

6 Perdas de Pacotes

O problema de perda de pacotes em rajadas nas redes IP e redes móveis é um dos fatores mais importantes a serem considerados na análise de sistemas de reconhecimento de voz distribuídos. Perdas de pacotes em rajadas causam uma redução drástica do desempenho do reconhecimento de voz. Neste capítulo, apresenta-se uma técnica nova para a reconstrução dos pacotes perdidos baseada em Redes Neurais e compara-se seu desempenho de reconhecimento com aqueles obtidos com as técnicas de inserção de zeros e interpolação linear.

Mesmo redes IP e redes móveis celulares sendo bastantes diferentes umas das outras, ambas sofrem de perdas de pacotes em rajadas. Em redes móveis celulares acontecem em momentos de atenuação acentuada do sinal (*signal fading*) enquanto em redes IP as perdas de pacote ou descartes ocorrem devido ao congestionamento.

Para levar em consideração as características do processo de perda em rajadas, o mesmo foi aproximado usando um processo Markoviano de dois estados, também conhecido como modelo de Gilbert [80]. Os estados se referem aos eventos de “PACOTES RECEBIDOS” e “PACOTES PERDIDOS”. Como mostrado na Fig. 6.1, p denota a probabilidade de transição do estado de “PACOTES RECEBIDOS” para o estado de “PACOTES PERDIDOS” e q denota a probabilidade de transição do estado “PACOTES PERDIDOS” para o estado de “PACOTES RECEBIDOS”.

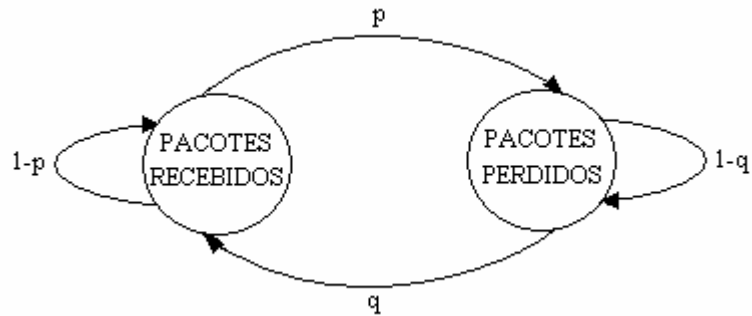


Figura 6.1 – Modelo de Gilbert

A Taxa de Perda de Pacote (TPP), também conhecida como Probabilidade de Perda Incondicional (PPI), é dada por

$$TPP = \frac{p}{p+q} \quad (6.1)$$

e a rajada de perda de pacote é medida pela Comprimento Médio da Rajada (CMR) que é dada por

$$CMR = \frac{1}{1-PCP} \quad (6.2)$$

onde PCP é a Probabilidade Condicional de Perda, que é a probabilidade de permanência no estado de “PACOTES PERDIDOS” (i.e. $PCP = 1 - q$) [81]. O modelo de perda de pacote foi simulado neste capítulo com as condições de rede usadas em [9] e apresentadas na Tab. 6.1.

| TPP(%) | PCP | CMR | p | q |
|--------|------|------|------|------|
| 0 | - | - | 0 | 0 |
| 10 | 0.15 | 1.18 | 0.10 | 0.85 |
| 20 | 0.30 | 1.43 | 0.20 | 0.70 |
| 30 | 0.35 | 1.54 | 0.30 | 0.65 |
| 40 | 0.50 | 2.00 | 0.30 | 0.50 |

Tabela 6.1 –Tabela de condições de rede para o modelo de Gilbert utilizado nas simulações

Na seção 6.1 deste capítulo é apresentada a reconstrução de pacotes perdidos usando Inserção de Zeros e Interpolação Linear. Na seção 6.2, é proposta uma nova técnica baseada em Redes Neurais, com o objetivo de reconstruir os pacotes perdidos. Na seção 6.3, são analisados os resultados de simulação. Finalmente, a seção 6.4 contém uma breve conclusão.

6.1. Inserção de Zeros e Interpolação Linear

Existem algumas aproximações para melhorar o desempenho do sistema de reconhecimento de voz, na presença de imperfeições do canal. Um exemplo é o apagamento de quadros. Uma solução simples para essa degradação é inserção dos zeros na posição dos pacotes perdidos. Uma outra aproximação é interpolação linear entre pacotes recebidos com sucesso (em nosso caso, quadros). O destino recebe, por exemplo, o primeiro conjunto de LSFs quantizadas. Entretanto, devido às imperfeições do canal, não é recebido o segundo conjunto. Na chegada do terceiro conjunto, o receptor pode aproximar o segundo pela interpolação linear do primeiro conjunto com o terceiro. Certamente, a interpolação de mais de um conjunto é praticável em troca de um incremento indesejável de atraso [80]. Para aplicações de redes IP, se n quadros consecutivos de duração t cada um, são perdidos, o atraso devido à interpolação é $D_i = nt + RTT / 2$, onde RTT (*Round-Trip Time*) é o tempo para um pacote ir da fonte ao destino e então de volta à fonte. Valores típicos para RTT variam de 10 a 700 ms e, de acordo com [80], atrasos aceitáveis para aplicações de VoIP não devem exceder 800 ms.

É importante notar que a primeira técnica (Inserção Zero) ignora as características do sinal. Consequentemente, não explora o conhecimento do sinal para melhorar o desempenho do reconhecimento. Por outro lado, o uso da segunda técnica (Interpolação Linear) implica geralmente em longo atraso nos pacotes reconstruídos.

6.2. Redes Neurais

Pelas razões expostas na seção anterior, foi proposta, nesta tese, uma nova técnica baseada em Redes Neurais [82] para reconstrução dos pacotes perdidos, com a vantagem de usar o conhecimento do comportamento do sinal e evitar o retardo significativo para a reconstrução do sinal. O atraso da técnica proposta é somente o tempo das Redes Neurais para computar a saída. Este cálculo está baseado nos quadros de LSFs recebidos antes do pacote perdido ou das LSFs interpoladas obtidas antes do pacote perdido que se deseja recuperar.

Na Fig. 6.2 é apresentada a topologia das Redes Neurais escolhida baseado em resultados de simulações obtidas em uma série de estudos preliminares e pela conhecido uso das Redes Neurais *Multi-Layer Perceptron* na previsão de séries temporais [82]. A camada escondida é composta de 3 neurônios (foram testados 1 à 8 neurônios na camada escondida, tendo o com 3 neurônios apresentado melhor desempenho) cuja função selecionada para o neurônio foi a tangente hiperbólica (não foi testada outra função ativação para a camada escondida). A função linear foi selecionada para o neurônio da camada da saída (não foram testadas outras funções de ativação para o neurônio da camada de saída). Foram utilizadas 10 Redes Neurais com esta topologia, cada uma delas usada para cada uma das 10 LSFs de cada quadro. As 4 entradas de cada Rede Neural são os valores das LSFs ou de seus valores reconstruídos em $T-4$, $T-3$, $T-2$ e $T-1$ (foram testadas janelas de 1 à 8 amostras passadas – o que levou a determinação do número de entradas da rede neural, – tendo a de quatro apresentado melhor desempenho) onde T é o instante em que um quadro é perdido. A saída é a LSF reconstruída em T . Ressalta-se, portanto, que este valor da LSF será usado no sistema de reconhecimento de voz e como uma entrada da rede neural se a LSF de $T+1$ for perdida também. Cada uma das 10 Redes Neurais são treinadas inicialmente com a mesma base de dados usada no treinamento do HMMs (*Hidden Markov Models*). É interessante observar que quando são recebidos 5 quadros sucessivamente com sucesso, são usados os primeiros 4 pacotes como entradas das Redes Neurais e o quinto pacote como sua saída. Este procedimento tem como

única finalidade re-treinar (re-estimar) as Redes Neurais, sendo realizado durante a fase de teste (“on-line”).

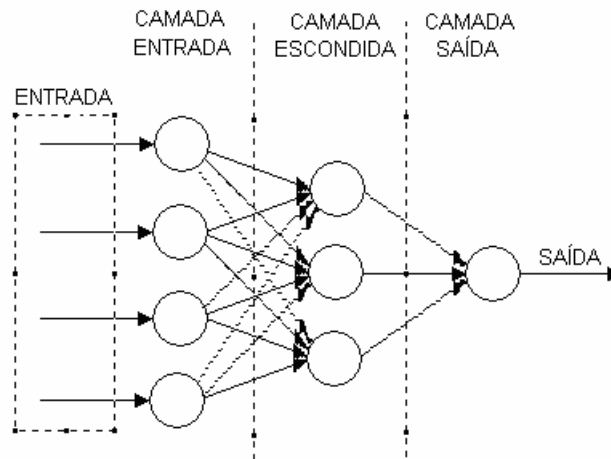


Figura 6.2 – Topologia da Rede Neural

Para o treinamento da rede Neural foi utilizado o algoritmo do Backpropagation (que utiliza a técnica do Gradiente Descendente para ajuste dos pesos da rede neural). Os parâmetros utilizados no treinamento foram:

- Taxa de aprendizado - 0,001
- Número de épocas de Treinamento – 5000
- Medida de erro - Mean Square Error (MSE) - valor alvo 0,0001

Não foi utilizado o fator de Momentum, pois não verificou-se durante o treinamento a captura da rede por mínimos locais que justificassem a utilização do mesmo. Não foi realizada validação cruzada pois devido ao grande volume de dados da base, avaliou-se que era pequeno o risco de overfitting (excesso de treinamento prejudicando a capacidade de generalização) da rede neural.

Não utilizou-se também o Filtro Digital, proposto no capítulo 5, para realizar a recuperação dos pacotes perdidos, pois o mesmo se baseia no espectro do sinal, sendo que neste caso o sinal já havia sido danificado pela perda de pacotes. Neste caso deseja-se recuperar esta informação e não aumentar a taxa com uma informação espectral danificada do sinal, por isso o uso das Redes

Neurais para a realização da restauração dos pacotes perdidos, tentando recuperar a informação perdida.

6.3.

Resultados de Simulação para o Codec ITU-T G.723.1 e AMR-NB

Os resultados do desempenho são apresentados em cinco tabelas, onde em cada tabela são mostrados o desempenho de reconhecimento para o atributo MPCEP obtido das LSF em diversas taxas dos codificadores ITU-T G.723.1 e AMR-NB para diferentes condições de rede. A Tab. 6.2 mostra os resultados do reconhecimento para uma rede ideal sem perda dos pacotes. As Tab. 6.3, 6.4, 6.5 e 6.6 mostram os desempenhos de reconhecimento para redes reais com taxas da perda de pacotes TPP e comprimento médio das rajadas CMR dados por $TPP = 0, 10, 20, 30$ e 40% e $CMR = 0, 1,18, 1,43, 1,54$ e $2,00$, respectivamente. Deve-se observar que em cada caso de teste, os parâmetros do modelo são treinados com o mesmo tipo de atributos (o mesmo tipo de reconstrução), isto é, treinamento e teste estão casados neste sentido. É também importante lembrar que o AMR-NB quando operando em 12,2 kbit/s gera LSFs em 100 Hz, o que evita a necessidade de interpolação das LSFs para esta taxa do codec (para as outras taxas do codec AMR-NB, as LSFs são geradas em 50 Hz e necessitam ser interpoladas para atingir os 100 Hz).

Nas simulações com os codificadores ITU-T G.723.1 e AMR-NB se utilizou a nova técnica de interpolação de parâmetros, utilizando filtros digitais para a realização do aumento de taxa (Capítulo 5) e se restringiu ao uso da base de vozes no cenário mais próximo da utilização real de um sistema de reconhecimento de voz distribuído, que é o reconhecimento independente do locutor e do texto usado no treinamento (Fig. 1.4, do capítulo 1 – cenário 3).

| Atributos | \overline{WRR} | σ | Intervalo de Confiança |
|---|------------------|----------|------------------------|
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 71,32% | 1,57% | [70,63% ; 72,01%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 74,10% | 1,53% | [73,42% ; 74,78%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 72,87% | 1,58% | [72,18% ; 73,56%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 72,89% | 1,58% | [72,20% ; 73,58%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 72,53% | 1,61% | [71,82% ; 73,24%] |

Tabela 6.2 – Tabela de desempenho de reconhecimento para redes sem perdas de pacotes ($TPP = 0\%$ e $CMR = 0$)

| Atributos | WRR | σ | Intervalo de Confiança |
|---|--------|----------|------------------------|
| Inserção de Zeros | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 66,21% | 1,59% | [65,51% ; 66,91%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 70,01% | 1,54% | [69,34% ; 70,69%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 68,12% | 1,60% | [67,42% ; 68,82%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 68,15% | 1,59% | [67,45% ; 68,85%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 67,09% | 1,64% | [66,37% ; 67,81%] |
| Interpolação Linear | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 67,52% | 1,59% | [66,82% ; 68,22%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 71,45% | 1,54% | [70,78% ; 72,13%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 69,53% | 1,59% | [68,83% ; 70,23%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 69,55% | 1,59% | [68,85% ; 70,25%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 68,51% | 1,63% | [67,80% ; 69,23%] |
| Redes Neurais | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 67,54% | 1,58% | [66,85% ; 68,23%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 71,49% | 1,54% | [70,82% ; 72,17%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 69,54% | 1,59% | [68,84% ; 70,24%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 69,58% | 1,59% | [68,88% ; 70,28%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 68,52% | 1,62% | [67,81% ; 69,23%] |

Tabela 6.3 – Tabela de desempenho de reconhecimento para rede com TPP= 10% e CMR=1,18

| Atributos | WRR | σ | Intervalo de Confiança |
|---|--------|----------|------------------------|
| Inserção de Zeros | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 61,57% | 1,64% | [60,85% ; 62,29%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 65,82% | 1,58% | [65,13% ; 66,51%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 63,71% | 1,65% | [62,99% ; 64,43%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 63,77% | 1,64% | [63,05% ; 64,49%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 62,22% | 1,70% | [61,48% ; 62,97%] |
| Interpolação Linear | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 62,63% | 1,63% | [61,92% ; 63,35%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 66,84% | 1,58% | [66,15% ; 67,53%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 64,72% | 1,65% | [64,00% ; 65,44%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 64,78% | 1,63% | [64,07% ; 65,50%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 63,21% | 1,70% | [62,47% ; 63,96%] |
| Redes Neurais | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 63,12% | 1,61% | [62,42% ; 63,83%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 67,37% | 1,56% | [66,69% ; 68,06%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 65,28% | 1,64% | [64,56% ; 66,00%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 65,34% | 1,61% | [64,64% ; 66,05%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 63,71% | 1,69% | [62,97% ; 64,45%] |

Tabela 6.4 – Tabela de desempenho de reconhecimento para rede com TPP= 20% e CMR=1,43

| Atributos | WRR | σ | Intervalo de Confiança |
|---|--------|----------|------------------------|
| Inserção de Zeros | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 57,79% | 1,68% | [57,06% ; 58,53%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 62,01% | 1,63% | [61,30% ; 62,73%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 59,64% | 1,69% | [58,90% ; 60,38%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 59,72% | 1,68% | [58,99% ; 60,46%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 58,10% | 1,75% | [57,33% ; 58,87%] |
| Interpolação Linear | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 58,57% | 1,68% | [57,84% ; 59,31%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 62,81% | 1,63% | [62,10% ; 63,53%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 60,27% | 1,69% | [59,53% ; 61,01%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 60,37% | 1,68% | [59,64% ; 61,11%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 58,68% | 1,75% | [57,91% ; 59,45%] |
| Redes Neurais | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 59,93% | 1,66% | [59,20% ; 60,66%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 64,49% | 1,60% | [63,79% ; 65,19%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 63,91% | 1,67% | [63,18% ; 64,64%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 63,99% | 1,66% | [63,26% ; 64,72%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 62,41% | 1,73% | [61,65% ; 63,17%] |

Tabela 6.5 – Tabela de desempenho de reconhecimento para rede com TPP= 30% e CMR=1,54

| Atributos | WRR | σ | Intervalo de Confiança |
|---|--------|----------|------------------------|
| Inserção de Zeros | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 49,20% | 1,75% | [48,43% ; 49,97%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 56,40% | 1,70% | [55,66% ; 57,15%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 52,99% | 1,77% | [52,22% ; 53,77%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 53,13% | 1,75% | [52,36% ; 53,90%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 51,06% | 1,84% | [50,25% ; 51,87%] |
| Interpolação Linear | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 49,31% | 1,75% | [48,54% ; 50,08%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 56,59% | 1,70% | [55,85% ; 57,34%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 53,27% | 1,76% | [52,50% ; 54,04%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 53,37% | 1,75% | [52,60% ; 54,14%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 51,32% | 1,84% | [50,51% ; 52,13%] |
| Redes Neurais | | | |
| MPCEP - ITU-T G723,1 (5,3 e 6,3 kbit/s) | 52,22% | 1,71% | [51,47% ; 52,97%] |
| MPCEP - AMR-NB (12,2 kbit/s) | 59,47% | 1,65% | [58,75% ; 60,19%] |
| MPCEP - AMR-NB (10,2, 7,40, 6,70 e 5,90 kbit/s) | 56,04% | 1,72% | [55,29% ; 56,80%] |
| MPCEP - AMR-NB (7,95 kbit/s) | 56,14% | 1,70% | [55,40% ; 56,89%] |
| MPCEP - AMR-NB (5,15 e 4,75 kbit/s) | 54,07% | 1,79% | [53,29% ; 54,86%] |

Tabela 6.6 – Tabela de desempenho de reconhecimento para rede com TPP= 40% e CMR=2,00

Dos resultados da simulação fica claro que a Inserção de Zeros é definitivamente a pior aproximação para a solução da perda de pacotes. Agora comparando a Inserção de Zeros, a Interpolação Linear e a proposta de Redes Neurais para a reconstrução de pacotes perdidos de LSFs nas Tab. 6.3, 6.4, 6.5 e 6.6, pode-se ver que a técnica proposta que usa Redes Neurais supera as duas outras técnicas em todos os casos. Entretanto, as melhorias são somente significativas nas Tab. 6.5 e 6.6, correspondendo à perda de pacotes -*TPP* - de 30% e 40%, respectivamente, onde as rajadas de perdas de pacotes das redes IP e das redes móveis celulares são mais severas. No caso onde *TPP* = 40% (Tab. 6.5), o esquema novo fornece ganhos de reconhecimento de aproximadamente 3% quando comparado com a técnica da Interpolação Linear.

Comparando o codec ITU-T G.723.1 nas taxas 6,3 e 5,3 kbit/s com o codec AMR-NB nas taxas similares (6,7 e 5,9 kbit/s), em toda as condições de rede, o AMR-NB fornece um ganho em torno de 1,50%. É também muito significativo notar que os intervalos de confiança de 95% não se sobrepõem. Além disso, é importante observar que o AMR-NB em suas taxas mais baixas (5,15 e 4,75 kbit/s) supera o codec de ITU-T G.723.1 que opera em 6,3 e 5,3 kbit/s para todos os valores de *TPP*. Outra vez, seus intervalos de confiança de 95% não se sobrepõem. Note que as LSFs, de onde os atributos de reconhecimento são extraídos, são codificados em uma taxa de bits mais elevada pelo AMR-NB em

comparação ao ITU-T G.723.1. Finalmente, está claro que as Redes Neurais são uma técnica atrativa para a reconstrução de pacotes perdidos para ambos os codificadores de voz.

6.4. Conclusão

Neste capítulo, foram realizadas diversas experiências importantes em Reconhecimento de Voz Contínuo Distribuído com amplo vocabulário no Português Brasileiro. Foi proposto o uso de Redes Neurais para a reconstrução de pacotes perdidos em sistemas Móveis e redes IP. Comparando com a Inserção de Zeros e a técnica de Interpolação Linear, as Redes Neurais mostraram ser o melhor método para reconstruir pacotes perdidos em sistemas de Reconhecimento de Voz Distribuído que empreguem os codecs ITU-T G.723.1 ou AMR-NB, especialmente em condições severas da perda do pacote. Além disso, foi mostrado que o AMR-NB que opera em uma taxa de bits mais baixa supera o codec ITU-T G.723.1 nas taxas de reconhecimento, sem sobreposição dos seus intervalos de confiança em 95%, em todas as condições da rede.

No próximo capítulo serão apresentadas as conclusões finais e as sugestões para trabalhos futuros.