLC (*packet loss concealment* - recuperação de perda).

Primeiramente, testes com a versão original do Modelo E foram feitos sem que o modelo obtivesse sucesso nos testes de conformidade da P.564. Em seguida, os autores do trabalho ajustaram parâmetros do modelo de uma maneira não especificada a fim de diminuir o erro quadrático médio dos testes. Ainda assim, o Modelo E não foi aprovado nos testes da P.564. Baseado nessas experiências, os autores concluem o trabalho afirmando que calibrar o Modelo E não é suficiente para melhorar sua performance o suficiente para que seja aprovado pela P.564.

Esse trabalho da ITU-T, porém, não apresenta a metodologia dos testes de forma transparente: o esquema de realização de chamadas não é apresentado e os dados

mostrados não acompanham confiança estatística alguma. Esses fatos diminuem a credibilidade dos testes e fazem com que o trabalho não seja uma resposta razoável a questão da precisão do Modelo E para medir qualidade de chamadas de voz em tempo real.

Além desse trabalho, Estepa (Estepa, A. and Estepa, R. and Vozmediano, J.M., 2002) também apresentou uma proposta de pequenas adaptações ao Modelo E para que ele conseguisse recolher parâmetros de qualidade de um equipamento VoIP. Porém, tais adaptações se limitaram apenas à medir a qualidade do equipamento acústico de um terminal VoIP e o trabalho não vai além das propostas, não indicando qualquer teste ou implementação.

1.2 Contribuições

O presente trabalho é um esforço no sentido de apresentar a teoria e a aplicação da P.564 da forma mais clara e transparente possível e de analisar e melhorar o desempenho do Modelo E e de suas extensões com o uso das ferramentas acumuladas com anos de experiência com modelos de medição de qualidade de chamadas.

Para tal, foram realizadas centenas de milhares de chamadas com diferentes *codecs* e *ptimes*, com uso ou não de VAD e PLC. As chamadas foram emuladas sob impacto de perdas aleatórias de pacotes, *jitter* e perdas em rajada. As rajadas de perda, em especial, são uma espécie de fenômeno comum na internet mas que não foram originalmente consideradas pela P.564. Contudo, o presente trabalho considera o suporte a tais tipos de perda uma capacidade essencial de um bom modelo de medição de qualidade, chegando a adaptar a metodologia de testes da P.564 para que testes envolvendo perdas em rajada pudessem ser feitos, mostrando também, de forma breve, uma técnica de emulação de perdas em rajada adaptada para a necessidade

atual. Além disso, ainda foi testado e comparado o suporte ao efeito de memória recente dos modelos, efeito que faz com que degradações no início de uma chamada tenham impacto menor que degradações no seu fim.

Para realizar os testes da P.564, foi construído um emulador de chamadas capaz de suportar toda a metodologia necessária ao emprego da P.564 através da codificação e decodificação de diferentes *codecs* e emulação de uma série de fatores de degradação de qualidade presentes nas redes reais. A biblioteca que implementa o Modelo E, a VQuality (LUSTOSA et al., 2005), passou por uma reestruturação, tendo sua arquitetura modificada a fim de corrigir algumas falhas e se tornar mais eficiente, sendo reimplementada por completo.

Depois de reimplementada, a VQuality foi posta aos testes de aderência da P.564 com o emprego do simulador construído. Com o emprego de simulações assistidas pela devida confiança estatística, diferentemente dos testes anteriores feitos pela ITU-T, o Modelo E passou com sucesso pelos testes da P.564 em alguns cenários. Contudo, os experimentos indicaram que o Modelo E não foi devidamente ajustado para ser usado com diferentes *ptimes* nem para se adaptar ao uso ou não de VAD e PLC, recursos que podem alterar drasticamente a qualidade de uma chamada. Mesmo assim, os resultados serviram para indicar os pontos nos quais o modelo pode ser melhorado.

Novos experimentos foram feitos simulando chamadas sob efeito de diferentes taxas de perdas de pacotes. Com base nesses resultados, novos parâmetros foram calculados para o Modelo E, que passa a ter parâmetros tabelados para uso com diferentes *ptimes*, VAD e PLC e para os *codecs* G.711 e GSM. Além disso, foi observado um certo padrão de comportamento da qualidade conforme o *ptime* utilizado, tendo sido modificada a fórmula do Modelo E para acondicionar novas variáveis referentes ao *ptime* e ao uso de VAD e PLC. Uma simplificação do cálculo do efeito do atraso

sobre a qualidade da chamada também foi incorporada ao modelo, diminuindo ainda mais seu custo computacional.

Depois de adaptado, o Modelo E novamente foi posto à prova usando a metodologia da P.564. A nova adaptação do Modelo E mostrou um desempenho melhor nos testes da P.564 em comparação ao modelo original, passando nos testes de aderência de diferentes cenários. Portanto, o modelo passa, agora, a poder ser usado em ambientes de produção com uma confiabilidade maior enquanto preserva sua característica de não precisar de acesso direto à mídia e com um custo computacional ainda mais baixo.

1.3 Organização da Dissertação

O presente trabalho está organizado de forma que o próximo capítulo, o segundo, apresenta a tecnologia mais atual de modelos de medição de qualidade de chamadas de voz. Nele, são discutidos o PESQ, o Modelo E, o Modelo E Estendido, a VQuality, uma biblioteca que implementa o Modelo E original e sua extensão, e, por fim, o algoritmo SEAM definido na P.563.

O terceiro capítulo fala sobre a precisão dos modelos de medição de qualidade. Nele, é apresentada a recomendação P.564, suas definições e seus testes de conformidade, e é discutida em detalhes a precisão do PESQ, ferramenta usada pela P.564 como método de referência em seus testes de conformidade.

Começa no quarto capítulo a contribuição efetiva deste trabalho. Nesta parte da dissertação é apresentado como implementar toda a metodologia da P.564 de forma estatisticamente confiável, já que sua recomendação é vaga em relação a uma série de seus aspectos práticos. Este capítulo discute como encontrar os cenários de emulação

de chamadas necessários, como simular os fatores de degradação de rede, apresenta a implementação do simulador de chamadas usada no restante do trabalho e como a P.564 foi aplicada especificamente ao Modelo E. Por fim, este capítulo contém os resultados dos testes de conformidade do Modelo E original à P.564.

O capítulo cinco aproveita os resultados dos testes de conformidade do capítulo anterior para calibrar os parâmetros e modificar as fórmulas do Modelo E, conseguindo melhorar drasticamente sua capacidade de medição de qualidade. O capítulo apresenta testes de conformidade com a versão adaptada do modelo e mostra como tais adaptações fazem com que o modelo seja aprovado numa ampla gama de configurações de codificação diferentes.

Por fim, o sexto e último capítulo conclui o trabalho resumindo os desafios enfrentados, as conclusões possíveis após tais experiências e indica os trabalhos futuros que poderão ser feitos com base nos resultados atuais.

2 MÉTODOS DE AVALIAÇÃO DE QUALIDADE DE CHAMADAS DE VOZ

Para controlar a qualidade dos serviços de telefonia, a experiência do planejamento da rede em sua fase de concepção é essencial para que a infraestrutura instaladas opere dentro de padrões recomendáveis de desempenho e confiabilidade, obtidos através de estudos de engenharia de tráfego. O bom desempenho de um serviço de telecomunicação pode ser medido pela satisfação de seus usuários, satisfação que pode ser afetada por problemas operacionais que podem ocorrer na prestação do serviço. Estabelecer limites aceitáveis para alguns parâmetros críticos da transmissão da voz passa a ser uma ação fundamental para viabilizar uma monitoração proativa em tempo real do serviço.

A complexidade das redes modernas, porém, faz com que a interação entre diversos parâmetros diferentes torne extremamente difícil determinar o nível de degradação do serviço baseando-se apenas na experiência dos gerentes da rede. Uma ação tardia pode levar ao aumento das reclamações dos usuários e à sua insatisfação com o serviço. Uma postura mais sistemática de acompanhar a evolução do serviço se torna imperiosa.

O método conceitualmente mais simples de acompanhar a qualidade de um serviço consiste em realizar pesquisas de qualidade diretamente com os usuários. Tal método é denominado subjetivo. Usualmente, é dado o nome de Pontuação de Opinião Média ao resultado de tais testes. Este termo, em inglês, *Mean Opinion Score*, deu origem à famosa sigla MOS.

Um método pelo qual o MOS pode ser medido de forma subjetiva foi padronizado na recomendação P.800 (ITU-T, 1996). Nele, avaliadores ouvintes atribuem uma pontuação de 1 a 5 à qualidade da fala reproduzida pelo sistema de comunicação em teste, sendo 5 excelente e 1 inaceitável.

Diz-se que, para a fase de planejamento de uma rede, é interessante o uso de estimativas da reação dos usuários ao sistema (JOHANNESSON, 1997). Assim, é estimada a porcentagem dos usuários que classificariam o sistema como bom ou melhor, em inglês, *good or better* (GoB), ou como ruim ou pior, em inglês, *poor or worse* (PoW). Uma terceira estimativa é a porcentagem dos usuários que terminaria a chamada devido à má qualidade, em inglês, *terminated early* (TME).

Porém, durante a fase de planejamento, quando a infraestrutura de fato ainda não foi instalada, o uso de métodos subjetivos só pode ser feita por simulação, onde os impactos na degradação da voz teriam que ser estimados frente aos diversos fatores que podem atuar numa chamada. Este fator torna esta tarefa difícil, senão, impossível de ser realizada a contento pela inexistência de conhecimento das distribuições temporais dos diversos fatores envolvidos, o que normalmente só pode ser concretamente levantado com uma infraestrutura já operacional. Por outro lado, durante a operação normal de um serviço, testes subjetivos podem ser conduzidos em laboratório para chamadas de teste pré-programadas. Todavia, a determinação efetiva da qualidade percebida pelos usuários não pode ser determinada pela violação de privacidade ao se ouvir conversas telefônicas de terceiros, ainda que a monitoração

constante e em tempo real das chamadas dos usuários seja o grande objetivo. Assim, deve ser usado um método que, através de cálculos matemáticos envolvendo os fatores de degradação de uma chamada estime o nível de degradação de qualidade percebida, como se este tivesse sido determinada pelo método subjetivo.

Segundo (MAHDI; PICOVICI, 2009): “Subjective tests are, however, slow and expensive to conduct, making them accessible only to a small number of laboratories and unsuitable for real-time monitoring of live networks. As an alternative, numerous objective voice quality measures, which provide automatic assessment of voice communication systems without the need for human listeners, have been made available over the last two decades. These objective measures, which are based on mathematical models and can be easily computerised, are becoming widely used particularly to supplement subjective test results.”

Em sistemas de telefonia construídos com uma grande variedade de elementos, diferentes tipos de danos à chamada podem ocorrer simultaneamente. Assim, um método objetivo deve também considerar o efeito combinado de vários parâmetros. Também é importante notar que é preferível que um método objetivo seja baseado em estimativas da qualidade da chamada fim-a-fim percebida ao invés de apenas especificações de faixa de variação para os parâmetros objetivos observados de forma individual. Os métodos objetivos mais utilizados e de mais interesse são o PESQ e o Modelo E, que serão apresentados nas próximas seções juntamente com o Modelo E Estendido e o SEAM (P.563). Todos esses métodos são padronizados pelo ITU-T, com exceção do Modelo E Estendido, padronizado pelo ETSI.

2.1 PESQ

O PESQ (*Perceptual Evaluation of Speech Quality*), definido na recomendação P.862 (ITU-T, 2001), introduz um método objetivo que compara o áudio original com a gravação do mesmo áudio após passar pelo sistema de comunicação em teste. O algoritmo do PESQ faz inicialmente um alinhamento temporal das duas gravações, para depois, então, comparar a degradação dos sinais de áudio. Os sinais são separados em trechos chamados elocuições e a comparação das elocuições do áudio original com as do áudio degradado fornece a perturbação sonora sofrida pelo som e também atrasos que a rotina de alinhamento temporal repassa ao modelo perceptual. O resultado após o alinhamento temporal é o valor e a confiança do atraso de cada elocução, levando em conta os períodos de silêncio. O algoritmo de alinhamento é baseado no princípio de comparar a confiança de ter dois atrasos num certo intervalo de tempo com a confiança de haver um único atraso para este intervalo. Essa confiança mede o quão bem feito é o alinhamento temporal de uma elocução e é normalizada como um valor entre 0 e 1 baseada na correlação cruzada entre o áudio original e o degradado. O algoritmo pode identificar variações de atraso tanto durante períodos de silêncio quanto em períodos de voz ativa.

Cada elocução pode ser dividida em pedaços menores e tal processo pode se repetir várias vezes em diferentes pontos da elocução para que seja escolhida a melhor configuração. Se a divisão de uma elocução em duas gerar uma confiança maior que a da elocução unida e as duas partes tiverem atrasos significativamente diferentes, a elocução é definitivamente dividida de acordo. Esse teste é repetido recursivamente dentro das elocuições.

O modelo perceptual do PESQ é usado para calcular a diferença entre o sinal original e o degradado. Os sinais da voz são mapeados para o domínio de tempo-frequência com a Transformada Rápida de Fourier usando janelas de 32 ms sobrepostas em

50% às janelas adjacentes. Filtros e compensações são aplicados aos sinais a fim de simular a percepção dos sinais sonoros pelo sistema auditivo humano. O modelo leva em consideração, por exemplo, que o aparelho auditivo humano possui uma resolução maior para frequências sonoras baixas que para frequências altas. A densidade sonora do sinal de voz degradado através do tempo é calculada desconsiderando os períodos de silêncio e comparada à densidade do sinal original. A diferença pode ser tanto positiva quanto negativa. Quando é positiva, significa que algum componente foi adicionado ao sinal, como ruído, por exemplo. Caso a diferença seja negativa, partes do sinal original estão faltando, possivelmente devido à perda de pacotes.

A diferença entre as densidades sonoras gera a densidade da perturbação sonora. O PESQ leva em conta também uma possível assimetria causada pela distorção do sinal no processo de codificação e decodificação, calculando, além da densidade de perturbação sonora normal (erro audível absoluto), a densidade de perturbação assimétrica, uma medida dos erros audíveis que são significativamente mais altos que o de referência. Quadros consecutivos com perturbação acima de um limite têm seus atrasos recalculados a fim de se tentar chegar a uma melhor sincronização dos sinais no tempo. O modelo perceptual do PESQ utiliza a noção de efeito de memória recente, onde distorções no fim do sinal têm um impacto maior sobre a qualidade que distorções no início. A avaliação final do PESQ é a combinação linear dos valores médio da perturbação e da perturbação assimétrica. Esse valor final pertence ao intervalo $[-0,5-4,5]$ e pode ser convertido à escala do MOS de $[1-5]$.

Apesar do PESQ retornar um MOS sem necessidade de avaliadores, existem claras desvantagens. O PESQ mede a qualidade apenas num sentido, sem considerar outros fatores que degradam a qualidade como o atraso total, que afeta a interatividade, e o eco. Além disso, ele requer acesso direto à gravação original e à gravação degradada, tornando-o inadequado para medições usuais de qualidade em redes de comunicação.

2.2 Modelo E

Para o uso em produção, é indicado um método objetivo computacionalmente eficiente, de implantação simples e que não precise analisar diretamente a mídia. Tal método é definido como não-intrusivo, ou seja, passivo, em relação à mídia. As vantagens de usar tal método podem ser listadas:

- Medidas de qualidade podem ser feitas em tempo real e próximas ao usuário final da chamada;
- Métricas como eco, sinal ou ruído e relatórios de *buffers* de compensação de *jitter* dos clientes finais podem ser usadas;
- O processo de monitorar e analisar chamadas de forma não-intrusiva é mais eficiente computacionalmente por não analisar a mídia;
- Não há necessidade de armazenar ou acessar o conteúdo da mídia, o que evita qualquer possível transtorno legal de invasão de privacidade.

Um modelo objetivo passivo pode obter os parâmetros necessários à medição dos protocolos que transmitem a voz pela rede, como o RTP e seus relatórios RTCP e RTCP-XR (FRIEDMAN; CACERES; CLARK, 2003), uma extensão dos relatórios originais do RTCP contendo muito mais dados importantes para o acompanhamento da qualidade de uma chamada. Tal modelo pode ser mais facilmente distribuído na rede em uma arquitetura escalável sem que qualquer impacto seja notado na rede ou nos elementos da rede. Atualmente, um dos métodos mais amplamente utilizados comercialmente e em trabalhos de pesquisa e desenvolvimento para avaliação de qualidade é o Modelo E, definido na recomendação G.107 (ITU-T, 2009). O Modelo E é um modelo objetivo passivo que atende aos requisitos acima, prevendo um

método para calcular uma métrica única de qualidade de voz chamada fator R , que pode ser convertido para a escala do MOS. O Modelo E e suas características principais são abordadas a seguir.

O Modelo E foi desenvolvido para ser empregado na fase de planejamento de sistemas de comunicação de voz, porém, seu estudo, hoje, o considera uma opção para ferramenta de monitoração da qualidade de voz em sistemas VoIP em tempo real.

O Modelo E usa algoritmos e conceitos de modelos simples mais antigos assim como incorpora conceitos mais próximos do comportamento humano observável nos testes subjetivos. Ele se adapta facilmente tanto à telefonia tradicional, onde os principais fatores degradantes de qualidade eram a perda e o ruído, quanto às redes multimídia modernas, cujos fatores degradantes principais são a perda na rede, eco, atraso, variação de atraso (*jitter*) e a codificação da voz.

O princípio fundamental do modelo é baseado num conceito estabelecido há mais de 30 anos atrás por J. Allnatt (ALLNATT, 1975): “fatores psicológicos na escala psicológica são aditivos.” Este princípio faz com que o Modelo E seja facilmente extensível. Neste sentido, o fator R é composto da soma dos seguintes termos:

$$R = R_o - I_s - I_d - I_e + A \quad (2.1)$$

O fator R foi projetado para variar no intervalo de 0 (extremamente ruim) a 100 (muito bom), mas pode chegar a valores além destes limites em casos extremos.

R_o representa a relação sinal-ruído básica, enquanto os fatores I_s , I_d , I_e e A são fatores que mapeiam a expectativa do usuário. I_s tenta mapear a degradação que pode ocorrer diretamente ao sinal de voz, como uma conexão de volume alto demais,

Tabela 2.1: Valores para o fator A (ITU-T, 2009)

Sistema de comunicação	Fator A
convencional (fixo)	0
rede de celular dentro de edifício	5
rede móvel numa área geográfica ou movendo dentro de um veículo	10
locais de difícil acesso via satélite	20

interferência da voz sobre o *headset* ou a distorção causada pela digitalização do sinal de voz. Já o fator I_d representa a degradação da qualidade devido ao atraso, como eco do transmissor e do receptor e o atraso da chamada se ele for longo, o que pode atrapalhar a interatividade. O fator I_e , por outro lado, mapeia a degradação causada pelo equipamento, como *codecs* de baixo *bitrate* e perdas. Finalmente, o fator A é chamado de fator de expectativa, ou fator de vantagem. A qualidade da transmissão percebida pelo usuário é altamente influenciada pela facilidade ou dificuldade do estabelecimento de uma conexão. Em certas circunstâncias, a facilidade do uso de redes sem fio em relação às redes cabeadas compensa o efeito subjetivo da qualidade inferior da transmissão. Um bom exemplo é o uso de satélites para a comunicação com áreas de difícil acesso. O atraso e a perda nestas redes é maior que em redes cabeadas normais, mas o usuário sabe que a facilidade do uso destas compensa a perda de qualidade. Exemplos do fator de expectativa A são dados na Tabela 2.1. Para redes VoIP, o fator A é normalmente usado como 0.

A recomendação G.107 define valores padrões para cada uma das variáveis que compõem estes fatores. No caso do uso do Modelo E para medir em tempo real e de forma passiva a qualidade de chamadas de voz em redes IP, deve-se usar os valores padrões para o cálculo de R_o e I_s , já que eles não dependem diretamente da tecnologia VoIP. Existe um impacto do atraso no fator R_o , mas é desprezível. Assim, pode-se definir:

$$R_o = 94,77 \quad (2.2)$$

$$I_s = 1,41 \quad (2.3)$$

Já para o cálculo de I_d , usa-se:

$$I_d = I_{dte} + I_{dle} + I_{dd} \quad (2.4)$$

I_{dte} é a perda de qualidade devido ao eco causado pelo transmissor e I_{dle} se refere ao eco do lado do receptor. I_{dd} é a perda de qualidade graças a um atraso muito grande entre o transmissor e o receptor e começa a ser contado quando o atraso de ida e volta numa chamada passa de 200 ms. Todos estes três fatores dependem do atraso medido durante uma chamada, logo, costuma-se considerar I_d como função somente do atraso absoluto entre transmissor e receptor (T_a).

Deve-se lembrar que o atraso absoluto pode ser decomposto em três outras parcelas de atraso, o atraso de codificação/decodificação da voz (T_{codec}), o atraso do transporte da mídia pela rede (T_{rede}) e o atraso introduzido pelo *buffer* de compensação de *jitter* no lado do receptor. Logo:

$$T_a = T_{codec} + T_{rede} + T_{buffer} \quad (2.5)$$

T_{codec} é considerado como constante dependendo apenas do *codec* e *ptime* utilizados. T_{buffer} deve ser medido diretamente no *buffer* de compensação de *jitter* do lado do receptor. Pela falta de medições precisas num único sentido, T_{rede} é estimado através do emprego do RTT (*Round-Trip Time*) medido pelo RTCP ou RTCP-XR usando:

$$T_{rede} = \frac{RTT}{2} \quad (2.6)$$

Obviamente, é sabido que o atraso é assimétrico em grande parte das conexões na Internet. Em conversas interativas, acaba-se sendo otimista numa direção e pessimista na outra em termos de estimação do atraso com o emprego de $RTT/2$.

O cálculo das três parcelas de I_d como definido na G.107 pode ser considerado complexo, já que recorre a fórmulas extensas. Uma forma simplificada e aproximada foi sugerida por Clark (CLARK, 2001) e posteriormente corrigida por Lustosa (LUSTOSA et al., 2005) para simplificar o custo computacional do algoritmo. Abaixo,

segue a fórmula corrigida:

$$I_d = \begin{cases} 0,023 \cdot T_a & , \text{ se } T_a < 175 \text{ ms} \\ 0,111 \cdot T_a - 15,444 & , \text{ se } T_a \geq 175 \text{ ms} \end{cases} \quad (2.7)$$

O fator I_e depende do equipamento utilizado ou, em outras palavras, da tecnologia usada para digitalizar a voz. Perdas de quadros de voz também degradam a qualidade através deste fator. A mais nova versão da recomendação G.107 introduz ao Modelo E a capacidade de medir o impacto de perdas em rajada na qualidade da chamada. O fator I_e foi substituído por I_{e-eff} , que significa I_e efetivo. O novo fator depende do fator antigo e de estatísticas sobre a perda de quadros de voz, conforme a seguinte equação:

$$I_{e-eff} = I_e + (95 - I_e) \times \frac{P_{pl}}{\frac{P_{pl}}{BurstR} + B_{pl}} \quad (2.8)$$

I_e é uma constante baseada no *codec* empregado e representa a qualidade básica de codificação de um *codec*. P_{pl} é a probabilidade de perda de pacotes na rede. Este fator é a porcentagem de pacotes perdidos durante uma chamada. Já o B_{pl} representa a robustez à perda de pacotes do *codec*. I_e e B_{pl} devem ser retirados de uma tabela fornecida pela ITU-T dependendo do *codec* utilizado.

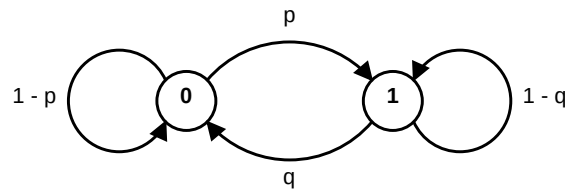


Figura 2.1: Modelo de Gilbert para perda em rajada

A última variável, $BurstR$, é a razão entre o tamanho médio da rajada de perda e o tamanho médio da rajada que seria observada num caso de perdas aleatórias. O modelo de perda aleatória, onde a perda de qualquer pacote tem a mesma probabilidade, pode não ser o mais próximo da realidade onde a probabilidade de perda de um pacote pode depender da perda ou não do pacote anterior. Num modelo

de simulação que considera a probabilidade de perda sendo $(1 - q)$ caso o pacote anterior tenha sido perdido e sendo p caso o pacote anterior tenha sido recebido com sucesso, é possível representar o modelo por uma cadeia de Markov (CM) com dois estados. Esse modelo é conhecido como modelo de Gilbert (GILBERT, 1960) e pode ser vista na Figura 2.1. Apesar de simples, esse modelo se mostra poderoso ao permitir uma métrica para a característica da rajada. Cada transição para o estado 0 significa que um pacote chegou com sucesso ao destino. Por outro lado, cada transição para o estado 1 significa que um pacote foi perdido na rede. Quando a cadeia está no estado de SUCESSO (estado 0), o próximo pacote pode ser recebido com sucesso com probabilidade $(1 - p)$ ou pode sofrer perda com probabilidade p , o que faz com que a cadeia mude para o estado de PERDA (estado 1). No estado de PERDA, o próximo pacote é também perdido com probabilidade $(1 - q)$ ou é recebido com probabilidade q , forçando a transição para o estado de SUCESSO.

Resolvendo a CM, obtém-se facilmente que as probabilidades de estado em equilíbrio são $\pi_{perda} = \frac{p}{p+q}$ e $\pi_{sucesso} = \frac{q}{p+q}$. O número de transições no estado PERDA tem uma distribuição geométrica com média dada por $\frac{1}{q}$, que é o tamanho médio de rajada de perda neste modelo. π_{perda} representa a porcentagem de pacotes perdidos ou $\frac{Ppl}{100}$.

Assumindo π_{perda} como sendo a probabilidade de perda no modelo aleatório, o tamanho de rajada também teria uma distribuição geométrica com média igual a $\frac{1}{1 - \pi_{perda}} = \frac{p+q}{q}$. Dessa forma, o cálculo de *BurstR* seria dado por:

$$\begin{aligned}
 BurstR &= \frac{E[\text{tamanho de rajada}]}{E[\text{tamanho de rajada com perda aleatória}]} \\
 &= \frac{1}{p+q} = \frac{Ppl/100}{p} = \frac{1 - Ppl/100}{q}
 \end{aligned} \tag{2.9}$$

Tabela 2.2: Valores de I_e e Bpl para uma série de diferentes *codecs* (ITU-T, 2003a)

<i>Codec</i>	I_e	Bpl	<i>Codec</i>	I_e	Bpl	<i>Codec</i>	I_e	Bpl
G.711	0	10	G.726 16k	40	69	G.728 16k	16	27
G.711 c/ PLC	0	34	G.726 24k	25	38	G.729	10	18
G.723.1 53k	19	24	G.726 32k	12	24	G.729A	11	17
G.723.1 63k	15	20	G.726 40k	7	24	GSM FR 610	26	43

O texto da recomendação G.107 não cita os *codecs* suportados pelo Modelo E nem seus parâmetros, como I_e e Bpl . Ao invés disso, é mantida outra recomendação, a G.113 (ITU-T, 2007b), cujo objetivo é reunir dados somente sobre degradação devido ao processamento da voz. Essa recomendação, porém, não lista os valores de forma adequada, contemplando somente alguns poucos *codecs*. Além disso, são mostrados diferentes parâmetros de um mesmo *codec* para diferentes cenários diferentes, não sendo claros os valores indicados para uso normal.

Uma contribuição foi posteriormente feita à ITU-T listando de forma mais objetiva uma gama maior de *codecs* (ITU-T, 2003a). A Tabela 2.2 relaciona os valores de I_e e Bpl de tal documento. Somente são contemplados *codecs* tradicionais da própria ITU-T, como G.711 e G.729, tendo sido deixados de fora *codecs* mais modernos e sem padronização da ITU-T, como Speex e iLBC. Os valores tabelados foram calculados somente em função do *codec*, não considerando o *ptime*, por exemplo.

Juntando as definições feitas até aqui, pode-se remodelar a equação do fator R com a seguinte equação, como já proposto por Lustosa (LUSTOSA et al., 2005):

$$R = 93,36 - I_d(T_a) - I_{e-eff}(\text{codec}, \text{perda}) \quad (2.10)$$

2.2.1 Conversão do Fator R para Outras Medidas

A princípio, uma rede pode ser planejada analisando apenas o fator R , mas, geralmente, é mais elucidativo para um não-*expert* quando os limites estão na forma de

reações típicas de usuários. Assim, o fator R pode ser transformado numa série de medidas de qualidade.

Conforme a recomendação G.107 (ITU-T, 2009), as grandezas GoB, PoW e TME são obtidas do fator R como abaixo:

$$\text{GoB} = 100 \cdot E_{rf} \left(\frac{R - 60}{16} \right) \% \quad (2.11)$$

$$\text{PoW} = 100 \cdot E_{rf} \left(\frac{45 - R}{16} \right) \% \quad (2.12)$$

$$\text{TME} = 100 \cdot E_{rf} \left(\frac{36 - R}{16} \right) \% \quad (2.13)$$

Usando a seguinte função de erro E_{rf} :

$$E_{rf}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (2.14)$$

Já o MOS é obtido a partir das seguintes relações:

$$\text{MOS}(R) = \begin{cases} 1 & , \text{ se } R < 0 \\ 1 + 0.035 \cdot R + 7 \cdot 10^{-6} R(R - 60)(100 - R) & , \text{ se } 0 \leq R \leq 100 \\ 4.5 & , \text{ se } R > 100 \end{cases} \quad (2.15)$$

Já que é possível traçar a relação entre o fator R , o MOS, GoB e PoW, pode-se usar a Tabela 2.3 como guia da satisfação dos usuários.

Tabela 2.3: Relação entre o fator R e a satisfação dos usuários (ITU-T, 2009)

Fator R (limite inferior)	MOS (limite inferior)	GoB (%) (limite inferior)	PoW (%) (limite superior)	Satisfação dos usuários
90	4.34	97	~ 0	muito satisfeitos
80	4.03	89	~ 0	satisfeitos
70	3.60	73	6	alguns usuários insatisfeitos
60	3.10	50	17	muitos usuários insatisfeitos
50	2.58	27	38	quase todos os usuários insatisfeitos

2.3 Modelo E Estendido

Uma importante contribuição ao Modelo E foi feita por Allan Clark (CLARK, 2001). Ela foi adotada pelo grupo TIPHON na especificação técnica ETSI TS 102 024-5 V4.1.1 (ETSI, 2003), no que se chamou de Modelo E Estendido. Apesar de a extensão ter sido proposta à ITU-T, ela não foi aprovada (ITU-T, 2003b). O Modelo E Estendido altera o cálculo de I_d e I_e para considerar novos fatores como qualidade instantânea, comportamento de perda de pacotes de forma alternante entre períodos de perdas isoladas e em rajadas e o efeito de memória recente. Com as extensões, a qualidade medida pode se aproximar mais daquela percebida pelos usuários.

Enquanto o modelo original somente mede a qualidade ao final da chamada, o Modelo E Estendido diferencia a qualidade final daquela percebida num dado instante. As alterações na qualidade instantânea, para melhor ou para pior, não são percebidas imediatamente pelo usuário. O Modelo E Estendido usa curvas exponenciais com constantes de tempo de 5 s para a transição de bom para ruim e de 15 s de ruim para bom.

O comportamento de perdas alternantes usa uma CM com quatro estados para modelar a perda ocorrendo em rajada ou de forma isolada. Uma rajada se inicia quando o número de pacotes recebidos até a próxima perda é menor do que oito, caso contrário a perda é isolada. À medida que perdas ocorrem, as taxas entre os

estados da CM são dinamicamente recalculadas.

O efeito da memória recente reflete a maneira como um usuário se lembra da chamada após o seu término. Um intervalo de qualidade baixa tende a ter um impacto negativo maior quando acontece próximo do fim da chamada. Num exemplo dado por Clark (CLARK, 2001), um período de 15 s de ruído foi deslocado do início para o final de uma chamada de 60 s de duração, provocando uma queda de 3,82 para 3,18 no MOS.

O maior diferencial do Modelo E Estendido em relação ao modelo original é sua capacidade de medir adequadamente o impacto das rajadas de perda sobre a qualidade. Na época de seu desenvolvimento, o Modelo E ainda não possuía a variável *BurstR* em sua fórmula e, portanto, o Modelo E Estendido era a única opção para mapear o efeito das rajadas de perda. Mesmo que seu uso não tenha sido aprovado pela ITU-T, sua proposta parece ter levado à incorporação de *BurstR* à fórmula do modelo. Até o momento, nenhum trabalho se dedicou a testar o desempenho de qualquer uma das duas alternativas de medir o impacto das perdas em rajada comparando a testes subjetivos.

2.4 VQuality: Uma Implementação do Modelo E

A VQuality é uma biblioteca que implementa tanto o Modelo E quanto o Modelo E Estendido visando seu uso em medições de qualidade de chamadas em tempo real. A primeira versão da biblioteca foi desenvolvida por Lustosa (LUSTOSA et al., 2005) e, além de implementar os modelos de medições, continha facilidades como geração e envio de CDRs (*Call Detail Record* – relatórios sobre a chamada guardados em bancos de dados). Durante o desenvolvimento do presente trabalho, com um estudo mais minucioso sobre os detalhes do uso do Modelo E para a avaliação de

Tabela 2.4: Dados usados como entrada pela VQuality

Dados da rede	RTT	quadros por pacote
Dados do buffer de compensação de jitter	tamanho do <i>buffer</i>	-
Dados da mídia	<i>codec</i>	<i>ptime</i>
Dados de reprodução	quadros tocados	quadros faltando

qualidade em tempo real, algumas inconsistências foram encontradas na concepção da VQuality. Tais problemas são discutidos e soluções são apresentadas mais adiante, no Capítulo 4, onde são mostradas em detalhes todas as ações tomadas para a realização dos testes de chamadas.

Tanto na nova versão da VQuality quanto na antiga, é esperado que a biblioteca seja incorporada ao software do cliente final a fim de receber dados sobre o fluxo de mídia no local mais próximo possível do usuário final. Assim, a VQuality pode ter acesso a quatro tipos distintos de dados que lhe servem de entrada. A biblioteca pode receber eventos e avaliar a qualidade a qualquer instante, conseguindo acompanhar a evolução da qualidade da chamada com o passar do tempo. Os tipos de dados são listados na Tabela 2.4 ao lado dos dados usados como entrada pela VQuality para fazer a medição de qualidade. Os dados de entrada são iguais tanto para o uso do Modelo E original quanto do Estendido.

Os dados da rede podem ser facilmente obtidos da pilha RTP/RTCP, sendo o RTT usado para estimar a parte do atraso da chamada provocado pela rede. O único dado do *buffer* de compensação de *jitter* necessário é o seu tamanho, ou seja, o atraso que ele causa ao tentar compensar o *jitter* da rede. Os dados da mídia, *i.e.* *codec* e *ptime*, não costumam mudar durante uma chamada. Já os dados de reprodução são os eventos que devem ser reportados à biblioteca assim que acontecem no momento de tocar a mídia, se o quadro é tocado com sucesso ou se está faltando.

Como tanto o Modelo E original quanto o Modelo E Estendido estão implementados

na VQuality, sua arquitetura é suficientemente genérica para suportar dois diferentes modelos através de herança de classes, podendo, futuramente suportar novos modelos facilmente. Cada modelo é implementado através de duas classes, uma que recebe as chamadas dos eventos do cliente e guarda os dados e gera as estatísticas necessárias pelo modelo e outra classe que guarda o modelo matemático de avaliação propriamente dito. A primeira camada é mais externa, servindo de interface para a camada mais interna que implementa o modelo. A camada interna não precisa implementar rotinas que geram e armazenam as estatísticas da chamada, ficando essa tarefa para a camada externa.

A antiga versão da VQuality já havia sido integrada ao GnomeMeeting (atual Ekiga), no Linux, e ao OpenPhone, no Windows (LUSTOSA et al., 2005). A biblioteca também foi integrada à ferramenta ACME (LUSTOSA et al., 2008), uma solução distribuída para a realização de testes automatizados de qualidade em ambiente VoIP, e ao Asterisk (SOUZA, 2008), viabilizando a aferição de qualidade neste ambiente livre favorito de telefonia IP. Mais recentemente, a reimplementação da biblioteca foi integrada novamente ao Asterisk.

A biblioteca tem sido usada no ambiente modelo na UFRJ do fone@RNP (RNP, 2011) como apoio para as práticas de QoS da instituição. Determinadas chamadas VoIP que passam pelo *gateway* da instituição têm sua qualidade medida pela VQuality, tendo seu MOS guardado num banco de dados local. Planeja-se expandir as ferramentas de estatísticas do ambiente para que a biblioteca seja usada como uma ferramenta chave para a gerência das instituições do serviço. A partir dos diferentes dados coletados pela biblioteca, além do próprio MOS, é possível traçar todo o perfil da qualidade de um serviço de telefonia IP em função da hora do dia ou da qualidade da rede que o serve.

2.5 P.563

Entre 2002 e 2004, a ITU-T promoveu uma competição para padronizar um modelo não intrusivo de medição de qualidade baseado na análise do sinal de voz (RIX et al., 2006). Em termos de condições da rede, o escopo do modelo foi definido de forma mais ampla que no PESQ, incluindo, em particular, testes subjetivos com uma gama maior de variações acústicas, medições de condições de rede e diferentes tipos de ruído. Um dos modelos propostos, o SEAM (*Single Ended Assessment Model*) foi adotado pela ITU-T como a recomendação P.563 (ITU-T, 2004).

Nesse algoritmo, vários parâmetros são extraídos para análise de quão natural é a voz, ruído e interrupções na fala e efeito de *clipping*, que são pequenos cortes no início de um período de fala. Com base nesses parâmetros, uma qualidade parcial de fala é estimada para cada classe de distorção e a qualidade final é uma combinação linear dessas classes com uma série de outros fatores menos importantes medidos a partir do sinal da voz.

Durante os testes de padronização da P.563, o coeficiente de correlação de Pearson medido entre o SEAM e 24 testes de plateia chegou a 0,88, abaixo do valor de 0,93 obtido pelo PESQ (RIX et al., 2006). Assim como o PESQ, o SEAM não leva em conta os efeitos do eco ou o impacto do atraso sobre a interatividade.

3 PRECISÃO DOS MODELOS DE MEDIÇÃO DE QUALIDADE

Qualquer modelo objetivo de medição ou predição de qualidade não passa de uma tentativa de se aproximar da qualidade como percebida por um ser humano, sendo impossível garantir precisão exata das estimativas de qualidade. Além disso, nenhum modelo é completo o suficiente a ponto de ter seu uso contemplado em todos os cenários possíveis. Assim, o melhor que se pode fazer é definir os cenários de uso indicado para o modelo objetivo e medir sua precisão em relação a avaliações reais de uma plateia.

O uso de uma plateia para a aferição de qualidade, porém, é uma tarefa custosa, consumindo tempo e recursos. Dessa forma, muitas vezes a plateia deixa de ser a referência para testes de precisão e adota-se, em seu lugar, um modelo objetivo de referência. O PESQ tem sido usado em diversas recomendações da própria ITU-T e em trabalhos acadêmicos como referência devido à sua boa precisão previamente testada em outros trabalhos da literatura, como a recomendação P.564 (ITU-T, 2007a) que traz uma metodologia de teste de conformidade para modelos objetivos e não-intrusivos de medição de qualidade.

Neste capítulo, é brevemente discutida a precisão do PESQ de forma específica, já que este serve de referência para os testes de precisão do Modelo E. Ao final do capítulo, é apresentada uma metodologia de emulação de chamadas para a realização de testes de precisão na P.564 para avaliação dos modelos objetivos E e E Estendido, qualificando-os para uso em tempo real.

3.1 Precisão do PESQ

Desde seu lançamento em 2001, o PESQ tem sido analisado por diversos trabalhos, tendo sua utilidade muitas vezes confirmada, mas também recebendo diversas críticas. A recomendação deixa claro que o algoritmo do PESQ não provê uma avaliação completa da qualidade de transmissão de uma chamada, mas somente mede o efeito da distorção e ruído sobre a fala em um sentido da chamada. Efeitos como perda devido ao volume da fala, atraso, eco e outros fatores de degradação relacionados à interação entre os dois sentidos da chamada não são refletidos na medição do PESQ.

Alguns outros trabalhos, porém, apontam para situações onde a precisão do PESQ pode ser questionada. Em 2002, Pennock (PENNOCK; HOLUB; SMID, 2002) concluiu que o PESQ é uma ferramenta útil para identificar problemas no sistema telefônico, mas não boa suficiente para ser usada como parâmetro em contratos (SLA) ou para tomar decisões importantes, pois ele pode levar a conclusões erradas. Esse mesmo trabalho constatou que a precisão do PESQ cai conforme o MOS também cai, sendo o modelo menos preciso em cenários de baixa qualidade. Pennock também chegou à conclusão que o PESQ pode ter problemas em medir com precisão a qualidade em cenários com perdas em rajada. Ding (DING et al., 2007) conclui que o PESQ é bom suficiente para medir MOS de chamadas sob perda aleatórias de pacotes, mesmo com o uso de diferentes valores de *p_{time}*. Ele também conclui, porém, que o PESQ não mede bem o efeito da perda de pacotes para G.711

sem PLC.

Manjunath (MANJUNATH, 2009) afirma que, como o PESQ somente foi verificado para um pequeno subconjunto entre todos os cenários aos quais ele é aplicado hoje, ele não foi devidamente validado para uso com *buffers* de compensação de *jitter* adaptativos, *clipping* temporal da voz, *clipping* da amplitude da voz e múltiplas vozes simultâneas. Também é dito que ele não deve ser usado para prever a qualidade perceptual da voz quando recursos avançados de processamento de voz estão ativos no sistema sob teste. Afirma-se ainda que o método é sensível demais a variações no atraso, enquanto a qualidade percebida por testes subjetivos não é. Um *white paper* da empresa Ditech Networks (DITECH NETWORKS, 2007) aponta problemas parecidos, listando possíveis problemas do PESQ com uso de *buffer* de compensação de *jitter* e PLC.

Apesar de qualquer desvantagem do PESQ, ele ainda é o melhor modelo de medição de qualidade, sendo usado como modelo de referência em várias recomendações da própria ITU-T (ITU-T, 2002, 2007a) e em diferentes trabalhos acadêmicos. Como exemplos, pode-se citar o uso do PESQ como mecanismo de medição de inteligibilidade da voz (BEERENDS et al., 2004), mecanismo para medição de qualidade de voz sintética (CERNAK; RUSKO, 2005), referência para criação e calibração de algoritmo de PLC para *codecs* AMR (XYDEAS; ZAFEIROPOULOS, 2003) e CELP (LEE et al., 2003), além de vários outros trabalhos que estudam a qualidade de VoIP sobre redes sem fio (QIAO; SUN; IFEACHOR, 2008; GOUDARZI; SUN; IFEACHOR, 2009).

Está no final do processo de adoção uma nova recomendação da ITU-T com um novo método de medição intrusiva de qualidade que irá substituir o PESQ. Este algoritmo, chamado de P.OLQA (*Objective Listening Quality Assessment*), ganhará o código P.863 (ITU-T, 2011). A recomendação, depois de anos de desenvolvimento,

foi finalmente aprovada em janeiro de 2011, mas seu texto final ainda não foi tornado público.

3.2 A Recomendação P.564

A recomendação P.564 (ITU-T, 2007a) foi criada com o intuito de servir de guia de implementação e teste de validação de modelos objetivos de monitoramento de qualidade de voz em tempo real. Ela define as características básicas que um modelo deve implementar como quais parâmetros podem ser usados como entrada, além dos testes específicos a que o modelo deva ser submetido e os resultados mínimos necessários para conformidade. A recomendação aplica-se a modelos tanto de banda estreita (300 a 3400 hertz) quanto de banda larga (30 hertz a 7000 hertz) e que especificamente usam as pilhas IP/UDP/RTP para o transporte da mídia. Um modelo objetivo que contempla a P.564 pode ser usado em diferentes cenários como monitoramento da qualidade de transmissão em operação ou manutenção, monitoração de SLAs entre provedores de telefonia IP ou uso em laboratório para testes de aferição de qualidade de equipamentos.

Os testes de conformidade medem o desempenho do modelo sob teste em comparação ao PESQ em situações envolvendo diferentes valores de fatores de degradação, chamados pontos de interesse. Esses pontos representam cenários montados para contemplar uma gama variada de combinações de fatores de degradação.

3.2.1 Definição do Modelo Aderente

Um modelo aderente à recomendação P.564 deve fazer sua avaliação baseando-se apenas em parâmetros do fluxo de mídia pela rede e de sua manipulação no cliente,

Tabela 3.1: Parâmetros de entrada permitidos pela P.564 para uso de um modelo aderente à recomendação

Modo de operação	Parâmetros permitidos
Dinâmico	- Tempo de chegada dos pacotes - Cabeçalhos RTP, UDP e IP - Informações dos relatórios RTCP/RTCP-XR - Informações do <i>codec</i>
Estático	- Tempo de chegada dos pacotes - Cabeçalhos RTP, UDP e IP - Informações <i>a priori</i> sobre o funcionamento do cliente final - Informações do <i>codec</i>
Embutido	- Tempo de chegada dos pacotes - Cabeçalhos RTP, UDP e IP - Perdas e atraso do <i>buffer</i> de compensação de <i>jitter</i> - Informações do <i>codec</i>

mas sem qualquer acesso à mídia em si. Um modelo aderente pode trabalhar nos seguintes modos de operação: dinâmico, estático e embutido. Dependendo do modo de operação, a recomendação lista diferentes fontes das quais o modelo pode retirar dados a fim de inferir a qualidade de uma chamada. A Tabela 3.1 lista os parâmetros que a recomendação permite que sejam usados pelo modelo conforme o tipo de operação exercido.

A partir da análise dos dados ao seu alcance, o modelo de avaliação deve produzir, ao final da chamada, o MOS relativo à experiência do usuário final sob as condições observadas. Fatores como volume da fala, ruído de fundo, atraso e eco não devem ser refletidos no cômputo do MOS. Mesmo assim, diagnósticos adicionais podem ser feitos baseados na análise da mídia da chamada ou no atraso. No caso do atraso, especificamente, é compreensível que a P.564 não considere o impacto do fator já que o PESQ, seu método de referência, não é capaz de medir o impacto do atraso sobre a qualidade. Todavia, como o PESQ ainda é capaz de medir em pequena escala o impacto da variação do atraso sobre a qualidade, a P.564 permite que o atraso do *buffer* de compensação de *jitter* seja considerado. Mesmo deixado de fora da avaliação principal de um modelo aderente à P.564, um fator tão importante

quanto o atraso deve ter seu impacto levado em conta em diagnósticos adicionais do modelo, principalmente quando a monitoração ocorre na internet, onde o atraso pode ser significativo.

3.2.2 Testes de Conformidade

Os testes de conformidade consistem em comparar o MOS gerado pelo modelo sob teste ao MOS gerado pelo PESQ em uma série de variados cenários, envolvendo sempre o impacto simultâneo de dois fatores de degradação sobre o fluxo de mídia da chamada de voz. A recomendação define quatro fatores de degradação: perda aleatória de pacotes, *jitter* e dois tipos diferentes de ocupação de banda.

A P.564 estipula apenas perdas aleatórias, o que é uma limitação. Na Internet, as perdas em roteadores congestionados podem ocorrer em rajadas associadas a um alto *jitter*, o que motivou à extensão dos testes da P.564 para cenários de perda em rajada.

Deve-se salientar que, conforme a própria recomendação, os testes devem ser repetidos para cada configuração de *codec*, *ptime* e uso de PLC ou VAD, opções configuráveis pelos usuários em chamadas VoIP. É possível que um modelo não passe nos testes de uma determinada combinação de *codec* e *ptime*, o que não significa que ele não possa ser usado com outras configurações aprovadas nos testes.

Cada fator de degradação deve ser configurado para variar de acordo com um parâmetro único. Por exemplo, no caso da perda aleatória, o parâmetro é a porcentagem de pacotes perdidos na chamada. Isso permite representar os cenários dos testes num plano cartesiano cujos eixos são os valores dos parâmetros dos dois fatores de degradação. A Figura 3.1 mostra o plano formado pelo *jitter* e pela porcentagem

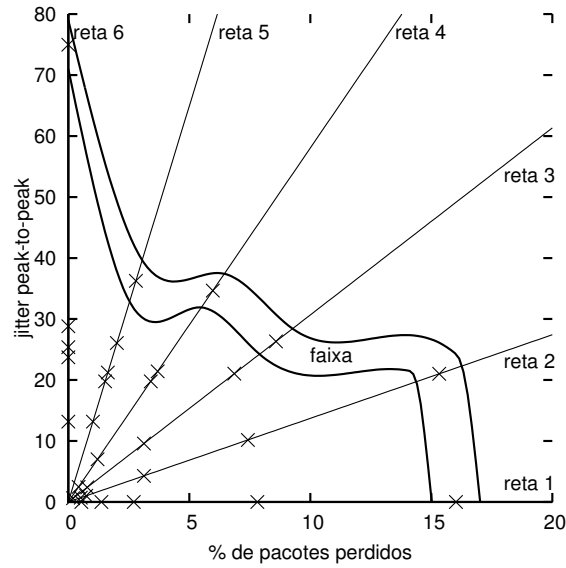


Figura 3.1: Pontos de interesse calculados para o *codec* G.711 μ -law e *ptime* de 20 ms

de perda aleatória, fatores de degradação no caso do *codec* G.711 μ -law com *ptime* 20 ms e sem uso de VAD ou PLC. Essa figura será usada para exemplificar como os pontos de interesse são encontrados.

Para cada fator de degradação deve ser encontrado o limite I_{max} , um valor que sem a interferência de outro fator faz com que o MOS da chamada fique no intervalo [1,3–1,5]. Uma vez calculado o I_{max} para os dois fatores de degradação, as escalas máximas dos eixos ficam definidas como mostrado na Figura 3.1. No caso, o I_{max} do *jitter* é próximo de 80 ms e o I_{max} da perda aleatória é próximo de 20%. Com as escalas ajustadas, a identificação de pontos de interesse pode ser distribuída em semirretas regularmente distribuídas no quadrante do plano, separadas por 18 graus. Na Figura 3.1, a primeira e a última reta correspondem aos eixos.

A recomendação preza que os pontos de interesse selecionados explorem várias combinações diferentes dos fatores de degradação. Os pontos são definidos como aquelas

combinações de parâmetros de fatores de degradação que fazem com que o MOS calculado pelo PESQ para uma chamada fique num determinado intervalo. Os intervalos são calculados em função da qualidade máxima medida pelo PESQ que um *codec* pode oferecer. Portanto, para descobrir os pontos de interesse, somente a avaliação do PESQ é necessária. No caso do *codec* G.711, cujo MOS máximo é 4,5, usando *p_{time}* de 20 ms sem PLC ou VAD, por exemplo, os 5 pontos de interesse definidos pela P.564 para cada reta devem produzir MOS nos seguintes intervalos: [1,80–2,10], [2,70–3,00], [3,45–3,75], [3,90–4,20] e [4,20–4,50]. Todos os 6 pontos de interesse selecionados e que estão no interior da faixa da Figura 3.1 produzem MOS no intervalo correspondente, [1,80–2,10].

É importante ressaltar que dificilmente um aumento linear em fator de degradação fará com que o MOS também caia linearmente, dificultando a localização de pontos no plano que representam cenários com MOS em intervalo desejado. Geralmente, a influência de um fator sobre a qualidade é complexa e a disposição dos pontos de interesse e seus intervalos no plano tende a ser assimétrica, como acontece na Figura 3.1.

Todos os experimentos devem ser realizados sobre quatro arquivos de áudio fornecidos pela recomendação. Cada arquivo possui duas frases sem conexão lógica entre elas, espaçadas por alguns segundos, num total de oito segundos de áudio. Os áudios são de uma mulher falando inglês britânico, uma mulher falando francês e dois são de um mesmo homem falando espanhol. O tempo de fala e de silêncio é aproximadamente o mesmo durante os áudios. A recomendação sugere criar novos arquivos de áudio a partir da combinação dos quatro dados a fim de se obter arquivos mais extensos. Ao se usar gravações mais extensas, a chamada emulada passa a ter uma duração mais próxima de uma chamada real. Dessa forma, são usados nas chamadas emuladas quatros novos arquivos compostos de concatenações dos quatro arquivos originais em diferentes ordens, totalizando gravações de 32 segundos.

A recomendação deixa a critério do aplicador realizar os experimentos através de chamadas reais ou simuladas/emuladas. Em ambos os casos, os mesmos quatro arquivos de áudio devem ser usados e o resultado de cada rodada é a média da qualidade dos áudios. Mais de uma rodada pode ser feita para cada cenário e, neste caso, a avaliação é a média das avaliações das rodadas. Nenhum rigor estatístico é exigido pela recomendação.

Depois de definidos todos os 30 pontos de interesse, deve-se fazer a avaliação de todos os pontos novamente, agora utilizando o modelo sob teste. Com os resultados das avaliações, pode-se calcular, para cada ponto de interesse, a diferença entre a qualidade medida pelo PESQ e pelo modelo sob teste. Essa diferença é chamada de desvio de previsão e é dada por $MOS_{modelo} - MOS_{PESQ}$, onde o sinal do resultado (negativo ou positivo) é importante.

A primeira medida de comparação exigida pela P.564 é o coeficiente de correlação de Pearson entre as duas séries de avaliações, que serve para medir a dependência linear entre elas. Sabendo que as duas séries de previsões encontram-se na mesma escala (MOS) e são previsões de um mesmo fenômeno, espera-se uma correlação alta. Em seguida, deve-se classificar os pontos de interesse em dois grupos. No grupo 1 ficam os pontos nos quais $MOS_{PESQ} \geq 2,8$. No grupo 2 ficam aqueles com $MOS_{PESQ} < 2,8$, ou seja, a qualidade foi ruim ou péssima.

Para cada grupo, os desvios de previsão do modelo sob teste devem ser classificados em diferentes faixas, que servem para observar a sua distribuição. Uma faixa engloba todos os desvios cujos valores estão entre os limites inferior e superior mostrados na Tabela 3.2. Por exemplo, caso o erro da previsão de um ponto de interesse do Grupo 1 seja -0.3, aquele ponto não se enquadra na faixa 1, porém se encaixa nas faixas de 2 a 6. Também deve ser calculada dentro de cada grupo a porcentagem de pontos de interesse cujo erro de previsão se enquadra em cada faixa. Pode-se observar que

Tabela 3.2: Faixas de classificação do desvio de previsão e resultados necessários para aderência de um modelo à P.564

Medida	Limite inferior	Limite superior	Critério	
			Classe C1	Classe C2
Correlação	-	-	$> 0,90$	$> 0,85$
Falsos negativos (%)	-	-	$< 5,0$	$< 5,0$
Falsos positivos (%)	-	-	$< 3,0$	$< 3,0$
Grupo 1 ($MOS_{PESQ} \geq 2,8$)				
% de desvios na faixa 1	-0,25	+0,25	$\geq 95,0$	$\geq 75,0$
% de desvios na faixa 2	-0,35	+0,5	$\geq 97,9$	
% de desvios na faixa 3	-0,5	+0,5		$\geq 95,0$
% de desvios na faixa 4	-0,6	+1,0	$\geq 99,0$	
% de desvios na faixa 5	-1,0	+1,0		$\geq 97,9$
% de desvios na faixa 6	-1,5	+1,5		$\geq 99,0$
Grupo 2 ($MOS_{PESQ} < 2,8$)				
% de desvios na faixa 7	-0,4	+0,4	$\geq 90,0$	
% de desvios na faixa 8	-0,5	+0,5		$\geq 90,0$
% de desvios na faixa 9	-0,7	+0,7	$\geq 95,0$	
% de desvios na faixa 10	-1,0	+1,0		$\geq 95,0$
% de desvios na faixa 11	-1,2	+1,2	$\geq 99,0$	
% de desvios na faixa 12	-1,5	+1,5		$\geq 99,0$

toda vez que um ponto se enquadra numa faixa mais restritiva, ele automaticamente se enquadra em todas as outras faixas mais amplas.

Por último, devem ser calculadas as porcentagens de medições classificadas como falsos negativos e falsos positivos. Por definição, os falsos negativos são aqueles casos onde $MOS_{PESQ} < 2,0$ e $MOS_{modelo} \geq 3,0$. Analogamente, os casos de falsos positivos são aqueles onde $MOS_{PESQ} \geq 3,0$ e $MOS_{modelo} < 2,0$.

Finalmente, as medidas de comparação são checadas contra os limites na Tabela 3.2 para classificação em uma de duas classes de conformidade. A classe C1 demonstra uma maior qualidade do modelo, mas é mais difícil de ser alcançada. Já a classe C2 é mais flexível e mais fácil de ser alcançada, porém indica uma qualidade de previsão inferior.

Como exemplo, pode-se considerar a sexta linha da Tabela 3.2, que começa com “% de desvios na faixa 1.” Caso o desvio de um ponto de interesse, ou seja, a diferença entre o MOS medido pelo PESQ e o MOS medido pelo modelo sob avaliação, esteja entre o limite inferior (segunda coluna) de -0,25 e o limite superior (terceira coluna) de +0,25 e MOS do PESQ seja maior ou igual a 2,8 (critério do grupo 1), esse desvio fará parte da faixa 1. Caso 95% ou mais dos pontos de interesse do grupo 1 façam parte da faixa 1, um dos critérios para aprovação da classe C1 será satisfeito (quarta coluna). Para que o modelo seja aprovado na classe C1, uma série de outros critérios ainda devem ser satisfeitos. Já para a classe C2, o critério relativo à faixa 1 é que pelo menos 75% dos pontos de interesse façam parte dessa faixa (quinta coluna).

Na tabela, correlação e falsos positivos e negativos não possuem limites inferiores ou superiores por serem simples medidas e não faixas, mesmo assim critérios relativos a essas medidas devem ser satisfeitos. Algumas faixas não possuem critérios em uma das classes de aprovação, já que uma classe é mais exigente que a outra.

4 EMULAÇÃO DE CHAMADAS E RESULTADOS DOS TESTES DE CONFORMIDADE

Apesar de definir matematicamente como os pontos de interesse devem ser formados, a recomendação P.564 não define qualquer metodologia para encontrar os pontos. Dessa forma, procurou-se desenvolver uma metodologia algorítmica apropriada a essa tarefa, que foi implementada no emulador de chamadas. Diz-se que as chamadas são emuladas pois toda a codificação e decodificação da mídia é feita da mesma maneira como seria feita numa chamada real. Contudo, de forma diferente do que ocorre numa chamada real, a mídia codificada não é transmitida pela rede. No software, a mídia passa por um módulo responsável por emular comportamentos das redes reais como atraso, perdas e *jitter*.

Neste capítulo, além de haver uma discussão sobre as metodologias desenvolvidas para o emprego da P.564, é discutido como os fatores de degradação foram implementados e como funciona o emulador que dá suporte à aplicação em si da P.564. Por fim, também são mostrados e comentados os resultados dos testes de conformidade do Modelo E e do Modelo E Estendido à P.564.

4.1 Cálculo dos Pontos de Interesse

Como a P.564 não define uma maneira de calcular os pontos de interesse, foi escolhido usar busca binária sobre as retas para encontrar valores dos parâmetros dos fatores de degradação que implicam num MOS conforme desejado. A fim de agregar confiabilidade, foi decidido usar certo rigor estatístico nos cálculos. Dessa forma, o cálculo dos pontos de interesse e das médias do MOS do Modelo E e do PESQ nos pontos de interesse foram todos realizados e repetidos até que fosse alcançado um intervalo de confiança de 95% menor que 10% do valor absoluto da média. Para encontrar um ponto de interesse em uma reta, é feita uma busca sobre a reta variando a distância à origem. Cada ponto buscado é testado para que se descubra se é um ponto de interesse ou não, ou seja, se o seu MOS satisfaz um dos intervalos desejados. A busca só para quando tanto a média quanto todo o intervalo de confiança do MOS correspondente estão dentro do intervalo alvo de MOS que define o ponto de interesse. Uma busca é feita por reta para cada um dos 5 pontos de interesse que se deseja encontrar.

4.2 Emulação de Fatores de Degradação

Pode-se dizer que o tráfego VoIP é bem comportado. Ele manda pacotes geralmente pequenos em intervalos de tempo regulares. Porém, a concorrência do tráfego telefônico nas redes IP com outros tráfegos menos comportados pode causar degradação na qualidade de entrega de uma chamada.

É de amplo conhecimento que a principal causa da degradação das redes IP atuais é o congestionamento de roteadores em algum ponto da rede. É normal que, por alguns instantes, algum equipamento fique congestionado e atrase a entrega de

pacotes ou até mesmo descartá-los. Esse fenômeno é o principal causador tanto de perdas na internet quanto de variação no atraso de entrega de pacotes, o *jitter*. Como os roteadores costumam ficar congestionados somente por um breve período de tempo, tanto as perdas quanto o *jitter* acabam tendo uma tendência de acontecer ao mesmo tempo e concentrados nos mesmos intervalos de tempo que acontecem os congestionamentos.

Apesar de a P.564 definir quatro tipos diferentes de fatores de degradação, foi escolhido usar apenas a perda aleatória e o *jitter*. Os outros dois fatores de degradação envolvem largura de banda disponível e em uso. Foi desconsiderada a ocupação da banda, pois seu efeito indireto é sentido na perda e no *jitter* que ocorrem em gargalos.

Também foi adotado um modelo de perdas em rajada, característica comum nas redes reais, para os testes de conformidade. Mesmo fora do âmbito da P.564, emulações de chamadas com esse fator de degradação servem para ampliar os estudos de precisão dos modelos de medição de qualidade.

4.2.1 Jitter

O parâmetro do *jitter* é a variação máxima que um pacote pode sofrer em relação ao atraso do restante do período de fala. Assim, o parâmetro deste fator de degradação não é o valor do *jitter* propriamente dito, mas o desvio máximo que o momento da entrega de um pacote pode sofrer em relação ao momento esperado de sua entrega, medida conhecida como *peak-to-peak jitter*.

Esse fator é emulado sorteando um atraso aleatório no intervalo $[0-j]$, onde j é o parâmetro do *jitter*. O Algoritmo 1 mostra o procedimento usado: a cada pacote

de um fluxo de mídia, é sorteado um atraso qualquer entre 0 e j , o parâmetro deste fator de degradação. A emulação do *jitter* pode causar reordenação de pacotes, fenômeno que só pode ser desfeito pelo uso de um *buffer* de compensação de *jitter*.

Dados: Vetor v de pacotes de voz; *jitter peak-to-peak* j

Resultado: Vetor v modificado após sofrer *jitter*

$l \leftarrow |v|;$

para $i \leftarrow 1$ **até** l **faça**

 //*rand()* retorna um número real no intervalo $[0,1[$;

$v[i].atraso \leftarrow rand() \cdot j;$

fim

Algoritmo 1: Algoritmo de emulação de *jitter*

4.2.2 Perdas Aleatórias

A perda aleatória se caracteriza pelo fato de os eventos de perda serem independente entre si, fazendo com que as perdas possam ocorrer de forma isolada e independente. A interpretação de tal fenômeno pode levar a duas formas distintas de emulá-las. A primeira forma assume que cada pacote de um fluxo na rede tem uma certa probabilidade de ser perdido, usualmente igual para todos os pacotes de um mesmo fluxo. Tal algoritmo não pode garantir que a porcentagem de pacotes perdidos será fixa, uma vez que cada evento de perda é aleatório e independente.

Contudo, seguindo o texto da recomendação P.564, as perdas aleatórias foram implementadas descartando uma porcentagem fixa dos pacotes do fluxo igual ao parâmetro do fator de degradação é descartada. Os pacotes podem ser sorteados de qualquer parte do fluxo, tanto do início, meio ou fim, sem nenhuma ordem particular. O parâmetro do fator de degradação de perdas aleatórias é, neste caso, a porcentagem de pacotes perdidos. Tendo o total de pacotes de uma chamada, escolhe-se aleatoriamente os descartes para obter a porcentagem definida, garantindo a fração desejada de pacotes perdidos e também ocorrências de perda distribuídas uniformemente pelo

fluxo.

O Algoritmo 2 descreve como uma fração f , um número real entre 0 e 1 que representa a porcentagem e quais pacotes serão de fato perdidos, é usada para calcular a quantidade de pacotes a serem perdidos. No algoritmo, tal quantidade de pacotes a descartar é representada pela variável x . A cada pacote descartado com sucesso pelo algoritmo, x é decrementado de uma unidade. Quando x chega a zero, o algoritmo chega ao fim.

Dados: Vetor v de pacotes de voz; Fração $f \in [0, 1[$ de pacotes de pacotes a serem perdidos

Resultado: Vetor v modificado após sofrer perdas aleatórias

$l \leftarrow |v|;$

$x \leftarrow l \cdot f;$

enquanto $x > 0$ **faça**

$i \leftarrow rand() \cdot l;$

se $v[i].status = OK$ **então**

$V[i].status = PERDIDO;$

$x \leftarrow x - 1;$

fim

fim

Algoritmo 2: Algoritmo de emulação de perdas aleatórias

4.2.3 Perdas em Rajada

Somente perdas aleatórias são contempladas na recomendação P.564. Sabe-se, porém, que as perdas na internet têm tendência a acontecer em rajada (ITU-T, 2003c). Dessa forma, mesmo estando fora do escopo dos testes de conformidade da P.564, a medição da precisão de modelos em situações de perdas em rajada agrega valor aos testes ao usar cenários mais naturais às redes reais que simples perdas aleatórias.

Diversos modelos foram desenvolvidos com o objetivo de emular o comportamento

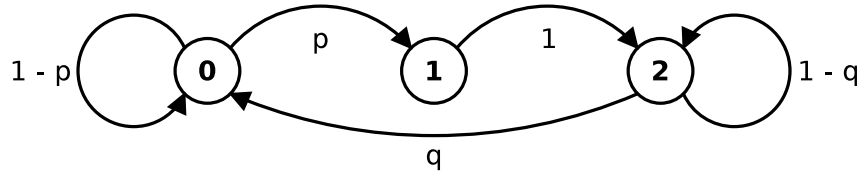


Figura 4.1: Cadeia de Markov de modelo de perdas em rajada de tamanho > 1

das perdas da internet. Alguns são baseados em cadeias de Markov simples, como o modelo de Gilbert (GILBERT, 1960) e outros envolvem análise Bayesiana de históricos de rede (ROSSI et al., 2003). Para testar o Modelo E frente a perdas em rajada, foi decidido usar uma CM que sempre forçasse perdas em rajada, ou seja, não há perdas isoladas. A Figura 4.1 mostra a CM de três estados capaz de emular rajadas de perda. O primeiro estado, estado 0, é o estado de SUCESSO. Somente neste estado um pacote é entregue normalmente. Os estados 1 e 2 são ambos estados de PERDA. O primeiro destes estado marca o início de uma rajada, e a transição para o estado 2 marca a segunda perda obrigatória da rajada. A partir do estado 2, a cadeia pode continuar gerando perdas e, por consequência, aumentar o tamanho da rajada, ou pode fazer com que a rajada termine e o próximo pacote seja entregue com sucesso ao transitar para o estado 0.

Ao resolver o estado estacionário da cadeia, chega-se ao seguinte conjunto de equações:

$$\pi_0 = \frac{q}{p + pq + q} \quad ; \quad \pi_1 = \frac{pq}{p + pq + q} \quad ; \quad \pi_2 = \frac{p}{p + pq + q} \quad (4.1)$$

Dessa forma, sabendo que $\pi_{perda} = \pi_1 + \pi_2$, tem-se que:

$$\Pi_{perda} = \frac{p + pq}{p + pq + q} \quad (4.2)$$

Ressaltando que não pode haver perdas isoladas, *i.e.* tamanho de rajada igual a 1, a probabilidade de que uma rajada de perda tenha tamanho k é dada pela distribuição

geométrica p_k :

$$\begin{aligned} p_1 &= 0 \\ p_k &= (1 - q)^{k-2} \cdot q, \quad k=2, 3, 4... \end{aligned} \quad (4.3)$$

E a partir da distribuição, conclui-se que a esperança do tamanho de uma rajada é dada por:

$$E[k] = \sum_{k=2}^{\infty} p_k \times k = 1 + \frac{1}{q} \quad (4.4)$$

Para os testes de conformidade manterem o esquema de calcular os pontos de interesse, deve-se manter o uso de dois parâmetros de fatores de degradação. Dado que a cadeia recém apresentada somente pode ser descrita com o uso de dois parâmetros, não é possível utilizar o esquema para emular perdas em rajada e para calcular pontos de interesse no âmbito da P.564 da mesma maneira que são usados as perdas aleatórias e o *jitter*. Uma maneira de adaptar o seu uso aos testes de conformidade da P.564 é usar as perdas em rajada com único fator de degradação na emulação e calcular, para cada *codec* e *ptime*, pontos de interesse num plano composto pelos seus dois parâmetros.

Foi escolhido usar os parâmetros p e q para definir os pontos de interesse para cenários com perdas em rajada. Deve-se notar, porém, dois aspectos relevantes ao emprego dessa alternativa. O primeiro é que a definição dos pontos de interesse assume que quanto maior o parâmetro de um fator de degradação, menor a qualidade. Sabendo que a diminuição da qualidade é provocada pelo aumento da variável q e, consequentemente, pela diminuição de $(1 - q)$, utiliza-se o fator $(1 - q)$ como parâmetro de modo a manter a compatibilidade com as fórmulas do teste de conformidade da P.564. Dessa forma, para manter a compatibilidade com as fórmulas dos testes de conformidade da P.564, deve-se usar $(1 - q)$ como parâmetro.

Outra particularidade é o fato que a reta definida para $p = 0$ não tem validade. Isso ocorre porque, seja qual for o valor de q , não haverá qualquer chance de ocorrer perda de pacote. Dessa forma, deve-se substituir toda a reta por um único ponto de interesse, por simplicidade, $p = q = 0$, no qual se espera MOS máximo de qualquer modelo de medição de qualidade.

As rajadas de perdas são emuladas usando o Algoritmo 3. O algoritmo recebe as probabilidades de transição p e q descritas anteriormente. A variável *estado* guarda o estado no qual o algoritmo se encontra: 0, 1 ou 2. O algoritmo começa sempre no estado 0 e a cada pacote do fluxo de voz um número real entre 0 e 1 é sorteado para decidir a transição de estados. Quando ocorre uma transição para o estado 1 ou 2, um pacote é perdido.

4.3 VQSim: Emulador de Chamadas VoIP para Testes de Qualidade

Os experimentos necessários à aplicação da P.564 foram realizados com um emulador criado especificamente para esta tarefa. Ele foi escrito em C++ a fim de utilizar a facilidade da orientação a objetos e, ao mesmo tempo, usar a biblioteca PJSIP (PRIJONO, 2011), escrita em C, para o tratamento da mídia. A PJSIP é usada apenas para codificar/decodificar a mídia e para emular o *buffer* de compensação de *jitter*, cujo tamanho é modificado dinamicamente conforme a variação do *jitter*. O algoritmo de PLC usado pela PJSIP é conhecido como *Waveform Similarity Based Overlap-Add* (VERHELST; ROELANDS, 1993).

O software foi criado seguindo as indicações da P.564 de como chamadas devem ser emuladas. Mesmo que a medição de qualidade não seja feita em tempo real, a maneira como a VQuality é usada para avaliar a qualidade das chamadas emuladas

Dados: Vetor v de pacotes de voz; Probabilidades p e q de transição na cadeia de Markov

Resultado: Vetor v modificado após sofrer perdas em rajada

```

 $l \leftarrow |v|;$ 
 $estado \leftarrow 0;$ 
para  $i \leftarrow 1$  até  $l$  faça
     $x \leftarrow rand();$ 
    selecione  $estado$  faça
        caso 0
            se  $x < p$  então
                 $v[i].status \leftarrow PERDIDO;$ 
                 $estado \leftarrow 1;$ 
            fim
        caso 1
             $v[i].status \leftarrow PERDIDO;$ 
             $estado \leftarrow 2;$ 
        caso 2
            se  $x < q$  então
                 $estado \leftarrow 0;$ 
            senão
                 $v[i].status \leftarrow PERDIDO;$ 
            fim
    fim
fim

```

Algoritmo 3: Algoritmo de emulação de perdas em rajada

é exatamente igual àquela usada para avaliar chamadas reais, fazendo com que os testes sejam equivalentes. Também é importante ressaltar que, apesar de a VQuality somente avaliar a qualidade ao final de uma chamada do emulador, ela pode avaliar a qualidade de chamadas reais a qualquer instante, sendo possível acompanhar a evolução da qualidade da chamada com o passar do tempo.

O emulador é composto de uma série de componentes que trocam fluxos de áudio entre si, como pode ser visto no diagrama de componentes da Figura 4.2. No diagrama de componentes, cada componente de um software é representado por uma caixa que possui interfaces de entrada e de saída, chamadas conectores ou *sockets*.

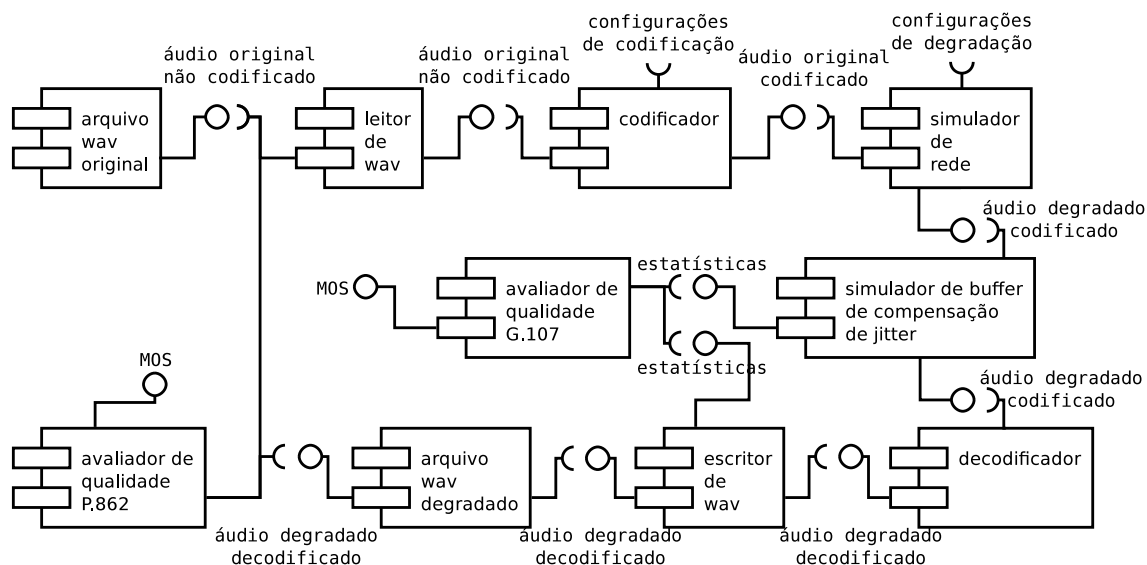


Figura 4.2: Diagrama de componentes do emulador de chamadas

Próximo a um conector está escrito o nome do dado que ele representa. O conector com um círculo na ponta indica que aquele componente oferece um tipo de dado para ser usado por outro componente. De forma análoga, o conector com a ponta aberta significa que aquele componente precisa daquele dado como entrada.

Os arquivos .wav usados, apesar de serem componentes do sistema, são partes passivas e são vistos na figura apenas como dependências dos componentes que precisam ler ou escrever neles. Os componentes do software trabalham de maneira completamente serial, de forma que, em cada passo da emulação, um componente processa todos os quadros de voz da chamada antes de passá-los ao componente seguinte.

A emulação começa com o **leitor de wav** lendo o **arquivo wav original**. O leitor lê a mídia e a transforma num vetor de pacotes de voz com informações como *ptime*, *timestamp* etc. Esse vetor é passado ao **codificador**, que é responsável por codificar cada quadro do vetor que recebe e transformá-los num outro vetor com a mídia completamente codificada. No **codificador**, o processo de codificação já

representa uma pequena degradação de qualidade, já que o áudio perde algumas características ao ser comprimido pelo *codec*.

Depois de codificado, o fluxo é passado ao **emulador de rede** que, conforme as configurações de emulação, emula fatores de degradação como *jitter* e perdas aleatórias. Esses fatores de degradação têm acesso aos pacotes do vetor e podem modificar campos referentes ao atraso dos pacotes ou marcá-los como perdidos. Em seguida, os pacotes passam pelo **emulador de buffer de compensação de jitter**, componente responsável por considerar o atraso de cada pacote e garantir que os quadros de voz sejam tocados em ordem e em sincronia. Esse componente é responsável por transformar o *jitter* em perdas e atrasos, dados estatísticos que são passados para o **avaliador de qualidade G.107**, componente responsável por medir a qualidade da chamada usando o Modelo E ou o Modelo E Estendido através da VQuality.

Finalmente, os pacotes chegam ao **decodificador** que decodifica os quadros conforme passados pelo *buffer* de compensação de *jitter*. Este componente é o responsável por empregar o PLC, caso esse recurso esteja ativado. Caso não esteja, o decodificador substitui quadros perdidos por silêncio. O vetor contendo os quadros decodificados é em seguida passado para o **escritor de wav** que simplesmente escreve os quadros num novo arquivo, temporário, representado pelo componente **arquivo wav degradado** no diagrama da Figura 4.2. Este componente emula a tarefa de tocar o áudio para o usuário final da chamada, de forma que estatísticas necessárias para o funcionamento do **avaliador de qualidade G.107**, como pacotes perdidos ou a distribuição de tais perdas, se encontram neste componente. Por fim, o **avaliador de qualidade P.862** é capaz de comparar o **arquivo wav original** ao **arquivo wav degradado** a fim de gerar uma métrica de qualidade para a chamada usando o PESQ. Assim, o PESQ estará comparando o áudio que seria efetivamente ouvido com o áudio original.

O emulador também é capaz de realizar as buscas necessárias e calcular, por conta própria, os pontos de interesse para cada *codec* e *ptime*, além de fornecer, ao final, um relatório de conformidade e aderência do Modelo E e do Modelo E Estendido à P.564.

4.4 Nova Versão da VQuality

Apesar do amplo uso da biblioteca VQuality em trabalhos anteriores, o software apresentava alguns problemas e inconsistências que foram notadas durante a execução do presente trabalho. Um exemplo era a noção de que se deveria considerar apenas metade do atraso gerado pelo *buffer* de compensação de *jitter*. Esse equívoco provavelmente ocorreu devido ao trecho abaixo da recomendação G.114 (ITU-T, 2003), que trata de como medir o atraso de transmissão em redes telefônicas e faz parte da série de recomendações que acompanham o Modelo E (grifos no original):

“For PLANNING PURPOSES, IT IS RECOMMENDED to assume that a de-jitter buffer adds one half of its peak delay to the mean network delay.”

Outro problema da primeira versão da biblioteca era que o módulo responsável pelo uso do Modelo E original não era acessível, ou seja, apesar de implementado não podia ser usado. Além disso, a biblioteca não considerava o atraso do empacotamento da mídia, ou seja, o atraso gerado pelo uso de determinado *ptime*. Por exemplo, para que sejam usados pacotes contendo 20 ms de mídia, deve-se esperar que 20 ms de áudio sejam capturados, gerando um atraso igual ao *ptime*. Como esse atraso de empacotamento era desconsiderado, a estimativa da degradação devida ao atraso era sempre ligeiramente otimista, o que poderia causar um erro de precisão prin-

principalmente quando *ptimes* maiores são usados e quando o atraso da rede eleva o atraso total próximo a 100 ms, valor que faz com que o Modelo E penalize de forma mais dura a degradação devido ao atraso.

A fim de resolver tais inconsistências e com o objetivo de definir uma interface mais simples para o usuário, a biblioteca foi totalmente reimplementada. A reimplementação dos módulos relativos à geração de CDRs, porém, foi deixada para uma futura revisão do software.

A antiga versão da VQuality diferenciava os quadros perdidos na rede dos quadros de voz perdidos no *buffer* de compensação de *jitter*, mesmo que, na prática, os dois eventos tivessem o mesmo impacto sobre a qualidade percebida pelo usuário. Esse fato gerava um problema de contabilizar a perda de um quadro duas vezes, já que um quadro perdido na rede também poderia ser contabilizado como perdido no *buffer* de compensação de *jitter*, pois sua ausência era percebida pelo *buffer* e reportada à VQuality. Mudando a filosofia da biblioteca, o software agora não distingue onde o pacote foi perdido, e sim os quadros que são tocados com sucesso ao usuário final e os que estão faltando no momento de serem tocados. Isso elimina o problema de ter que distinguir pacotes perdidos de quadros perdidos, já que um pacote pode conter mais de um quadro, e distinguir quadros perdidos na rede dos descartados no *buffer* de compensação de *jitter*.

Mesmo contabilizando as perdas da rede e do *buffer* de forma única no modelo de qualidade, a VQuality ainda é capaz de manter estatísticas sobre as duas taxas de perda. A diferenciação em quadro perdido na rede e quadro perdido no *buffer* de compensação de *jitter* é importante para permitir o uso da VQuality como ferramenta de apoio à monitoração e gerência de qualidade, com a indicação de possível ponto de falha ocorrendo.

A nova filosofia da VQuality é que não importa onde houve perda de pacotes ou atraso, mas sim o seu impacto sobre a qualidade percebida pelo usuário final, sendo a base de eventos da nova versão se os quadros são tocados ou não.

A fim de facilitar o trabalho de integração da biblioteca a um software VoIP, foi criada uma classe para servir de interface aos dois modelos suportados, a classe **VQualityWrapper**. Ela oferece os seguintes métodos para serem usados para a integração:

```
VQualityWrapper(ModelType rModelType );
void OnFramePlay(void );
void OnFrameMiss(void );
void OnNetworkDelayChange(double rNetworkDelayMsecs );
void OnJitterBufferResize(double rJitterBufferSizeMsecs );
void OnCodecChange(CodecType rCodecType );
void OnFrameSizeChange(int rFrameSizeMsecs );
void OnFramesPerPacketChange(int rFramesPerPacket );
double GetRFactor(void );
double GetMos(void );
```

O construtor recebe como argumento o tipo de modelo de medição de qualidade que deve ser usado, o Modelo E original ou o Estendido. Estão disponíveis sete métodos para que as entradas necessárias sejam passadas à VQuality no formato de eventos. A qualquer momento durante uma chamada, pode ser pedido à classe o MOS ou o fator R medidos pela VQuality.

4.5 Aplicação da P.564 ao Modelo E

Em relação a sua arquitetura, o Modelo E implementado pela VQuality se enquadra no modo de operação embutido da P.564, obtendo informações passadas pelo software do cliente final. O cliente repassa à VQuality dados sobre os diferentes tipos

de atraso que a mídia sofre (atraso da rede, do *buffer* de compensação de *jitter* e de codificação) e eventos de quando a mídia é tocada ou perdida.

Tanto o Modelo E original da ITU-T quanto o Modelo E Estendido por Clark foram submetidos ao teste de conformidade da P.564. Como o uso de ambos é idêntico, as particularidades aqui descritas sobre a aplicação da P.564 servem para os dois modelos.

Mesmo que a P.564 não permita que sejam usados parâmetros de atraso no modelo sob teste, ele permite explicitamente acesso e uso dos dados do *buffer* de compensação de *jitter* incluindo perdas e atraso. Isso faz com que o I_d do Modelo E seja usado mas resumido ao atraso do *buffer*. Todos os outros termos do cálculo do fator R continuam sendo usados da mesma maneira.

O suporte do Modelo E a diferentes *codecs* se dá com o uso de algumas constantes em suas fórmulas. Os *codecs* suportados oficialmente pelo Modelo E são o G.711, G.723, G.728, G.729 e o GSM. *Codecs* mais modernos como Speex e iLBC não são ainda suportados.

Os *codecs* suportados pela versão *open source* da biblioteca PJSIP são G.711, G.722, GSM, iLBC e Speex. Como outros *codecs* proprietários exigem alguma espécie de licença, sua codificação e decodificação não são feitas pela biblioteca. Por outro lado, caso seja instalada uma segunda biblioteca com suporte a outros *codecs*, o PJSIP pode usá-la para oferecer os *codecs* proprietários através da sua interface.

Tendo em vista as limitações do Modelo E e da biblioteca em relação à oferta de *codecs*, os testes se limitaram à avaliação da conformidade do Modelo E à recomendação P.564 apenas com os *codecs* G.711 μ -law e GSM. Os testes foram feitos para diferentes valores de *ptime*, variando de 10 a 100 ms, com e sem PLC e com e sem

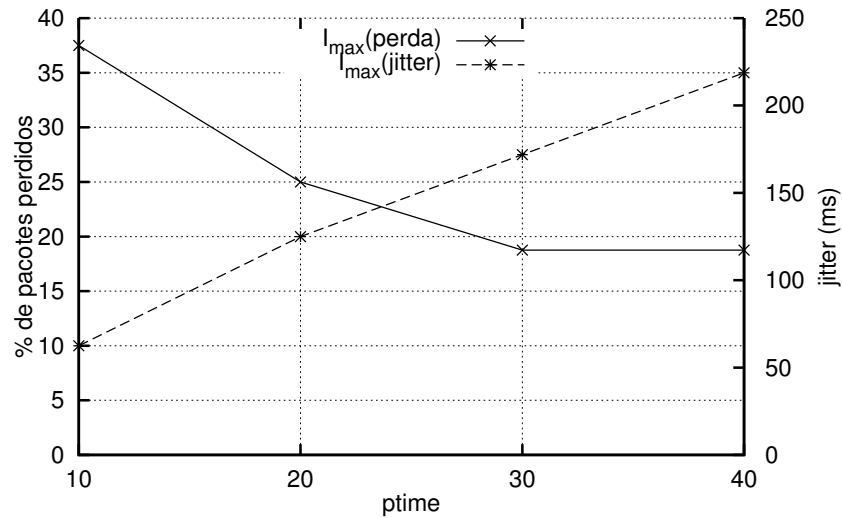


Figura 4.3: I_{max} da perda e do *jitter* para diferentes *ptimes*

VAD.

Os testes também serviram para demonstrar que diferentes valores de *ptime* em uso significam diferentes efeitos da perda e do *jitter* sobre a qualidade. Definindo como I_{max} o valor máximo do parâmetro do fator de degradação tal como usado anteriormente, a Figura 4.3 mostra que, conforme o *ptime* aumenta, menores índices de perda e maiores de *jitter* são necessários para degradar o MOS medido pelo PESQ ao intervalo $[1,3-1,5]$. Enquanto, em média, é necessário uma taxa de perda aleatória de quase 40% para diminuir o MOS ao intervalo anterior quando se usa *ptime* de 10 ms, uma taxa próxima de 20% é capaz de fazer o mesmo quando um *ptime* de 40 ms é empregado. Já quando se degrada uma chamada usando apenas *jitter*, uma média um pouco acima de 50 ms deste fator de degradação é suficiente para diminuir bastante a qualidade de uma chamada que usa um *ptime* de 10 ms, enquanto somente um *jitter* de mais de 200 ms consegue para provocar o mesmo efeito quando se usa um *ptime* de 40 ms. Assim, ao se analisar a evolução dos valores de I_{max} conforme se aumenta o *ptime*, percebe-se que *ptimes* menores podem ser mais indicados para redes com mais perda e *ptimes* maiores para redes com mais *jitter*.

Tabela 4.1: Resultado dos testes de conformidade para o G.711 sem VAD nem PLC

Medida	Modelo E					Modelo E Estendido				
	<i>ptime</i>					<i>ptime</i>				
	10	20	30	40	50	10	20	30	40	50
Correlação	0,99	0,98	0,97	0,98	0,97	0,98	0,98	0,98	0,98	0,98
Falsos negativos (%)	0	0	0	0	0	0	0	0	0	0
Falsos positivos (%)	0	0	0	0	0	0	0	0	0	0
Grupo 1 ($MOS_{PESQ} \geq 2,8$)										
% de desvios na faixa 1	78,2	94,7	82,6	65,2	87,5	73,9	94,7	82,6	82,6	95,8
% de desvios na faixa 2	82,6	100	95,6	100	100	78,2	100	100	95,6	100
% de desvios na faixa 3	91,3	100	100	100	100	78,2	100	100	100	100
% de desvios na faixa 4	91,3	100	100	100	100	78,2	100	100	100	100
% de desvios na faixa 5	100	100	100	100	100	100	100	100	100	100
% de desvios na faixa 6	100	100	100	100	100	100	100	100	100	100
Grupo 2 ($MOS_{PESQ} < 2,8$)										
% de desvios na faixa 7	14,2	90,9	71,4	71,4	83,3	28,5	72,7	100	85,7	83,3
% de desvios na faixa 8	71,4	100	85,7	100	83,3	71,4	81,8	100	100	100
% de desvios na faixa 9	85,7	100	85,7	100	100	100	100	100	100	100
% de desvios na faixa 10	100	100	100	100	100	100	100	100	100	100
% de desvios na faixa 11	100	100	100	100	100	100	100	100	100	100
% de desvios na faixa 12	100	100	100	100	100	100	100	100	100	100
Resultado (C2)	✗	✓	✗	✗	✗	✗	✗	✓	✓	✓

4.6 Conformidade à P.564

Depois de serem realizadas todas as emulações para os pontos de interesse de todas as combinações de *codec*, *ptime* e uso de VAD e PLC, sempre com rigor estatístico, foram montadas tabelas como a Tabela 4.1, que representam o relatório final de teste de conformidade para os Modelos E original e E Estendido para G.711 μ -law sem PLC e sem VAD. O relatório de conformidade para um cenário deve incluir todos os parâmetros a serem comparados com os requisitos mínimos para aprovação, ou seja, o coeficiente de correlação de Pearson dos testes, a porcentagem de falsos positivos e negativos e a porcentagem de desvios de previsão pertencentes a cada uma das 12 faixas anteriormente definidas. Quando usado o Modelo E original nos testes de conformidade do *ptime* de 20 ms, por exemplo, o modelo foi aprovado no teste por seu bom desempenho. A correlação chegou ao alto valor de 0,98 e poucos pontos de interesse tiveram desvios de previsão significativos, com apenas 5,3% dos casos

do Grupo 1 com desvio maior que $+0,25$ ou menor que $-0,25$. Os casos restantes, ou 94,7% dos casos, do Grupo 1 se mantiveram dentro da faixa 1, enquanto 90,9 os casos do Grupo 2 se mantiveram dentro da melhor faixa desse grupo, a faixa 7.

Pode-se observar na Tabela 4.1 a distribuição dos desvios para cinco *ptimes* diferentes. As células hachuradas indicam os valores medidos que fizeram com que a configuração fosse reprovada nos critérios da classe C2. Nota-se que, entre os casos nos quais o Modelo E original foi testado, somente os casos com *ptime* de 20 ms conseguiu resultados bons suficientes para ser aprovados pelos critérios de C2. Os cenário de *ptime* 30, 40 e 50 ms medidos pelo Modelo E Estendido também passaram nos testes de conformidade. Nenhum caso foi aprovado para a classe C1.

Na Tabela 4.1, há uma aparente recuperação do desempenho do Modelo E para os casos com *ptime* de 50 ms, tanto que o Modelo E Estendido foi aprovado com classe C2 no cenário com tal *ptime*. Esse fenômeno, porém, parece ser causado por imprecisão do PESQ quando quadros maiores são perdidos e substituídos por silêncio. O modelo de avaliação do PESQ parece ter problemas em medir a qualidade em casos com grandes períodos de silêncio quando não é usado o PLC.

Pennock (PENNOCK; HOLUB; SMID, 2002) já havia relatado resultados ruins do PESQ em casos de perdas em rajada, porém sem discriminar o valor de *ptime* utilizado. O mesmo trabalho indica que a qualidade das avaliações do PESQ diminui conforme aumenta o *ptime* e que a baixa precisão com perdas em rajada ocorre também com *ptimes* grandes. Pode-se observar nesse trabalho que, quando *ptime* pequenos (10 ms, 20 ms ou 30 ms) são usados, uma perda em rajada pode ser vista como equivalente a uma perda isolada de um *ptime* grande.

Uma possível justificativa para tal comportamento é a dinâmica entre o tamanho do *ptime* e da elocução empregada pelo PESQ, que é de 32 ms. A perda de um pa-

cote com $ptime$ maior que 32 ms significa pelo menos uma elocução inteira perdida. Com o aumento do $ptime$, mais elocuições consecutivas são inteiramente perdidas, causando uma perturbação no sinal que força o algoritmo de alinhamento temporal do PESQ tentar realinhar a fala com maior frequência, piorando a precisão da avaliação de qualidade. Apesar de ser normal o PESQ realinhar quando encontra uma grande diferença entre o sinal original e o degradado, quando um intervalo significativo de áudio é substituído por silêncio, o realinhamento pode não ter condições de melhorar a previsão. Mesmo que a P.564 não leve isso em consideração, a consequência de tais problemas com o PESQ é que os resultados de conformidade para $ptimes$ maiores que 40 ms são progressivamente menos confiáveis.

Uma forma de facilitar o diagnóstico dos problemas de precisão dos modelos sob teste é acompanhar sua precisão visualmente. Dessa forma, foi preparada uma série de gráficos que refletem a diferença entre a medição de qualidade dos modelos e o PESQ para cada um dos dois modelos e para os três menores $ptimes$.

A Figura 4.4 mostra graficamente como ocorre a distribuição dos desvios de previsão dos modelos em diferentes níveis de qualidade medidos pelo PESQ. Os gráficos mostram a diferença entre o modelo em questão e para $ptimes$ de 10, 20 e 30 ms. Os segmentos de retas partindo do eixo das abscissas representam o desvio da previsão: $(MOS_{modelo} - MOS_{PESQ})$. Assim, um segmento de reta voltado para cima significa que o MOS previsto pelo modelo em questão foi maior que aquele medido pelo PESQ, ou seja, o modelo foi otimista. Analogamente, quando o segmento está voltado para baixo, o modelo sob teste foi pessimista. A distribuição do erro nas figuras ajuda a entender se o modelo de avaliação é pessimista ou otimista em relação ao PESQ, além de permitir que se diferencie o comportamento dos modelos em casos de MOS baixo ou alto.

Mais uma vez, o comportamento observado do Modelo E original é muito parecido

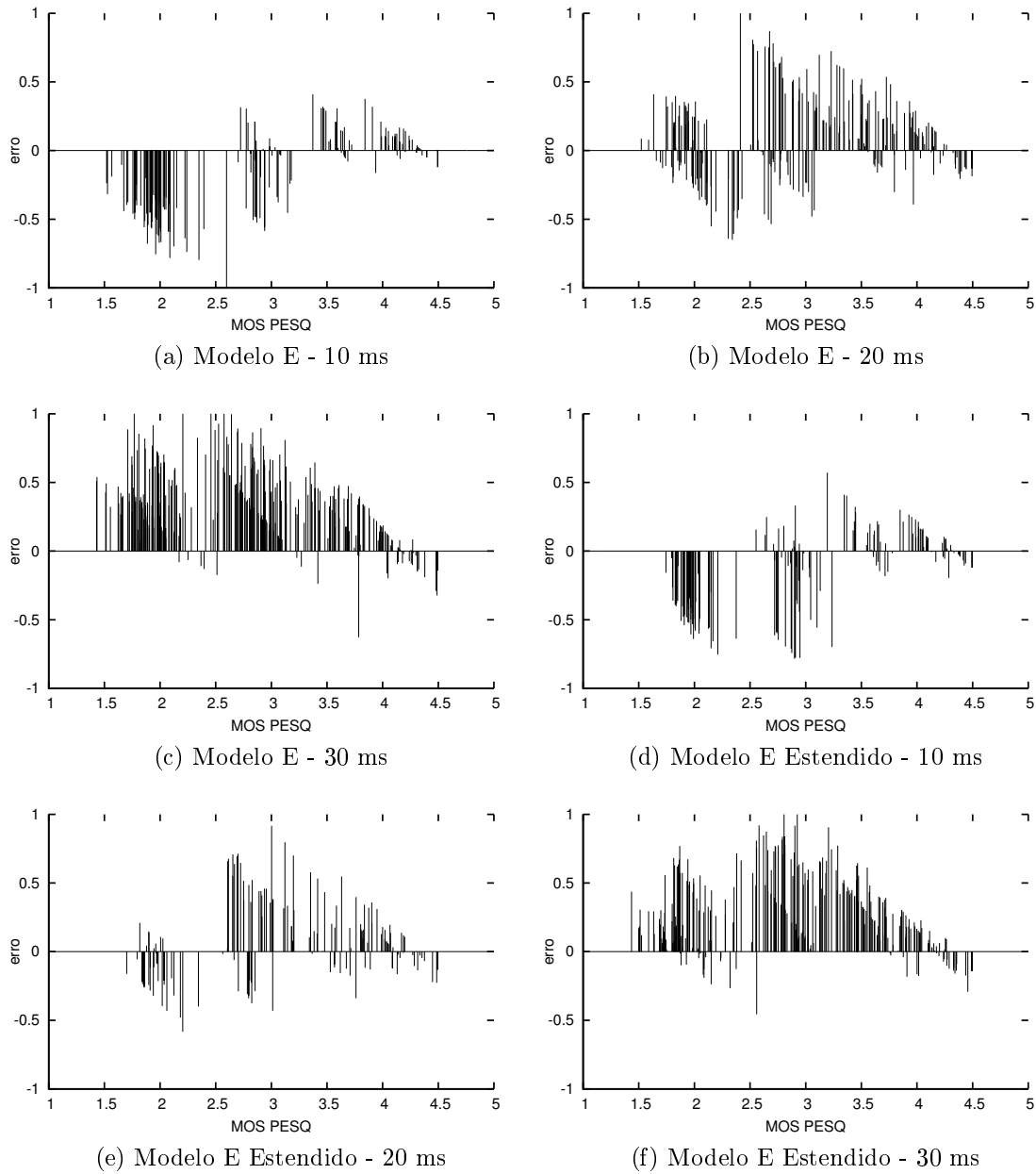


Figura 4.4: Distribuição dos desvios de previsão do Modelo E original e Estendido quando se usa G.711 sem PLC e sem VAD

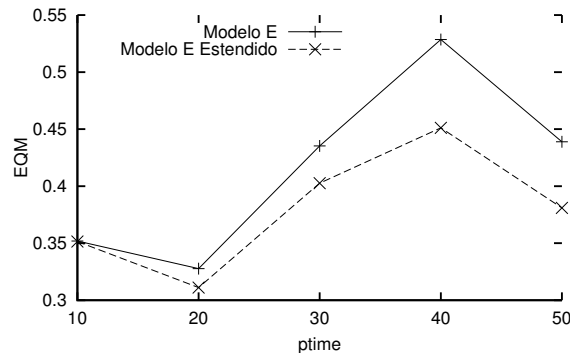


Figura 4.5: Erro quadrático médio de todas as emulações nos pontos de interesse sem VAD e PLC para o Modelo E original e o Estendido com *codec* G.711 μ -law

com o comportamento do Modelo E Estendido. No caso de *ptime* 10 ms, o erro dos gráficos indica que ambos os modelos são otimistas em cenários de MOS alto, aproximadamente maiores que 3,5, e muito pessimistas para cenários de MOS baixo. Nos cenários que usam *ptime* 20 ms, porém, há uma maior variação entre pessimismo e otimismo, prevalecendo o otimismo. A menor quantidade de desvios pequenos, traços baixos na figura, reflete a melhor precisão dos modelos no caso de 20 ms, único que passou nos testes de conformidade da P.564. Finalmente, no caso de 30 ms, os dois modelos são extremamente otimistas. Esse comportamento pode ser causado pela subestimação do poder das perdas ou por um pessimismo por parte do PESQ. Como já foi indicado antes através da análise da Figura 4.3, *ptimes* maiores estão mais suscetíveis a perda de qualidade quando há perda de pacotes. Logo, foram encontradas evidências que tanto o Modelo E original quanto o estendido não foram planejados para o uso com *ptimes* diferentes de 20 ms.

A diferença entre o desempenho do Modelos E e do Estendido pode ser medida através da análise do erro quadrático médio (EQM) obtido dos diferentes cenários usados. A Figura 4.5 mostra os resultados dos EQM medidos para todas as emulações feitas em cada um dos pontos de interesse, sem VAD e sem PLC. É visível como no caso de *ptime* de 20 ms o erro é menor que nos outros casos. A diminuição do

EQM ao ir de *p_{time}* 40 ms para 50 ms mostra mais uma vez um descompasso pela imprecisão do PESQ, já que contraria a tendência do aumento de erro de previsão conforme o *p_{time}* se afasta de 20 ms.

Enquanto o PJSIP consegue codificar qualquer *p_{time}* múltiplo de 10 ms no caso do G.711, para o GSM só é possível usar *p_{times}* múltiplos de 20 ms. As emulações foram feitas com *p_{times}* até 40 ms devido à desconfiança dos resultados do PESQ para *p_{times}* maiores.

As Tabelas 4.2, 4.3, 4.4 e 4.5 resumem os resultados de conformidade de todos os cenários propostos usando os *codecs* G.711 e GSM. Os relatórios referentes aos testes de conformidade dessas tabelas podem ser encontrados na seção A.1 do Apêndice A.

Em relação ao desempenho do G.711, a Tabela 4.2 mostra que o Modelo E somente foi aprovado com uso de perdas aleatória e *jitter* quando foi usado um *p_{time}* de 20 ms. Já a Tabela 4.3 mostra que o Modelo E não passou em nenhum teste de conformidade quando sob efeito de rajadas de perdas. Tal resultado sugere que o Modelo E foi projetado e testado usando o *p_{time}* de 20 ms, sem PLC e sem VAD. Parece razoável assumir que a questão da variação do *p_{time}* foi deixada de lado, já que os textos do Modelo E e do Modelo E Estendido não tratam do tema. De fato, mesmo que os resultados do G.113 não sejam conclusivos sobre os valores de I_e e B_{pl} a serem usados pelo Modelo E, na recomendação constam vários valores calculados para diferentes *p_{times}*, a grande maioria usando 20 ms.

Já o Modelo E Estendido conseguiu passar em diferentes testes de conformidade, todos eles sem PLC habilitado. Nos testes com perda aleatória e *jitter*, na Tabela 4.2, o Modelo E Estendido passou em 4 dos 16 cenários testados, sempre na classe C2. Naqueles cenários com rajadas de perda, Tabela 4.3, o modelo estendido conse-

Tabela 4.2: Resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas aleatórias e *jitter*

<i>p</i> time	G.711 μ-law – Perdas aleatórias e <i>jitter</i>							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
10	-	-	-	-	-	-	-	-
20	C2	-	-	-	-	-	-	-
30	-	-	-	-	C2	-	C2	-
40	-	-	-	-	C2	-	C2	-

Tabela 4.3: Resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas em rajada

<i>p</i> time	G.711 μ-law – Perdas em rajada							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
10	-	-	-	-	-	-	-	-
20	-	-	-	-	C1	-	C2	-
30	-	-	-	-	C1	-	C1	-
40	-	-	-	-	C1	-	C1	-

Tabela 4.4: Resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas aleatórias e *jitter*

<i>p</i> time	GSM – Perdas aleatórias e <i>jitter</i>							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
20	C2	C1	C1	C1	C1	C1	C1	C1
40	-	C2	C1	C1	-	C1	-	C1

Tabela 4.5: Resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas em rajada

<i>p</i> time	GSM – Perdas em rajada							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
20	C2	C1	C1	C1	C1	C1	C1	C1
40	-	C2	C1	C1	-	C1	-	C1

guiu ser aprovado em mais casos, 6 dos 16, e com classificações melhores dentro da classe C1. Tudo isso mostra uma capacidade significativamente melhor do Modelo E Estendido em medir a qualidade mesmo em cenários sem perdas em rajada.

A Figura 4.3 mostra como o I_{max} da perda aleatória diminui conforme a taxa de perda aumenta, o que indica que $ptime$ menores são mais resistentes à perda. Esse fato aliado à observação que, na Figura 4.4, o Modelo E é pessimista para o $ptime$ 10 ms e otimista para o $ptime$ 30 ms serve como a maior indicação que os valores de I_e e Bpl para o G.711 foram calculados usando um $ptime$ de 20 ms.

Já era esperado um melhor desempenho do Modelo E Estendido que do Modelo E original nos casos com $ptime$ maior, uma vez que o Modelo E Estendido considera, ainda que de forma limitada, o $ptime$ no cálculo do MOS. Afinal, o Modelo E Estendido usa a informação do tamanho do $ptime$ para calcular a duração dos períodos de perda em rajada e de perdas isoladas. Na Tabela 4.1, pode-se observar claramente como a distribuição de desvios do Modelo E Estendido tende a ser melhor que a do Modelo E original.

Nos cenários com GSM, os resultados foram significativamente melhores. Como pode ser visto na Tabelas 4.4 e 4.5, os modelos foram aprovados na maior parte dos casos, muitas vezes dentro da classe C1. Dos 32 cenários presentes nas tabelas, os modelos não foram aprovados com o GSM em apenas 6 deles, o que significa uma aprovação em 81,25% dos casos. Além disso, entre os 26 casos de aprovação, 22 passaram na classe C1. Com o GSM, o $ptime$ parece influenciar menos no desempenho do Modelo E que no caso do G.711, indicando que a qualidade do GSM é menos sensível à variação do $ptime$.

Nos cenários com GSM, todos os casos nos quais um modelo não foi aprovado aconteceram com $ptime$ de 40 ms e sem PLC. Tal fato faz com que seja razoável afirmar

que o Modelo E com GSM foi calibrado para ser usado com *ptime* de 20 ms com PLC. Como nem o Modelo E nem o Modelo E Estendido são sensíveis ao uso de VAD ou PLC, o fato de tantas aprovações terem se concretizado indica que a qualidade não varia muito em função do uso de tais recursos pelo GSM, diferentemente do que acontece com o G.711.

Nos casos de perdas em rajada, principalmente, era esperado um resultado muito melhor por parte do Modelo E Estendido em relação ao Modelo E original, já que uma de suas principais características é sua suposta capacidade de modelar o efeitos negativos sobre a qualidade causada pelas rajadas de perdas. Afinal, mesmo que o Modelo E original atualmente suporte medir o impacto de rajadas através de *BurstR*, seu mecanismo é simplista quando comparado à cadeia de Markov do Modelo E Estendido. O Modelo E Estendido conseguiu resultados melhores que o Modelo E original tanto nos cenários com *jitter* perdas aleatórias quanto nos cenários de perdas em rajada, tanto com o G.711 quanto com o GSM. Nas Tabelas 4.4 e 4.5, por exemplo, todos os casos que passaram nos testes de conformidade na classe C2 com o Modelo E original foram aprovados na classe C1 quando os mesmos cenários foram testados com o Modelo E Estendido.

De uma maneira geral, observou-se que os modelos não foram devidamente ajustados para o uso de diferentes *ptimes*, VAD e PLC. Mesmo que pareça que eles foram calibrados para uso em determinada configuração de mídia, os resultados não são sempre concretos, com variações no desempenho dos modelos na P.564. O maior exemplo é o caso de *ptime* 20 ms, sem VAD nem PLC. O Modelo E original passou nos testes de conformidade quando essa configuração de codificação foi emulada sob perdas aleatórias e *jitter*, porém, o Modelo E Estendido não passou nos testes com essa mesma configuração. Portanto, a conclusão de tais testes é que o Modelo E e sua extensão não foram devidamente preparados para lidar com tantas configurações diferentes de *ptimes* e uso de VAD e PLC.

5 ADAPTAÇÃO DO MODELO E

Partindo da observação que o Modelo E original não foi planejado para ser usado nos diferentes cenários presentes nos testes da P.564, para que se melhore o seu desempenho nos testes, deve-se calibrar adequadamente o modelo considerando os diferentes *ptimes* e o uso de VAD e PLC. Como no modelo I_e e Bpl são as únicas variáveis que retratam características de um *codec*, os valores dessas variáveis devem ser alterados para que o modelo se adeque aos diferentes cenários aqui considerados. A metodologia para encontrar valores melhores para as variáveis de robustez é usar um método de regressão para que a curva do Modelo E fique mais próxima de cada uma das curvas observadas com o PESQ.

A Figura 5.1 apresenta o gráfico da evolução da qualidade de diferentes *ptimes* em função de diferentes taxas de perdas aleatórias. As chamadas da figura foram feitas usando G.711 sem VAD e sem PLC. É visível a vantagem de *ptimes* menores de lidar com perdas aleatórias. Também é grande a diferença de qualidade conforme a taxa de perda aumenta. A diferença dos MOS chega a 1,93 entre o *ptime* de 10 ms e o de 40 ms quando há 10% de perdas.

A figura também mostra a curva da qualidade medida pelo Modelo E original em

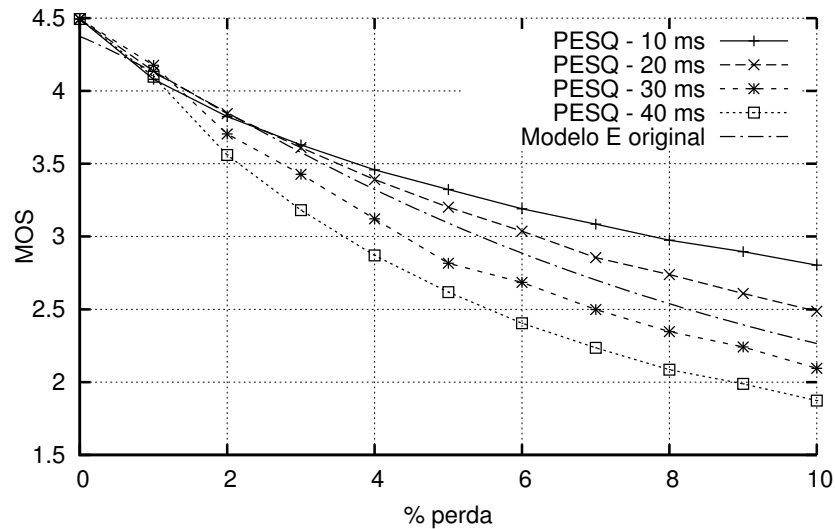


Figura 5.1: Qualidade de diferentes *ptimes* para G.711 sem VAD e sem PLC em função de perdas aleatórias

função da taxa de perda. Na figura, observa-se que a curva de qualidade do Modelo E observada encontra-se próxima às curvas dos *ptime* 20 e 30 ms. Como o modelo original é insensível ao *ptime*, isso significa que o Modelo E produz medições de qualidade próximas somente dos *ptimes* de 20 e 30 ms do PESQ.

Enquanto I_e representa a qualidade de codificação de um *codec* sob nenhuma perda, o Bpl representa a robustez do *codec* a perdas aleatórias. São esses os dois únicos fatores do Modelo E dependentes do *codec* utilizado e que devem passar a ser também sensíveis ao *ptime* e ao uso de VAD e PLC. Conforme o documento G.107 (ITU-T, 2009), o valor do I_e deve pertencer ao intervalo $[0-40]$, enquanto o Bpl deve pertencer ao intervalo $[4,3-40]$. Tais valores podem ser calculados usando emulações de chamadas cuja qualidade é aferida com o PESQ. Assim, esses valores são tão precisos quanto a avaliação do PESQ pode ser. Contudo, como o PESQ é a melhor ferramenta de medição de qualidade atual e é ela mesma que testa os novos valores obtidos através da P.564, sua escolha como modelo de referência é razoável.

Esses novos valores são escolhidos de forma a melhor adaptar o Modelo E à qualidade medida pelo PESQ em cada um dos cenários de codificação. Seria difícil adaptar os parâmetros do Modelo E caso fosse usado o Modelo E Estendido para esta tarefa, uma vez que isso exigiria trabalhar com sua cadeia de Markov, resultando em muitas variáveis a serem consideradas. Para que resultados fossem obtidos em tempo hábil e sem criar maior complexidade de tratamento e análise, os novos valores dessas variáveis são calculados a partir da manipulação matemática das fórmulas do Modelo E original.

5.1 Ajustando o I_e

O valor do I_e não depende do uso ou não de PLC, já que, sob condições ideais sem perdas, tal tecnologia não precisa ser empregada. Porém, o VAD representa, por si só, um fator de degradação e pode fazer com que a qualidade mude conforme o *ptime*.

O uso do VAD traz consigo um efeito chamado *clipping*. No caso da voz, esse termo se refere à perda de milissegundos de áudio devido ao emprego da detecção de voz ativa. Como o VAD ignora tanto o silêncio como ruídos de fundo, essa perda pode ocorrer caso a voz do usuário seja confundida com ruído. Outro caso possível é que o mecanismo de detecção de voz demore alguns milissegundos para detectar que um período de silêncio acabou e a fala está começando.

Portanto, devido ao fato do emprego do VAD alterar a qualidade de uma chamada, para que se obtenha uma melhor previsão da qualidade de chamadas que usam VAD, deve-se ter um I_e específico para o caso. Assim, cada *codec* deve passar a ter dois valores distintos de I_e , um para VAD habilitado e outro para VAD desabilitado.

A recomendação P.834 (ITU-T, 2002) padroniza uma metodologia para o cálculo do I_e de novos *codecs*. Essa recomendação descreve um processo cujo primeiro passo é codificar alguns arquivos de áudio com o *codec* que se deseja calcular o I_e . Depois de codificados, os arquivos devem ser avaliados pelo PESQ. O MOS calculado é então convertido para o fator R do Modelo E e o I_e é calculado ao se inverter a equação de R , dando origem a um I_e chamado cru. Esse mesmo cálculo deve ser feito para uma série de outros *codecs* cujo I_e já foi padronizado. Em seguida, para cada *codec* já padronizado, é criado um ponto num plano cujos eixos são o I_e cru e o I_e padrão. Usando uma técnica de regressão linear, deve ser calculada uma reta que descreva com o menor erro possível os pontos daquele plano. Finalmente, usando a função dessa reta, o I_e cru do novo *codec* pode ser adaptado ao I_e definitivo para o *codec* em questão. Portanto, conforme a P.834, o I_e de um *codec* deve ser adaptado em função dos valores de outros *codecs* para ser calibrado.

Contudo, a P.834 não argumenta porque deve ser criado um novo I_e em função do R de vários *codecs* diferentes ao invés de analisar individualmente a qualidade de cada *codec*. O I_e cru calculado através do uso do PESQ é aquele que resultará em melhores resultados para o Modelo E quando comparado ao PESQ, afinal, o valor teve origem através do uso do próprio PESQ. Alterar o I_e de acordo com a tendência dos outros *codecs*, portanto, somente fará a previsão do Modelo E se afastar da previsão do PESQ.

A recomendação P.834 define ainda que o fator R para o *codec* G.711 sob condições ideais é 87,2 (equivalente a um MOS 4,26). Porém, essa definição é um tanto pessimista pois assumindo um *ptime* de 20 ms e um *buffer* de compensação de *jitter* de tamanho 40 ms chega-se a um valor de R de 91,52 (equivalente a um MOS 4,37). Considerando um atraso nulo e condições perfeitas, o maior valor de MOS que a fórmula do Modelo E pode dar origem é 4,41. Outro erro presente na definição da metodologia, uma fórmula escrita errada, somente foi corrigido numa errata de 2005,

três anos depois do seu lançamento.

Portanto, para que se calcule valores de I_e da forma mais harmônica possível com o PESQ, a adaptação do I_e padronizada na P.834 não deve ser usada. Assim, o processo de cálculo de novos valores de I_e empregado neste trabalho se resume a codificar arquivos de voz com o *codec* em questão, medir a qualidade com o PESQ e ajustar o valor de I_e para que o MOS previsto pelo Modelo E seja equivalente ao medido pelo PESQ. Para tal, deve-se usar a fórmula invertida daquela que converte o fator R ao MOS, como segue abaixo (ITU-T, 2009):

$$R = \frac{20}{3} \left(8 - \sqrt{226} \cos \left(h + \frac{\pi}{3} \right) \right) \quad (5.1)$$

Onde h representa:

$$h = \frac{\arctan2 \left(18566 - 6750MOS, 15\sqrt{-903522 + 1113960MOS - 202500MOS^2} \right)}{3} \quad (5.2)$$

E a função *arctan2* é uma maneira de se calcular o arcotangente da proporção entre x e y e é definida como:

$$\arctan2(x, y) = \begin{cases} \arctan \left(\frac{y}{x} \right) & , \text{ se } x \geq 0 \\ \pi - \arctan \left(\frac{y}{-x} \right) & , \text{ se } x < 0 \end{cases} \quad (5.3)$$

Usando esse conjunto de fórmulas definidas na própria recomendação G.107, pode-se facilmente converter o MOS do PESQ para um valor do fator R . A partir daí, assu-

minindo os valores padrão dos parâmetros do Modelo E, chega-se a seguinte equação:

$$I_e = 93,36 - I_d - R \quad (5.4)$$

Para calcular o valor de R , ainda é necessário que se obtenha um valor para I_d . Nesse caso, sabendo que o atraso da rede não deve ser considerado, o valor do I_d depende apenas do *ptime* empregado e do *buffer* de compensação de *jitter*. Um *ptime* provoca um atraso igual ao seu tamanho, devido ao tempo de amostragem. Para estimar o tamanho do *buffer* de compensação de *jitter* deve-se ter em mente que nos testes usados para calibrar o Modelo E, não é introduzido *jitter* e, logo, o tamanho do *buffer* não varia. Um valor razoável para se estimar o tamanho do *buffer* de compensação de *jitter* numa situação sem atraso ou perdas na rede é de dois pacotes, o que faz com que seu atraso seja aproximadamente o dobro do valor do *ptime*. Usando essas suposições, pode-se facilmente estimar um valor do I_d como sendo igual ao triplo do valor do *ptime*.

Uma vez definida a metodologia de se obter novos valores de I_e , é chegado o momento de exercitar mais uma vez as emulações de chamadas. Cada cenário usa as mesmas permutações dos áudios da P.564 usadas anteriormente. O MOS de cada um desses cenários de codificação é a média dos quatro MOS medidos pelo PESQ. Como não há emulação de nenhum fator de degradação, o MOS medido para cada arquivo não varia, não sendo necessário repetir as emulações em busca de um intervalo de confiança razoável. Novos valores de I_e foram calculados para os *codecs* G.711 μ -law e GSM usando *ptimes* de 10 ms a 40 ms com e sem VAD.

A Tabela 5.1 reúne os valores usados no processo de cálculo dos novos valores de I_e . No caso no qual não se usa VAD, para facilitar os cálculos, é assumido um *ptime* de 20 ms a fim de estimar um valor para I_d . Cada coluna representa o processo

Tabela 5.1: Cálculos de novos valores de I_e

VAD	sem		com					
codec	G.711*	GSM*	G.711				GSM	
ptime	-	-	10	20	30	40	20	40
MOS	4,49	3,62	4,31	4,33	4,32	4,30	3,56	3,52
R	98,78	70,39	88,83	89,68	89,16	88,64	69,30	68,29
I_d	1,83	1,83	1,10	1,83	2,49	3,09	1,83	3,09
I_e	-7,25	21,14	3,43	1,85	1,71	1,63	22,23	21,98
média	-7,25	21,14	2,16				22,11	

* assumindo um $ptime$ de 20 ms.

de cálculo de I_e para cada configuração de *codec*, *ptime* e uso ou não de VAD. As linhas mostram os valores usados em cada passo do processo. O primeiro passo é o cálculo do MOS usando emulações de chamadas com PESQ. Chamadas são feitas sem qualquer degradação além da codificação com o *codec* cujo I_e está sendo calculado. Em seguida, é usada a equação (5.1) para converter o MOS para a escala do fator R . Porém, antes de se calcular o I_e , deve-se estimar um valor para I_d . Como dito anteriormente, um valor razoável para este fator é igual ao triplo do *ptime* usado. Finalmente, usando os valores de R e I_d , pode-se usar a equação (5.4) para estimar um valor de I_e .

Ao analisar a Tabela 5.1, observa-se que o I_e calculado para o G.711 sem VAD é negativo. Isso significa que o Modelo E não consegue chegar à qualidade medida pelo PESQ. De fato, enquanto o PESQ estima a qualidade de codificação do G.711 em 4,49, o MOS máximo atingível pelo sistema de equações do Modelo E pelo modelo é 4,41. Logo, conclui-se que o Modelo E é pessimista quanto à qualidade do G.711. Contudo, o I_e do G.711 sem VAD deve permanecer sendo zero a fim de atender a premissa de que o I_e deve pertencer ao intervalo $[0-40]$.

A qualidade de um *codec* sob o uso de VAD pode variar muito em função do áudio e do *ptime* utilizado, de tal forma que tabelar o I_e para cada *ptime* não é o suficiente para mapear com exatidão sua qualidade. Ao se analisar a evolução do I_e conforme

se aumenta o p_{time} , não se encontra uma natureza previsível já que a série nem mesmo é crescente ou decrescente. Isso aliado ao fato que os I_e calculados tiveram valores próximos faz com que seja razoável adotar um I_e único para todos os p_{times} . Assim, foi decidido usar a média dos I_e calculados para todos os p_{times} de um *codec*. Nesse caso, o I_e do G.711 com VAD passa a ser 2,16 e o do GSM 22,11.

5.2 Representação da Robustez de um *Codec*

Enquanto o I_e é relativamente fácil de ser calculado por refletir a característica de qualidade de um *codec* num ambiente sem perdas de pacotes, os fatores de robustez são mais difíceis de serem obtidos pois é necessário analisar o impacto sobre o *codec* de uma série de taxas de perdas diferentes. Como, por definição, a robustez no Modelo E é em relação a perdas aleatórias, qualquer emulação com o fim de apoiar seu cálculo deve necessariamente usar tais perdas.

Como já foi observado empiricamente durante o andamento deste trabalho, diferentes combinações de p_{time} e de VAD e PLC, sob a ação de uma mesma taxa de perda, podem dar origem a diferentes medidas de qualidade. Assim, percebe-se que, para que o modelo se adapte adequadamente aos diferentes cenários, uma solução possível é calcular um valor de Bpl (fator de robustez de um *codec*) para cada combinação das características de codificação. Porém, somente variar o valor do Bpl pode não ser suficiente para fazer com que o Modelo E seja bem adaptado a uma determinada configuração de codificação.

Para melhor explicar a questão, é apresentada a equação abaixo, uma adaptação da Equação (2.8), a equação original do I_{e-eff} , com a constante 95 trocada por uma nova variável chamada Bpl_1 . A variável Bpl original é substituída por uma nova chamada Bpl_2 .

$$I_{e-eff} = I_e + (Bpl_1 - I_e) \cdot \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl_2} \quad (5.5)$$

Usar duas variáveis para caracterizar a robustez de um *codec* às perdas permite que a curva do MOS medido pelo Modelo E em função das perdas aleatórias possa ser ajustada com mais flexibilidade. A Figura 5.2 ilustra mudanças na curva do MOS com diferentes valores de Bpl_1 e Bpl_2 e serve para mostrar as características da curva observadas com a variação das variáveis. Enquanto Bpl_1 representa de forma mais geral a robustez do *codec*, Bpl_2 representa o quão rápida é a degradação da qualidade do *codec* em função do aumento da taxa de perda.

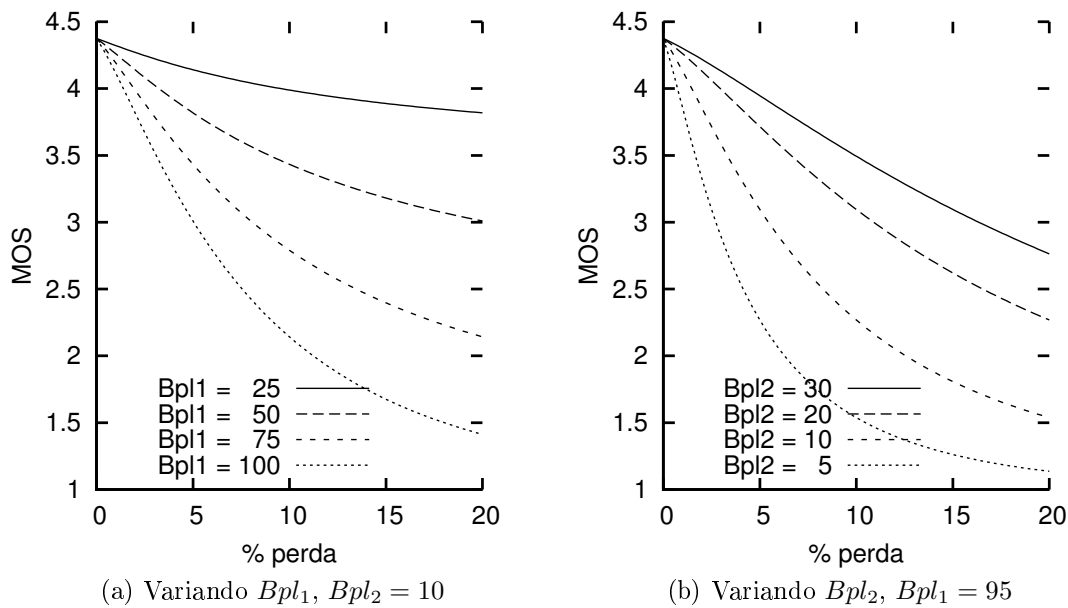


Figura 5.2: Efeitos sobre a curva do MOS medido pelo Modelo E conforme se varia Bpl_1 e Bpl_2

Na Figura 5.2a, usando $Bpl = 10$, o valor padrão do G.711, no lugar da variável Bpl_2 , percebe-se que valores mais baixos de Bpl_1 produzem curvas com quedas menos acentuadas apesar do Bpl_2 ter a capacidade de induzir uma queda rápida da

qualidade. Diz-se que um valor de Bpl_2 de 10 tem tal capacidade já que, no seu contexto original como Bpl , ele representa um *codec* com pequena robustez à perda, o que pode ser observado ao se comparar o valor do Bpl do G.711 aos outros *codecs* na Tabela 2.2, onde o G.711 tem o menor valor de Bpl .

Na Figura 5.2b, pode-se mais uma vez ver que um Bpl_2 igual a 10 provoca uma curva com perda acentuada de qualidade. Nesta mesma figura, na qual se usa Bpl_1 igual a 95, seu valor original, percebe-se como *codecs* com maior robustez, ou seja, um maior Bpl no contexto original, produzem curvas com diminuição de qualidade menos aguda.

Enquanto um valor maior de Bpl_1 equivale a um *codec* de menor robustez à perda, um valor maior de Bpl_2 caracteriza um *codec* mais robusto. A desvantagem de se usar duas variáveis para se modelar a robustez é que se torna complexo comparar a robustez de dois *codecs*, já que agora duas características diferentes da robustez são levadas em conta. Porém, a vantagem de se considerar tal modelo é a flexibilidade de representar mais comportamentos de *codecs*.

Uma prova da melhoria de tal abordagem é a Figura 5.3. Na figura, há uma série de pontos que representam a evolução da qualidade medida pelo PESQ conforme se aumenta a taxa de perdas aleatórias. Foi usado um método de regressão, que será tratado mais adiante, para encontrar os melhores valores das variáveis do Modelo E para representar a qualidade medida pelo PESQ. Foi usada a regressão para encontrar o melhor valor de Bpl na equação original do Modelo E e também para encontrar os melhores valores para o par de variáveis Bpl_1 e Bpl_2 . O cenário de codificação usado foi o *codec* G.711 com *ptime* de 10 ms sem VAD e com PLC. Neste caso, é visualmente perceptível como o modelos com duas variáveis representando a robustez consegue acompanhar a curva do PESQ muito melhor.

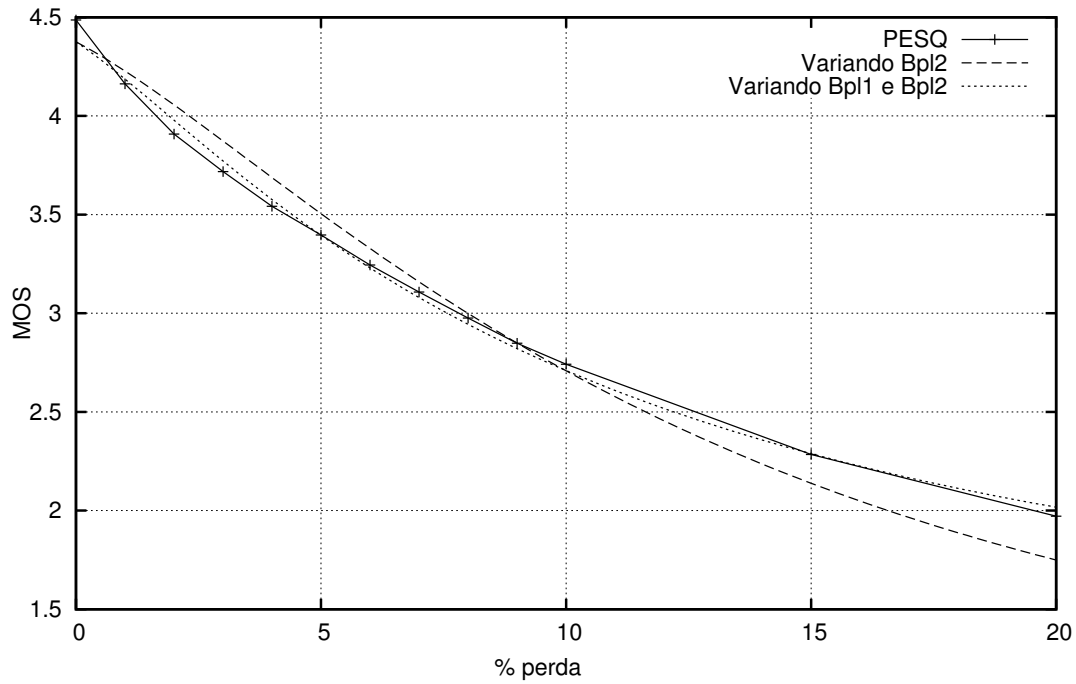


Figura 5.3: Comparação das curvas do Modelo E em função de perda aleatória usando uma ou duas variáveis para representar a robustez do *codec*

Valores de Bpl_1 e Bpl_2 podem ser calculados para cada cenário composto de *codec*, *ptime* e uso ou não de VAD ou PLC. Para cada um desses cenários, deve ser analisada a natureza da qualidade frente a diferentes taxas de perdas aleatórias. Assim, foram emuladas chamadas com perdas aleatórias de 0 a 20% para G.711 e GSM, *ptimes* de 10 ms a 40 ms e com e sem o uso de VAD e PLC e sua qualidade foi medida usando o PESQ. Tal faixa de taxas de perda é definida pelo ITU-T como a faixa de operação do parâmetro Ppl da fórmula do Modelo E. Ou seja, o Modelo E não foi preparado para trabalhar com perdas maiores que 20%. Foram emuladas as seguintes taxas de erro: $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20\}$. Tais taxas foram assim escolhidas para que perdas menores tivessem uma maior representatividade, já que são mais comuns nas redes atuais. Dessa forma, espera-se que os valores calculados sejam mais precisos nos casos de pequenas perdas.

Em cada caso de configuração de codificação e taxa de perdas, as emulações foram repetidas até que se obtivesse um intervalo de confiança de 99% menor que 0,1 na escala do MOS. Tal rigor estatístico foi aplicado com o intuito de calcular os melhores valores possíveis para sua aplicação futura em produção, o que necessita de bastante confiabilidade.

O software Gnuplot (GNUPLLOT, 2011), além de gerar figuras e gráficos, possui uma ferramenta de regressão não linear de mínimos quadrados, uma implementação do algoritmo de Marquardt-Levenberg (LEVENBERG, 1944). Tal ferramenta foi usada para, com base nos dados de emulação com o PESQ, encontrar os melhores valores de Bpl_1 e Bpl_2 para cada cenário, minimizando o erro da curva do Modelo E em relação aos pontos medidos.

Tabela 5.2: Valores de Bpl_1 e Bpl_2 calculados usando regressão

VAD	PLC	$ptime$	G.711 μ -law			GSM		
			Bpl_1	Bpl_2	EQM	Bpl_1	Bpl_2	EQM
sem	sem	10	63,53	6,34	0,066	-	-	-
		20	85,41	9,49	0,047	93,37	14,86	0,013
		30	88,26	7,63	0,061	-	-	-
		40	88,74	6,38	0,065	91,74	9,62	0,048
	com	10	57,19	8,04	0,045	-	-	-
		20	65,61	9,18	0,046	75,21	15,67	0,012
		30	77,31	12,56	0,051	-	-	-
		40	79,58	12,54	0,051	100,52	30,11	0,031
com	sem	10	63,78	7,08	0,062	-	-	-
		20	96,48	15,37	0,043	99,58	20,08	0,037
		30	96,22	11,71	0,043	-	-	-
		40	97,07	10,71	0,058	100,46	14,87	0,050
	com	10	57,07	8,16	0,045	-	-	-
		20	71,11	13,11	0,035	76,66	17,60	0,028
		30	79,78	25,83	0,007	-	-	-
		40	82,09	16,36	0,027	93,36	28,19	0,023

A Tabela 5.2 resume os resultados dos valores de Bpl_1 e Bpl_2 calculados para os *codecs* G.711 e GSM. Os valores das duas variáveis são listados ao lado da configuração de VAD, PLC e $ptime$, totalizando 16 valores para o G.711 e 8 para o GSM, que possui suporte a uma quantidade menor de valores de $ptime$.

Cada ponto de qualidade medido pelo PESQ é comparado ao valor previsto pelo Modelo E. A diferença entre o previsto e o observado é chamada de resíduo. Assim, a coluna EQM sumariza o erro quadrático médio desses resíduos. Percebe-se que, em nenhum caso, o EQM teve uma grandeza maior que 10^{-2} , de onde se conclui que as novas curvas do Modelo E produzem erros de previsão duas ordens de grandeza menores que a escala do MOS, *i.e.* [1–5].

Uma clara desvantagem de se usar tal método é a grande quantidade de dados necessários para tornar o modelo completo. Na tabela, estão presentes 24 pares de dados, o que totaliza 48 valores que devem estar presentes na implementação de tal modelo.

Uma maneira de se reverter essa situação é estudar o comportamento da qualidade nos diferentes cenários de diferentes valores de *ptime* para descobrir se há uma relação matemática que dita a evolução da qualidade em função do *ptime*. Ou seja, deve-se procurar reescrever a fórmula do Modelo E para levar em consideração o *ptime*, levando a um modelo mais genérico e que necessita de menos valores tabelados. Nesse caso, cada *codec* estaria limitado a 4 pares de variáveis Bpl_1 e Bpl_2 para cada um dos cenários com ou sem VAD e PLC.

Uma maneira de melhor entender o conceito de adaptar o Modelo E para que funcione em função também do *ptime* é representar a qualidade medida pelo PESQ para perdas aleatórias num espaço de 3 dimensões. São estas as dimensões: a porcentagem de perdas aleatórias, o *ptime* e o MOS medido pelo PESQ. A partir desses pontos no $\mathbb{R}^3$, pode-se usar o Modelo E para encontrar uma superfície que represente tais pontos com o menor erro possível. A Figura 5.4 ilustra a qualidade medida pelo PESQ para diferentes taxas de perda e *ptimes* para o G.711 sem VAD e sem PLC. A figura também mostra a superfície definida pelo Modelo E depois de ter sua fórmula adaptada e ter novos valores de Bpl_1 e Bpl_2 calculados. A nova fórmula do Modelo

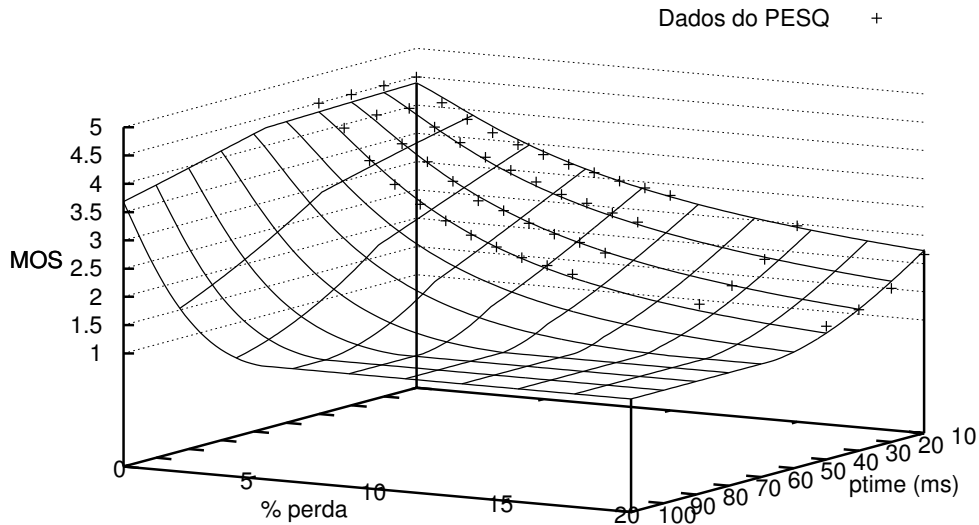


Figura 5.4: Superfície que representa o MOS em função da perda e do *ptime* em chamadas G.711 sem VAD e sem PLC

E será melhor discutida mais adiante.

A Figura 5.4 mostra como é possível definir uma superfície que passe bem perto dos pontos medidos pelo PESQ. Uma vez definida uma superfície como essa, pode-se usá-la até mesmo para estimar a qualidade de *ptimes* maiores, aqueles cuja qualidade medida pelo PESQ não é confiável. Lembrando que deve-se considerar o atraso do *ptime* e que se assume um tamanho de *buffer* de compensação de *jitter* de pelo menos dois quadros, como a fórmula do Modelo E também é sensível ao atraso, a superfície tende a diminuir a qualidade naturalmente com o aumento do *ptime*. A queda da qualidade é mais acentuada conforme o atraso aumenta, como já havia sido notado por Clark na sua simplificação do I_d na Equação (2.7). De fato, como estima-se o atraso de uma chamada como sendo, na prática, 3 vezes o seu *ptime*, um *ptime* de 60 ms equivale a um atraso de 180 ms, passando dos 175 ms que fazem

com que a equação de Clark aumente com maior intensidade a degradação devido ao atraso.

Através do estudo da variação da qualidade conforme o *ptime* e a perda aleatória, experimentou-se incluir o valor do *ptime* de várias formas diferentes na fórmula de I_{e-eff} . Como parece natural que o *ptime* altere a robustez do *codec*, o trabalho foi concentrado em estudar qual a relação entre o valor do *ptime* e os valores de Bpl_1 e Bpl_2 . Dessa forma, foi testado o efeito de usar a soma, produto ou alguma forma de exponenciação ou logaritmo do *ptime* na fórmula. A toda fórmula testada foi aplicada a regressão não linear do Gnuplot para que se calculasse os melhores valores de Bpl_1 e Bpl_2 para a equação para ajustar a fórmula do Modelo E aos dados do PESQ, como presente na Figura 5.4. Ao final, a fórmula que apresentou o menor EQM de seus resíduos foi esta abaixo, que subtrai o *ptime* de Bpl_1 e multiplica a raiz de *ptime* pelo Bpl_2 somente quando se usa PLC:

$$I_{e-eff} = I_e + (Bpl_1 - I_e + ptime) \cdot \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl_2 (\sqrt{ptime})^{PLC}} \quad (5.6)$$

O uso de PLC influencia bastante a robustez do *codec*, de forma que o Bpl_2 é ligeiramente aumentado conforme se aumenta o *ptime*. Mais precisamente, quando se usa PLC, foi notado que o Bpl_2 aumenta numa razão igual à raiz do *ptime*. Na Tabela 5.2, pode-se perceber que, quando não é usado PLC, não é claro se há um crescimento ou não de Bpl_2 nos casos com G.711. Já o caso de GSM, quando não é usado PLC, o Bpl_2 sempre diminui quando se aumenta o *ptime*.

Por outro lado, quando se usa PLC, em apenas dois casos o melhor Bpl_2 calculado usando regressão aumenta conforme o *ptime*. É o caso do G.711 com *ptime* 40 ms sem VAD e com PLC, onde o Bpl_2 é praticamente igual ao caso de *ptime* 30 ms e do

G.711 de $ptime$ 40 ms com VAD e com PLC, onde há um significativo decréscimo no valor de Bpl_2 em relação ao caso com $ptime$ igual a 30 ms. Em todos os outros casos com G.711 e em todos os casos com GSM, o valor de Bpl_2 aumenta conforme o $ptime$ também aumenta, o que acaba se mostrando uma tendência.

A variável PLC que aparece na equação (5.6) é 0 quando não se usa PLC e 1 quando este recurso é utilizado, fazendo com que a raiz do $ptime$ só seja considerada quando o PLC é usado. Foram testados outros valores na exponenciação do $ptime$ que multiplica Bpl_2 , mas o melhor valor encontrado usando regressão foi muito próximo de $\frac{1}{2}$, 0,48. O EQM do resíduo é praticamente o mesmo ao se usar a raiz quadrada do $ptime$ ou de elevá-lo a 0,48. Dessa forma, por simplicidade, decidiu-se manter a raiz.

Foi testado acrescentar constantes multiplicando o $ptime$, a fim de que ele ficasse numa ordem de grandeza diferente. Após usar regressão para encontrar os melhores valores das constantes para multiplicar o $ptime$ nas duas vezes em que é usado na equação, observou-se que as constantes eram sempre muito próximas de 1,0. Mais uma vez foi escolhido manter a fórmula simples ao não usar nenhuma constante multiplicando o $ptime$.

Tabela 5.3: Valores de Bpl_1 e Bpl_2 calculados usando regressão

VAD	PLC	G.711 μ -law			GSM		
		Bpl_1	Bpl_2	EQM	Bpl_1	Bpl_2	EQM
sem	sem	50,49	7,89	0,071	60,96	13,98	0,057
	com	39,73	2,34	0,052	63,63	5,91	0,035
com	sem	51,72	8,32	0,070	61,58	14,59	0,055
	com	40,49	2,37	0,053	61,08	5,56	0,032

Depois de concluída a escolha da nova fórmula de I_{e-eff} , foi mais uma vez usada a ferramenta de regressão do Gnuplot para encontrar os melhores valores de Bpl_1 e Bpl_2 a serem usados em conjunto com a equação (5.6). Mesmo com a nova fórmula, é necessário tabular os valores das duas variáveis para todos as combinações possíveis

de VAD e PLC. Dessa forma, para cada *codec*, ainda é necessário que se calcule 4 pares de valores das variáveis. A Tabela 5.3 mostra os valores calculados para G.711 e GSM da mesma maneira usada anteriormente na Tabela 5.2, com a diferença que agora não é mais necessário levar o *ptime* em conta na tabela.

A principal conclusão que se chega é que a nova fórmula do Modelo E continua dando origem a um EQM na mesma ordem de grandeza, 10^{-2} . De fato, os erros são muito parecidos apesar de a nova fórmula ser responsável por uma quantidade muito maior de pontos a ser considerados. Logo, conclui-se que a nova fórmula de I_{e-eff} , Equação 5.6, é tão precisa quanto a fórmula que não considera o *ptime*, Equação 2.8, apesar de ser mais genérica.

Contudo, tanto no caso do G.711 quanto no do GSM, os valores dos parâmetros da Tabela 5.2b variam pouco quando VAD passa a ser usado. Esse fato indica que o comportamento da queda de qualidade em função das perdas aleatórias é pouco influenciado pelo uso ou não do VAD, mesmo que haja uma queda do MOS quando o VAD passa a ser usado. Portanto, de forma geral, o uso do VAD diminui o MOS com um fator independente da perda, em função do *codec* utilizado.

Assim, uma nova adaptação parece adequada ao formulário do Modelo E: a incorporação de um fator relativo à perda de qualidade com o uso do VAD. Ao usar esse novo fator, não é mais necessário tabelar valores de Bpl_1 e Bpl_2 para os casos que usam VAD. Esse fator ainda trás outra vantagem: já que o novo fator será simplesmente somado à fórmula de I_{e-eff} quando VAD estiver em uso, o uso de um I_e específico para VAD se torna desnecessário, já que seu papel pode ser desempenhado pelo novo fator.

Dessa forma, a fórmula de I_{e-eff} incluindo o Fator de Compensação de VAD (FCV)

passa a ser escrita como a seguir:

$$I_{e-eff} = I_e + (Bpl_1 - I_e - ptime) \cdot \frac{Ppl}{\frac{Ppl}{BurstR} + Bpl_2 (\sqrt{ptime})^{PLC}} + VAD \cdot FCV \quad (5.7)$$

Assim como no caso do uso de PLC na fórmula, a presença da variável VAD serve apenas para limitar o uso de FCV a quando o recurso do VAD é realmente empregado. O valor de FCV deve ser calculado para cada *codec* e para ambos os caso de uso ou não de PLC. Prever o comportamento do PLC é uma tarefa complexa já que o algoritmo de recuperação de quadros perdidos pode depender de uma maneira imprevisível da qualidade da codificação, ou seja, da natureza do *codec*. Esse mesmo argumento pode ser usado para justificar a necessidade de se usar valores de Bpl_1 e Bpl_2 específicos para o uso ou não de PLC.

Tabela 5.4: Valores calculados de FCV e seu EQM associado

<i>codec</i>	PLC	FCV	EQM
G.711	sem	1,08	0,078
	com	1,63	0,058
GSM	sem	0,45	0,058
	com	1,06	0,032

O valor de FCV também é calculado usando regressões sobre os resultados das emulações de perdas aleatórias cuja qualidade foi medida usando o PESQ. Baseado nos valores de I_e , Bpl_1 e Bpl_2 do caso sem VAD, é calculado um valor de FCV que minimize o erro de precisão do Modelo E. Os valores calculados para G.711 e GSM estão na Tabela 5.4 juntamente com o EQM associado ao seu emprego. Mais uma vez, a alteração preserva o EQM na mesma ordem de grandeza de 10^{-2} .

Com o fim de resumir os valores do Modelo E adaptado descrito até aqui, é fornecida a Tabela 5.5 contendo todos os dados necessários para que o modelo seja usado com o *codec* G.711 ou GSM. A tabela é organizada de forma que os valores são apresentados conforme a configuração do cenário onde eles devem ser usados. Em cenários onde

Tabela 5.5: Todos os valores necessários para usar o Modelo E adaptado com os *codecs* G.711 e GSM

<i>codec</i>	VAD	PLC	I_e	Bpl_1	Bpl_2	FCV
G.711	-	sem	0	50,49	7,89	-
	-	com		39,73	2,34	-
	com	sem		-	-	1,08
	com	com		-	-	1,63
GSM	-	sem	21,14	60,69	13,98	-
	-	com		63,63	5,91	-
	com	sem		-	-	0,45
	com	com		-	-	1,06

determinado parâmetros não precisa ser usado, como o caso de FCV quando VAD não é usado, ou quando uma determinada informação do cenário é indiferente, como no caso do uso de VAD para caracterizar Bpl_1 e Bpl_2 , a célula é preenchida com um traço (-).

Os valores de I_e são únicos para um mesmo *codec* e é assumido 0 para o G.711. Já valores de Bpl_1 e Bpl_2 são usados independentemente do uso de VAD, dependendo apenas do *codec* e do uso ou não de PLC. Por último, FCV só é usado juntamente com VAD e possui um valor para cada situação de uso ou não de PLC.

5.3 Novas Simplificações do Modelo E

Usando todos os resultados atingidos até agora, pode-se chegar a uma proposta de uma simplificação conceitual do Modelo E usando os valores *default* dos fatores R_o e I_s e descartando o fator A , como visto na equação abaixo:

$$R = 93,36 - I_d - I_{e-eff} \quad (5.8)$$

Propõe-se também uma fórmula melhor para o cálculo aproximado do fator I_d , inspirado na Equação (2.7). As constantes usadas na fórmula foram calculadas usando a regressão não linear no Gnuplot para que a curva logarítmica fosse a melhor aproximação possível da curva original do Modelo E. A fórmula da simplificação proposta segue abaixo:

$$I_d = \begin{cases} 0,024 \cdot T_a + 0,337 & , \text{ se } T_a < 170 \text{ ms} \\ 29,823 \cdot \ln(T_a) - 148,608 & , \text{ se } T_a \geq 170 \text{ ms} \end{cases} \quad (5.9)$$

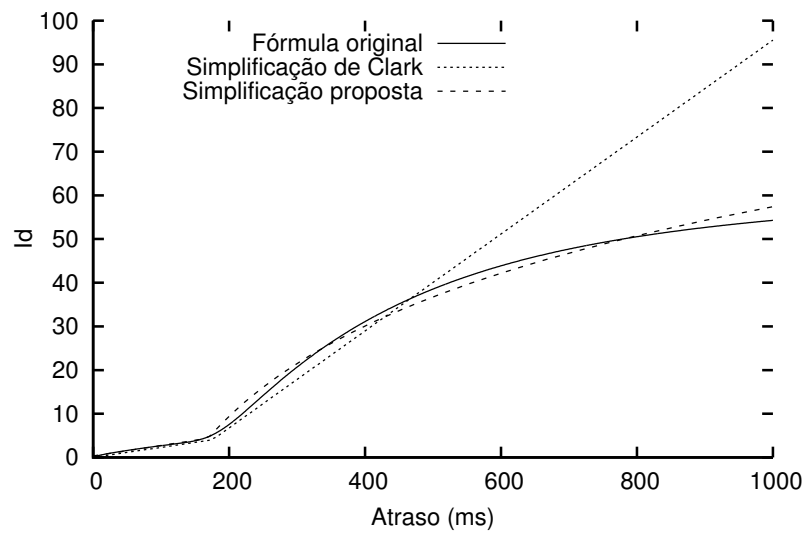


Figura 5.5: Comparação da curva de I_d original do Modelo E à simplificação de Clark e à uma nova simplificação proposta

Com o uso dessa nova fórmula, espera-se uma precisão maior para os casos de grandes atrasos, em relação à fórmula original do Modelo E. A Figura 5.5 compara graficamente as curvas do fator I_d original, da simplificação proposta por Clark e da nova simplificação aqui proposta. Através da figura, facilmente observa-se que, enquanto a aproximação de Clark torna-se crescentemente imprecisa conforme o atraso ultrapassa os 500 ms, a nova curva com sua função logarítmica acompanha de forma aproximada a curva original do I_d .

A esse modelo simplificado pode ainda ser incorporada a tecnologia do Modelo E Estendido, passando a contar com as funcionalidades do Modelo E Estendido como suporte a perdas em rajada além de uma fórmula simplificada de I_d e uma nova representação da robustez dos *codecs* a perdas aleatórias.

5.4 Novos Testes de Precisão

Uma vez definidos novos valores de I_e , Bpl_1 e Bpl_2 para todos os cenários propostos, pode-se passar o Modelo E adaptado e o Modelo E Estendido adaptado mais uma vez pelo crivo da P.564 a fim de mensurar a melhoria com os novos parâmetros. A metodologia da P.564 foi aplicada mais uma vez ao Modelo E e sua extensão seguindo o mesmo procedimento anterior. Os pontos de interesse, calculados baseados na avaliação do PESQ, foram aproveitados, tendo sido necessário apenas refazer emulações usando o Modelo E. O mesmo rigor estatístico anterior foi aplicado.

As Tabelas 5.6, 5.7, 5.8 e 5.9 resumem os resultados dos testes de aderência do Modelo E adaptado. Os relatórios referentes aos testes de conformidade dessas tabelas podem ser encontrados na seção A.2 do Apêndice A. Comparando com os resultados das Tabelas 4.2, 4.3, 4.4 e 4.5 aos novos resultados, é visível uma grande melhoria, com um número muito maior de cenários passando nos testes da P.564, tanto na categoria C1 quanto na categoria C2. Numa análise geral, em 39 dos 64 testes de conformidade com o G.711, o modelo alterado foi aprovado, resultado melhor que dos originais que só foram aprovados em 11 casos. No caso do GSM, em 30 casos de 32 os modelos alterados passaram nos testes de conformidade da P.564, melhorando o resultado anterior no qual os modelos haviam passado em 26 deles. Considerando os resultados dos dois *codecs*, enquanto o modelo original foi aprovado em apenas 38,5% dos cenários testados, o novo modelo passou em 71,9% dos casos. Entre todos os cenários nos quais o modelo passou nos testes, 37 deles passaram

Tabela 5.6: Novos resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas aleatórias e *jitter*

<i>p</i> time	G.711 μ-law – Perdas aleatórias e <i>jitter</i>							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
10	C2	-	C2	C2	C2	-	C2	C2
20	C2	-	C2	C2	C2	C1	C2	C2
30	C2	-	-	C2	C2	-	C2	C2
40	-	-	-	C2	C2	C2	-	C2

Tabela 5.7: Novos resultados de conformidade da P.564 para cenários com G.711 μ -law sob efeito de perdas em rajada

<i>p</i> time	G.711 μ-law – Perdas em rajada							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
10	C2	-	C2	-	C2	-	C1	C2
20	C1	-	C1	-	-	C1	-	C1
30	-	-	-	-	C1	C1	C1	C1
40	-	-	-	-	C1	C1	C2	C1

Tabela 5.8: Novos resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas aleatórias e *jitter*

<i>p</i> time	GSM – Perdas aleatórias e <i>jitter</i>							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
20	C1	C2	C1	C2	C1	C2	C1	C1
40	C1	C2	C1	C2	C1	C2	C1	C1

Tabela 5.9: Novos resultados de conformidade da P.564 para cenários com GSM sob efeito de perdas em rajada

<i>p</i> time	GSM – Perdas em rajada							
	Modelo E				Modelo E Estendido			
	—		VAD		—		VAD	
	—	PLC	—	PLC	—	PLC	—	PLC
20	C1	C1	C1	C1	C1	C1	C1	C1
40	-	C1	-	C1	C1	C1	C1	C1

na classe C1, ou seja, 53,6%. Em especial, é notável uma melhor capacidade de previsão de qualidade do Modelo E Estendido em relação ao Modelo E, ambos após as alterações, nos casos de perdas em rajada.

Apesar de todo o esforço para calibrar o Modelo E, muitos casos com G.711 não passaram nos testes de conformidade. Nos cenários que envolvem perdas em rajada, era de se esperar que o modelo sem sua extensão não tivesse um resultado significativamente favorável. Porém, o Modelo E adaptado foi aprovado em apenas 9 dos 16 cenários de perdas aleatórias e *jitter*. Em especial, nenhum cenário onde PLC foi usado sem VAD o modelo foi aprovado.

Ao analisar a fundo o que fez com que tantos casos não fossem aprovados nos testes de conformidade, descobriu-se que pontos de interesse com valores muito altos de *jitter* foram pontos nos quais os modelos tiveram um desempenho significativamente pior. Com o uso de um *buffer* de compensação de *jitter*, o *jitter* emulado é transformado em atraso e em perda de pacotes. O *buffer* de compensação de *jitter* usado no emulador é dinâmico, o que significa que variações de seu tamanho são comuns em casos de *jitter* alto. Como o Modelo E foi calibrado exclusivamente usando dados de emulações de perdas aleatórias sem *jitter* e como o Modelo E não considera a distribuição temporal, sua capacidade de medir o efeito do *jitter* é limitada. Porém, como o descarte de quadros de voz nos *buffers* costuma acontecer em rajada, o Modelo E Estendido possui uma capacidade maior de medir com precisão a qualidade nas emulações com muito *jitter*. Confirmando seu melhor suporte à medição do impacto de perdas em rajada, o desempenho superior do Modelo E Estendido pode ser visto claramente na Tabela 5.6, com um aumento de 44% no número de casos aprovados nos testes de conformidade, de 9 para 13.

Outro caso notável é o fraco desempenho do Modelo E em casos que se usa PLC, como pode ser visto nas Tabelas 5.6 e 5.7. Esse comportamento pode ser explicado

ao se considerar mais uma vez que o Modelo E foi calibrado usando perdas aleatórias uniformemente distribuídas durante a chamada. O desempenho dos algoritmos de PLC é muito melhor quando os quadros perdidos estão bem espaçados, o que permite que o PLC compense a perda de pequenos intervalos da conversa. No caso dos testes de perdas em rajada da Tabela 5.7 e das rajadas provocadas pelos *buffers* de compensação de *jitter* da Tabela 5.6, as rajadas dificultam bastante o trabalho do algoritmo de PLC, que passa a ter que compensar períodos significativamente maiores de quadros perdidos. Contudo, o Modelo E teve um desempenho melhor quando há a interação do VAD com o PLC, como pode ser visto na quarta coluna com os resultados do modelo na Tabela 5.6.

Os resultados dos testes de conformidade do Modelo E sem suas extensões tende a piorar na Tabela 5.6 conforme aumento o *ptime*. Comportamento semelhante ocorre na Tabela 5.7, no qual o Modelo E sem extensões não passou em nenhum teste nos *ptimes* de 30 e 40 ms. Essa dificuldade do Modelo E em lidar com *jitter* alto quando são usados *ptimes* maiores ocorre apesar dos resultados da Figura 4.3 indicarem que tais *ptimes* maiores são mais resistentes ao *jitter*, o que indica que tal interação deve ser estudada mais a fundo.

O desempenho dos modelos de medir com precisão a qualidade das chamadas emuladas usando GSM agora é ainda melhor, como pode ser visto nas Tabelas 5.8 e 5.9. Dos 30 cenários aprovados, 24 passaram nos testes na classe C1. Isso confirma uma facilidade maior do Modelo E de prever a qualidade desse *codec* em comparação ao G.711.

5.5 Testes de Efeito de Memória Recente

Testes adicionais foram feitos com o PESQ e com o Modelo E Estendido, depois de adaptado, para comparar suas capacidades de medir o efeito de memória recente, efeito que diz que que falhas próximas do final da chamada tem impactos maiores sobre a audiência que falhas no início. Para simular tal efeito, os arquivos de áudio usados pelo emulador foram divididos logicamente em 8 partes. Emulações foram feitas onde cada uma dessas partes era inteiramente perdida e a chamada tinha sua qualidade medida com o PESQ e com o Modelo E Estendido.

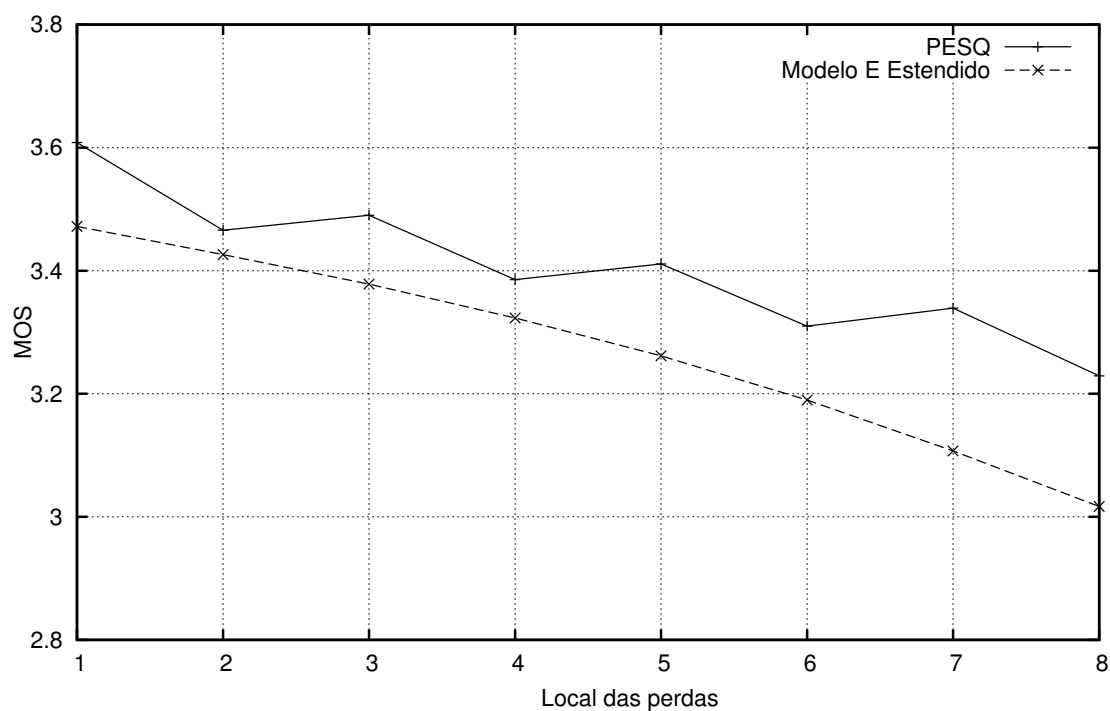


Figura 5.6: Qualidade medida pelo PESQ e pelo Modelo E Estendido para testar o efeito de memória recente

A Figura 5.6 mostra visualmente o resultado das emulações. Foi usado apenas o *codec* G.711 com *ptime* 20 ms sem VAD e sem PLC. Os testes foram repetidos até que obtivesse um intervalo de confiança de 99% menor que 0,1 na escala do MOS

para cada ponto. Na figura, são mostradas curvas que representam a evolução da qualidade medida por ambos os modelos conforme o seguimento perdido segue do início da chamada para o fim. O eixo das abcissas representa o local entre os oito possíveis onde ocorreu a perda. Como pode ser visto na figura, as curvas dos modelos são bem próximas e ambas mostram a diminuição da qualidade conforme as perdas chegam perto do fim da chamada. De fato, o erro quadrático médio entre as duas curvas, baseado nos oito pontos de cada, é de apenas 0,022 pontos na escala do MOS.

Enquanto a curva do Modelo E Estendido cai suavemente devido à sua natureza puramente estatística, a curva do PESQ possui um formato de serra. Esse formato da curva do PESQ ocorre porque, diferentemente do Modelo E Estendido, este modelo de avaliação de qualidade tem acesso direto à mídia e consegue observar que os seguimentos ímpares possuem mais períodos de silêncio que os pares. Isso ocorre devido à própria natureza dos arquivos de áudio usados, que são compostos de pequenos períodos de fala separados por períodos de silêncio. Apesar do formato de serra, a qualidade diminui de forma clara conforme o seguimento perdido se aproxima do fim da chamada.

Conforme o PESQ, a diferença de qualidade entre um período de perdas no primeiro oitavo de uma chamada e no último oitavo chega próxima de 0,4 pontos na escala do MOS, uma queda de um MOS aproximadamente 3,6 para um valor um pouco acima de 3,2, como pode ser visto na Figura 5.6. Uma variação parecida é medida pelo Modelo E Estendido, que estima uma queda de um valor ligeiramente acima de 3,4 para um valor de aproximadamente 3,0.

Mesmo que em todos os caso tenham sido perdidos sempre um oitavo dos pacotes da chamada, diferentes valores de MOS foram obtidos para diferentes localizações de um mesmo efeito de degradação. Portanto, esse experimento mostra que am-

bos os modelos reconhecem o efeito de memória recente e de forma semelhante, apresentando pequenas variações.

6 CONCLUSÃO

O Modelo E e sua extensão, o Modelo E Estendido, foram apresentados neste trabalho como opções de modelos de avaliação da qualidade de chamadas de voz em tempo real cuja eficiência precisava ser medida. Assim, foi apresentada a recomendação P.564 cuja finalidade é exatamente servir de teste de conformidade para tais tipos de modelos.

Após adequadamente apresentada a recomendação, foi discutido como aplicar sua metodologia ao caso do Modelo E. Em seguida, diversas emulações de chamadas foram feitas seguindo a P.564 a fim de comparar as avaliações do Modelo E às do PESQ, algoritmo de referência para avaliar a qualidade de chamadas de voz. Os testes de conformidade foram feitos usando o *codec* G.711 μ -law, diferentes valores de *p_{time}* e tanto com quanto sem o uso de PLC e VAD.

Os resultados das emulações de chamadas provaram que tanto o Modelo E original quanto o Estendido passaram nos testes de conformidade da P.564 quando usado G.711, porém, somente quando utilizado um *p_{time}* de 20 ms. Como os textos que definem os dois modelos não discutem a variação de seus parâmetros conforme o *p_{time}* utilizado, especula-se que os modelos foram calibrados apenas com o *p_{time}*

de 20 ms sem VAD e PLC.

Os modelos mostram melhor desempenho no caso do GSM, tendo o *ptime* um impacto menor sobre sua eficácia. Os resultados para esse *codec* não deixam claro para qual *ptime* ele foi calibrado, mas indicam que é assumido o uso do PLC.

Como os resultados dos testes de conformidade do Modelo E indicaram claramente que o modelo seria melhorado caso fosse sensível ao uso de diferentes *ptimes*, VAD e PLC. Dessa forma, emulações de chamadas foram feitas usando perdas aleatórias cuja qualidade foi medida com o PESQ. Com base na evolução da qualidade da perda com o aumento das taxas de perda, para cada configuração de *codec*, *ptime* e uso de VAD ou PLC, foram estimados valores de parâmetros para o Modelo E. Uma vez calculados tais parâmetros, foi possível desenvolver uma extensão ao modelo que representasse, matematicamente, a evolução da qualidade nesses cenários, o que deu origem a uma nova fórmula para o I_{eff} e gerou uma pequena relação de parâmetros que retratam a qualidade do *codec* e sua robustez à perda (novos I_e , Bpl_1 e Bpl_2).

Essa nova fórmula aliada aos novos parâmetros podem ser usados tanto no Modelo E original quanto aliada ao Modelo E Estendido. As alterações foram submetidas novamente aos testes de conformidade da P.564, obtendo resultados significativamente melhores que os modelos originais, comprovando o mérito de se diferenciar *ptimes* e uso de VAD e PLC ao se medir a qualidade de uma chamada.

Rajada de pacotes perdidos foram emuladas, forçando rajadas com tamanho de pelo menos dois pacotes consecutivos. Essas rajadas foram criadas a fim de confirmar se o Modelo E Estendido realmente apresentaria uma precisão melhor que o Modelo E original. Como esperado, ao serem forçadas rajadas de perdas de pacotes, o Modelo E Estendido apresentou resultados melhores nos testes de conformidade da P.564. De fato, mesmo em cenários onde não foram emuladas perdas em rajada diretamente,

a extensão do Modelo E se provou mais precisa devido a sua capacidade superior de avaliar a qualidade das perdas em rajada criadas pelo *buffer* de compensação de *jitter* quando a chamada está sob a influência de altos valores de *jitter*.

Foram feitas ainda emulações de chamadas para testar e comparar a capacidade dos modelos em lidar com o efeito de memória recente. A partir da comparação dos resultados do PESQ e do Modelo E Estendido, chegou-se à conclusão que ambos os métodos preveem o impacto de tal efeito de maneira muito semelhante, ambos penalizando de maneira clara eventos de perdas de pacotes que acontecem próximos ao fim da chamada.

A P.564, porém, não oferece um índice de qualidade do modelo sob teste, apenas um resultado de aprovação ou não do modelo para uma série de testes. A metodologia da recomendação peca por não permitir uma graduação de aprovação de um modelo à recomendação como, por exemplo, um índice medido por um número real que indica de forma graduada o desempenho do modelo na metodologia de avaliação. Um índice seria uma maneira mais justa que simplesmente medir a porcentagem de desvios de previsão em pontos de interesse que ficam dentro de determinadas faixas.

Sabendo que há apenas 30 pontos de interesse divididos entre dois grupos e considerando a tendência de que haja mais pontos de interesse no Grupo 1 (em média, 20 pontos de interesse pertencem ao Grupo 1 e 10 ao Grupo 2), pode-se perceber que um número pequeno de desvios é considerado ao se calcular a porcentagem dos desvios pertencentes às 12 faixas previstas. Assim, tomando por exemplo um caso onde há 20 pontos de interesse no Grupo 1, cada ponto de interesse equivale a 5 pontos na escala percentual, tornando completamente inúteis gradações como 97,9% ou 99% que existem como critério. Logo, percebe-se que a troca desse sistema de classificação por um sistema que gere um índice de conformidade conseguirá demonstrar de uma maneira mais completa e real a qualidade de um modelo.

Como os testes com o Modelo E realizados neste trabalho foram sempre comparados ao PESQ, não cabia estudar a eficácia do modelo em relação à sua percepção da degradação da qualidade em função do atraso de uma chamada, o que pode afetar a interatividade resultando em uma experiência muito negativa para o usuário. De tal forma, trabalhos futuros devem focar em realizar experiências semelhantes questionando o Modelo E em relação à sua medição do impacto do atraso.

Nesse sentido, uma grande contribuição ao Modelo E seria a alteração de suas fórmulas para que ele passe a reconhecer o caráter temporal do atraso durante uma chamada de voz. Atualmente, o modelo recebe como entrada apenas uma estimativa do atraso experimentado durante a chamada, atraso da rede ou do *buffer* de compensação de *jitter*. Porém, sabe-se que tais atrasos podem variar bastante durante uma chamada e fazer com que a avaliação do modelo atual seja de má qualidade. O fato de o modelo não considerar a variação do atraso do *buffer* de compensação de *jitter* pode ser um dos principais motivos para os resultados ruins de testes de conformidade neste trabalho. Logo, assim como o Modelo E Estendido reconheceu o caráter temporal das perdas ao se preocupar com a distribuição das mesmas durante uma chamada, trabalhos futuros devem tentar fazer com que a variação do atraso, *i.e.* *jitter*, seja também levada em conta.

Outras duas contribuições futuras são possíveis de serem feitas no mesmo âmbito deste trabalho. Não há garantia de que, com a publicação da P.863 – o P.OLQA, a P.564 seja atualizada mas, independentemente disso, o processo de adaptação dos parâmetros do Modelo E pode ser refeita, dessa vez baseada nos resultados de emulações e medições de qualidade pelo P.OLQA. A segunda possível contribuição é que se faça a adaptação do Modelo E Estendido diretamente a partir de emulações de perdas usando o PESQ ou o P.OLQA. A ideia é que esses modelos podem ser usados para calibrar e, possivelmente, adaptar parâmetros e comportamentos do Modelo E Estendido diretamente.

REFERÊNCIAS

ALLNATT, J. Subjective rating and apparent magnitude. **International Journal of Man-Machine Studies**, [S.l.], v.7, n.6, p.801–816, 1975.

BEERENDS, J.; LARSEN, E.; LYER, N.; VUGT, J. van. Measurement of speech intelligibility based on the PESQ approach. **In: Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN)** [S.l.: s.n.], 2004.

CERNAK, M.; RUSKO, M. An evaluation of synthetic speech using the PESQ measure. **In: Proceedings of the European Congress on Acoustics** [S.l.: s.n.], 2005. p.2725–272.

CLARK, A. D. Modeling the Effects of Burst Packet Loss and Recency on Subjective. **In: Proceedings of the Internet Telephony Workshop (IPTEL '01)** [S.l.: s.n.], 2001.

DING, L.; RADWAN, A.; EL-HENNAWEY, M.; GOUBRAN, R. Performance Study of Objective Voice Quality Measures in VoIP. **In: 12th IEEE Symposium on Computers and Communications, 2007. ISCC 2007** [S.l.: s.n.], 2007. p.197–202.

DITECH NETWORKS. **Limitations of PESQ for Measuring Voice Quality in Mobile and VoIP Networks**. [S.l.: s.n.], 2007. <http://www.ditechnetworks.com>.

Estepa, A. and Estepa, R. and Vozmediano, J.M. On the suitability of the E-model to VoIP networks. **In: Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on** [S.l.: s.n.], 2002. p.511–516.

ETSI. **ETSI TS 102 024-5 V4.1.1. Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Release 4, End-to-end Quality of Service in TIPHON systems; Part 5: quality of service (qos) measurement methodologies.** 2003.

FRIEDMAN, T.; CACERES, R.; CLARK, A. **RTP control protocol extended reports (RTCP XR).** [S.l.]: RFC 3611, 2003.

GILBERT, E. N. Capacity of a burst-noise channel. **Bell Systems Technical Journal**, [S.l.], v.39, p.1253–1265, setembro 1960.

GNUPLOT. **Gnuplot.** [S.l.: s.n.], 2011. Online: <http://www.gnuplot.info/> (acesso em 18/02/2011).

GOUDARZI, M.; SUN, L.; IFEACHOR, E. **PESQ and 3SQM measurement of voice quality over live 3G networks.** 2009.

ITU-T. **Recomendação P.800 — Methods for subjective determination of transmission quality.** 1996.

ITU-T. **Recomendação 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.** 2001.

ITU-T. **Recomendação P.834 — Methodology for the derivation of equipment impairment factors from instrumental models.** 2002.

ITU-T. **SG 12 - D 106 — Estimates of I_e and B_{pl} parameters for a range of CODEC types.** 2003.

ITU-T. **SG 12 - D 104 — Modelling Burst Packet Loss within the E Model.** 2003.

ITU-T. **SG 12 - D 97-E — Packet Loss Distributions and Packet Loss Models.** 2003.

ITU-T. **Recomendação G.114 — One-way transmission time.** 2003.

ITU-T. **Recomendação P.563 — Single-ended method for objective speech quality assessment in narrow-band telephony applications.** 2004.

ITU-T. **Recomendação P.564 — Conformance testing for voice over IP transmission quality assessment models.** 2007.

ITU-T. **SG 12 - C 100 — E-model P.564 Compliance Testing and E-model evolution proposals.** 2007.

ITU-T. **Recomendação G.113 — Transmission impairments due to speech processing.** 2007.

ITU-T. **Recomendação G.107 — The E-model: a computational model for use in transmission planning.** 2009.

ITU-T. **Recomendação P.863 — Perceptual objective listening quality assessment.** 2011.

JOHANNESSON, N. O. The ETSI computation model: a tool for transmission planning of telephone networks. **Communications Magazine, IEEE**, [S.l.], v.35, p.70–79, janeiro 1997.

LEE, M.-K.; JUNG, S.-K.; KANG, H.-G.; PARK, Y.-C.; YOUN, D.-H. A packet loss concealment algorithm based on time-scale modification for CELP-type speech coders. **In: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)** [S.l.: s.n.], 2003. v.1, p.I-116 – I-119.

LEVENBERG, K. A Method for the Solution of Certain Non-Linear Problems in Least Squares. **The Quarterly of Applied Mathematics**, [S.l.], v.2, p.164–168, 1944.

LUSTOSA, L.; RODRIGUES, P.; DAVID, F.; QUINELLATO, D. Arquitetura de Monitoração de Qualidade de Chamadas Telefônicas IP. **In: Proceedings of SBRC '05** [S.l.: s.n.], 2005. p.1073–1086.

LUSTOSA, L.; SOUZA, A.; A. RODRIGUES, P. de; QUINELLATO, D. ACME: an automated tool for generating and evaluating the quality of voip calls. **In: Management of Converged Multimedia Networks and Services**. [S.l.]: Springer Berlin / Heidelberg, 2008. p.91–103. (Lecture Notes in Computer Science, v.5274).

MAHDI, A.; PICOVICI, D. Advances in voice quality measurement in modern telecommunications. **Digital Signal Processing**, [S.l.], v.19, n.1, p.79–103, 2009.

MANJUNATH, T. Limitations of perceptual evaluation of speech quality on VoIP systems. **In: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09** [S.l.: s.n.], 2009. p.1–6.

PENNOCK, S.; HOLUB, J.; SMID, R. Accuracy of the perceptual evaluation of speech quality (PESQ) algorithm. **In: Proceedings of Online Workshop on Measurement of Speech and Audio Quality in Networks** [S.l.: s.n.], 2002. p.49–68.

PRIJONO, B. **Open source SIP stack and media stack for presence, im/instant messaging, and multimedia communication**. [S.l.: s.n.], 2011. Online: <http://www.pjsip.org> (acesso em 18/02/2011).

QIAO, Z.; SUN, L.; IFEACHOR, E. **Case study of PESQ performance in live wireless mobile VoIP environment**. 2008.

RIX, A.; BEERENDS, J.; KIM, D.-S.; KROON, P.; GHITZA, O. Objective Assessment of Speech and Audio Quality – Technology and Applications. **Audio, Speech,**

and Language Processing, IEEE Transactions on, [S.l.], v.14, n.6, p.1890–1901, 2006.

RNP. **fone@RNP**. [S.l.: s.n.], 2011. Online: <http://www.rnp.br/voip/> (acesso em 18/02/2011).

ROSSI, P.; ROMANO, G.; PALMIERI, F.; IANNELLO, G. Bayesian Modelling for Packet Channels. In: **Neural Nets**. [S.l.]: Springer Berlin / Heidelberg, 2003. p.218–225. (Lecture Notes in Computer Science, v.2859).

SOUZA, A. A. D. P. **Integração de Qualidade de Voz a Gateway Asterisk**. 2008.

VERHELST, W.; ROELANDS, M. An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech. In: **IEEE International Conference on Acoustics, Speech, and Signal Processing** [S.l.: s.n.], 1993. v.2, p.554–557.

XYDEAS, C.; ZAFEIROPOULOS, F. Model-based packet loss concealment for AMR coders. In: **IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)** [S.l.: s.n.], 2003. v.1, p.I – 112–15.

APÊNDICE A RELATÓRIOS DOS TESTES DE CONFORMIDADE

Tabela A.2: Relatório dos testes de conformidade para G.711 sem VAD e com PLC

[illegible]

Tabela A.3: Relatório dos testes de conformidade para G.711 com VAD e sem PLC

[illegible]

Tabela A.4: Relatório dos testes de conformidade para G.711 com VAD e com PLC

[illegible]

Tabela A.5: Relatório dos testes de conformidade para GSM sem VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.968	0.951	0.981	0.958
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	94.4	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	83.3	83.3	91.6	25.0
% de desvios na faixa 8	91.6	83.3	100.0	83.3
% de desvios na faixa 9	100.0	91.6	100.0	91.6
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.6: Relatório dos testes de conformidade para GSM sem VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.987	0.950	0.986	0.964
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	94.4	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	91.6	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.7: Relatório dos testes de conformidade para GSM com VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.975	0.944	0.965	0.961
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	91.6	91.6	91.6	75.0
% de desvios na faixa 8	100.0	91.6	100.0	75.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.8: Relatório dos testes de conformidade para GSM com VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.985	0.954	0.984	0.975
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	91.6	100.0	100.0
% de desvios na faixa 8	100.0	91.6	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

A.1.2 Perdas em Rajada

Tabela A.11: Relatório dos testes de conformidade para G.711 com VAD e sem PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.978	0.992	0.997	0.997	0.989	0.994	0.998	0.998
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	64.7	100.0	93.7	88.2	82.3	88.8	100.0	100.0
% de desvios na faixa 2	76.4	100.0	100.0	88.2	88.2	100.0	100.0	100.0
% de desvios na faixa 3	88.2	100.0	100.0	94.1	88.2	100.0	100.0	100.0
% de desvios na faixa 4	88.2	100.0	100.0	94.1	88.2	100.0	100.0	100.0
% de desvios na faixa 5	88.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	0.0	28.5	44.4	12.5	12.5	100.0	100.0	100.0
% de desvios na faixa 8	0.0	71.4	44.4	37.5	25.0	100.0	100.0	100.0
% de desvios na faixa 9	0.0	100.0	88.8	50.0	75.0	100.0	100.0	100.0
% de desvios na faixa 10	75.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 11	87.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.12: Relatório dos testes de conformidade para G.711 com VAD e com PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.912	0.949	0.962	0.971	0.973	0.980	0.983	0.983
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	58.8	58.8	62.5	58.8	58.8	58.8	62.5	58.8
% de desvios na faixa 2	58.8	58.8	62.5	58.8	58.8	58.8	62.5	58.8
% de desvios na faixa 3	58.8	58.8	62.5	58.8	58.8	64.7	81.2	76.4
% de desvios na faixa 4	58.8	58.8	62.5	58.8	58.8	64.7	81.2	76.4
% de desvios na faixa 5	88.2	88.2	93.7	94.1	88.2	100.0	100.0	100.0
% de desvios na faixa 6	88.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 8	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 9	12.5	0.0	0.0	0.0	0.0	12.5	22.2	62.5
% de desvios na faixa 10	62.5	62.5	55.5	62.5	62.5	100.0	100.0	100.0
% de desvios na faixa 11	62.5	62.5	55.5	62.5	100.0	100.0	100.0	100.0
% de desvios na faixa 12	62.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.13: Relatório dos testes de conformidade para GSM sem VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.991	0.990	0.995	0.998
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	50.0	100.0	50.0	100.0
% de desvios na faixa 8	50.0	100.0	90.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.14: Relatório dos testes de conformidade para GSM sem VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.996	0.993	0.996	0.996
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	93.3	100.0	93.3
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.15: Relatório dos testes de conformidade para GSM com VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.992	0.990	0.995	0.997
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	50.0	100.0	60.0	100.0
% de desvios na faixa 8	50.0	100.0	100.0	100.0
% de desvios na faixa 9	90.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.16: Relatório dos testes de conformidade para GSM com VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.995	0.996	0.996	0.994
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

[illegible]

Tabela A.19: Relatório dos testes de conformidade para G.711 com VAD e sem PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.985	0.966	0.960	0.952	0.986	0.984	0.986	0.977
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	94.7	85.0	88.8	90.0	94.7	95.0	94.4	90.0
% de desvios na faixa 2	100.0	95.0	94.4	95.0	100.0	100.0	100.0	95.0
% de desvios na faixa 3	100.0	100.0	100.0	95.0	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	95.0	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	70.0	83.3	80.0	100.0	100.0	100.0	80.0
% de desvios na faixa 8	100.0	100.0	83.3	80.0	100.0	100.0	100.0	80.0
% de desvios na faixa 9	100.0	100.0	91.6	80.0	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.20: Relatório dos testes de conformidade para G.711 com VAD e com PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.962	0.969	0.959	0.965	0.977	0.981	0.973	0.971
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	89.4	84.2	94.4	90.0	84.2	89.4	94.4	85.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	72.7	81.8	83.3	60.0	63.6	100.0	91.6	80.0
% de desvios na faixa 8	100.0	90.9	91.6	90.0	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	91.6	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.21: Relatório dos testes de conformidade para GSM sem VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.992	0.988	0.992	0.984
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.22: Relatório dos testes de conformidade para GSM sem VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.983	0.957	0.987	0.964
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	94.4	94.4	94.4	94.4
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.23: Relatório dos testes de conformidade para GSM com VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.975	0.987	0.986	0.991
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.24: Relatório dos testes de conformidade para GSM com VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.981	0.961	0.986	0.974
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	94.4	94.4	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	91.6	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

A.2.2 Perdas em Rajada

Tabela A.27: Relatório dos testes de conformidade para G.711 com VAD e sem PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.993	0.993	0.998	0.994	0.991	0.995	0.999	0.997
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	88.2	100.0	93.7	58.8	100.0	50.0	100.0	88.2
% de desvios na faixa 2	88.2	100.0	100.0	88.2	100.0	100.0	100.0	94.1
% de desvios na faixa 3	100.0	100.0	100.0	88.2	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	88.2	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	50.0	100.0	44.4	0.0	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	44.4	0.0	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	88.8	0.0	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.28: Relatório dos testes de conformidade para G.711 com VAD e com PLC

Medida	Modelo E				Modelo E Estendido			
	<i>ptime</i>				<i>ptime</i>			
	10	20	30	40	10	20	30	40
Correlação	0.933	0.987	0.997	0.998	0.986	0.993	0.994	0.995
Falsos negativos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% de desvios na faixa 1	88.2	88.2	93.7	88.2	100.0	100.0	100.0	100.0
% de desvios na faixa 2	88.2	88.2	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 3	88.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 4	88.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 7	50.0	87.5	77.7	37.5	75.0	100.0	100.0	100.0
% de desvios na faixa 8	62.5	87.5	88.8	75.0	100.0	100.0	100.0	100.0
% de desvios na faixa 9	75.0	100.0	100.0	87.5	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tabela A.29: Relatório dos testes de conformidade para GSM sem VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.990	0.999	0.996	0.999
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	50.0	100.0	100.0
% de desvios na faixa 8	100.0	80.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.30: Relatório dos testes de conformidade para GSM sem VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.998	0.994	0.994	0.996
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.31: Relatório dos testes de conformidade para GSM com VAD e sem PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.991	0.998	0.997	0.997
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	50.0	100.0	100.0
% de desvios na faixa 8	100.0	70.0	100.0	100.0
% de desvios na faixa 9	100.0	90.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0

Tabela A.32: Relatório dos testes de conformidade para GSM com VAD e com PLC

Medida	Modelo E		Modelo E Estendido	
	<i>ptime</i>		<i>ptime</i>	
	20	40	20	40
Correlação	0.998	0.996	0.993	0.992
Falsos negativos (%)	0.0	0.0	0.0	0.0
Falsos positivos (%)	0.0	0.0	0.0	0.0
% de desvios na faixa 1	100.0	100.0	100.0	100.0
% de desvios na faixa 2	100.0	100.0	100.0	100.0
% de desvios na faixa 3	100.0	100.0	100.0	100.0
% de desvios na faixa 4	100.0	100.0	100.0	100.0
% de desvios na faixa 5	100.0	100.0	100.0	100.0
% de desvios na faixa 6	100.0	100.0	100.0	100.0
% de desvios na faixa 7	100.0	100.0	100.0	100.0
% de desvios na faixa 8	100.0	100.0	100.0	100.0
% de desvios na faixa 9	100.0	100.0	100.0	100.0
% de desvios na faixa 10	100.0	100.0	100.0	100.0
% de desvios na faixa 11	100.0	100.0	100.0	100.0
% de desvios na faixa 12	100.0	100.0	100.0	100.0



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA
CCMN - Bloco C - Cidade Universitária - Ilha do Fundão
Rio de Janeiro - RJ CEP: 21941-916
www.ppgi.ufrj.br