

Impacto dos parâmetros STZCR e STE em uma comunicação VoIP com perdas de pacotes

Carlos Henrique Pereira, André Simões Aranda, Demóstenes Zegarra Rodríguez

Resumo—Este artigo apresenta um estudo de como os parâmetros intrínsecos de um sinal de voz, como o Short-Time Energy (STE) e o Short-Time Zero Crossing Rate (STZCR) estão relacionados com a qualidade de uma comunicação de voz sobre IP, que sofre perdas de pacotes no canal de transmissão.

Palavras-Chave—VoiP, Qualidade de Voz, MOS, P.862, P.563, STE, STZCR.

Abstract—This article presents a study regarding to the relationship between the intrinsic parameter of a voice signal, such as Short-Time Energy (STE) and Short-Time Zero Crossing Rate (STZCR) and the quality of a voice over IP communication with packet loss rate in the transmission channel.

Keywords—VoiP, Voice Quality, MOS, P.862, P.563, STE, STZCR.

I. INTRODUÇÃO

Os serviços de telecomunicações são essenciais em praticamente todos os âmbitos da atualidade. A qualidade desse serviço baseia-se nas experiências obtidas pelo usuário. Assim, as empresas têm que se focar em uma série de fatores, como a qualidade da comunicação por voz, para manter um alto nível de satisfação do usuário.

Estudos que geram metodologias de análise dos sinais de áudio podem ser usados para apoiar o desenvolvimento dos sistemas de comunicações. Em uma comunicação VoIP, as perdas de pacotes de dados deterioram muito a qualidade dos sinais de voz. Analisar como esse parâmetro afeta as características dos sinais de voz transmitida auxiliaria na melhora das metodologias de análise de qualidade de voz.

Existem vários métodos para analisar sinais de áudio. Porém em cenários reais com perdas de pacotes e degradações de rede, métricas como as recomendações ITU-T P.862 [2], POLQA [4] e ITU-T P.563 [3] não apresentam um desempenho adequado. Considerando as duas primeiras, por precisarem de um sinal de voz de referência para comparar com o sinal de voz degradado, não poderiam ser aplicadas em serviços em tempo real. Enquanto a métrica ITU-T P.563, existem algumas pesquisas que mostram que o desempenho apresentado por ela no serviço de VoIP pode ser melhorado [5], [6], [7], isto se deve ao seu algoritmo que não considera fatores externos, como perda de pacotes em uma rede IP.

Este trabalho tem como objetivo principal, analisar a influência dos parâmetros Short-Time Energy (STE) e Short-Time Zero Crossing Rate (STZCR) na qualidade de um sinal de voz, que sofre diferentes perdas de pacotes.

Para simular os diferentes cenários de perdas de pacotes foi

Carlos Henrique Pereira, André Simões Aranda e Demóstenes Zegarra Rodríguez, Departamento de Engenharia e Departamento de Ciência da Computação, Universidade Federal de Lavras, Lavras-MG, Brasil, E-mails: 201320483@computacao.ufla.br, andresimoesaranda@hotmail.com, demostenes.zegarra@dec.ufla.br. Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) através do Processo (under Grant) no.2012/24789-0. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant no. 307633/2011-0.

utilizado o software wav2rtp [1] que usa o modelo de cadeia de Markov. Os fundamentos teóricos para determinar os parâmetros STE e STZCR estão baseados em [8], [10] e os valores foram obtidos utilizando o software MATLAB.

As próximas seções desse artigo apresentam a metodologia utilizada nos testes, os resultados obtidos e finalmente as conclusões.

II. METODOLOGIA E RESULTADOS

Neste trabalho, a influência dos parâmetros STE e STZCR na qualidade de um sinal de voz que sofre perdas de pacotes é estudada. Conforme [8] e [10], esses parâmetros (STE e STZCR) podem ser utilizados para determinar se um determinado sinal é um sinal de voz.

Os testes foram realizados utilizando 6 arquivos de voz da base de dados ANITA [11], onde 3 oradores são mulheres e os outros 3 são homens. Todos pronunciam a sentença: “*It seemed as if a water ballon had been hurled at the pan because we were all wet*”. Com o software livre Audacity [12], foram criados 3 cenários com os arquivos da base de dados ANITA. No primeiro cenário manteve-se o arquivo original. No segundo cenário foi trocado o sinal da pronúncia da sentença: “*had been hurled at the pan*” por sinal de silêncio. No terceiro cenário foi inserido 42%, em relação ao arquivo original, de sinal de silêncio no arquivo, metade inserido antes da pronúncia da sentença mencionada no segundo cenário e a outra metade depois da pronúncia da mesma. Os cenários 2 e 3 possuem praticamente a mesma proporção de sinal de silêncio em relação ao tamanho do sinal resultante.

As perdas de pacotes foram aplicadas pelo software wav2rtp. Porém os arquivos da base de dados ANITA estão em uma taxa de amostragem de 16kHz e o software wav2rtp somente opera com taxa de amostragem de 8kHz. Portanto foi utilizado o software MATLAB para diminuir a taxa de amostragem dos arquivos utilizados. Foram realizados testes em 30 cenários de diferentes taxas de perdas de pacotes entre 0,5% e 20%. Até 10% a diferença de um cenário para o próximo é de 0,5%, acima de 10% essa diferença é de 1%. Foram geradas 100 amostras para cada cenário, onde o resultado MOS é a média destas amostras. Resultados dispersos foram ignorados para manter um desvio padrão amostral menor ou igual a 0,5.

Nas Figuras 1, 2 e 3 são apresentados os 3 cenários do arquivo f_33_en_c_se01, nas quais os valores dos parâmetros STE e STZCR, calculados pelas funções [9] no MATLAB, também são mostrados.

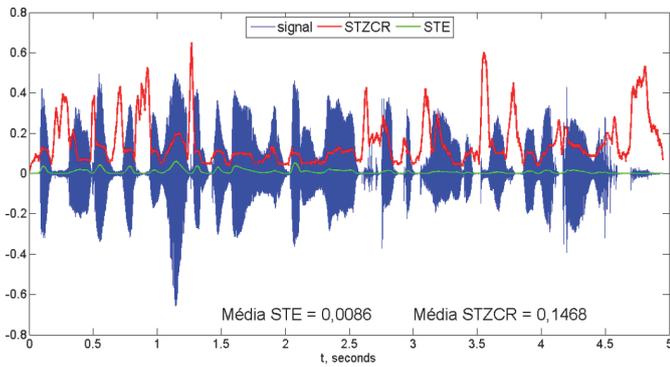


Fig. 1. Cenário 1 – Sinal de voz original vs. STE e STZCR.

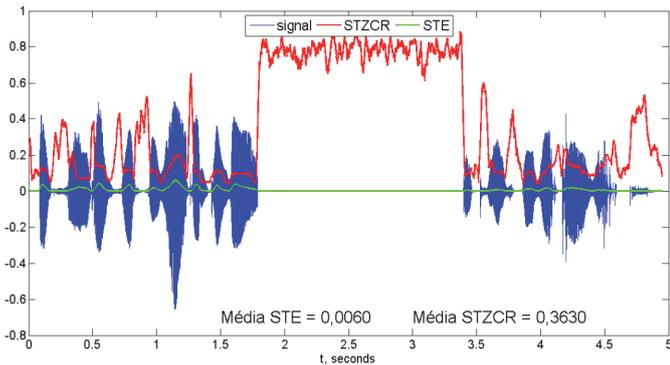


Fig. 2. Cenário 2 – Sinal de voz com um silêncio inserido vs. STE e STZCR.

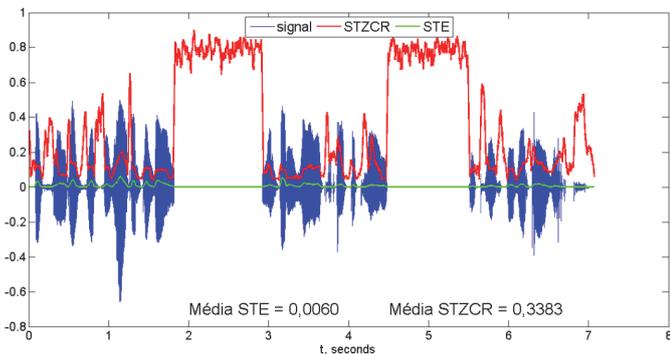


Fig. 3. Cenário 3 – Sinal de voz com dois silêncios inseridos vs. STE e STZCR.

A tabela I apresenta os valores MOS, obtidos pela recomendação P.862, nos 3 cenários do arquivo f_33_en_c_se01 com 5 cenários diferentes de taxas de perdas de pacotes. Observando a tabela I e as 3 figuras acima, percebe-se o impacto dos parâmetros STE e STZCR no comportamento dos valores MOS de acordo com o aumento das taxas de perdas de pacotes, para taxas superiores a 4%. No gráfico da Fig. 4 se pode observar melhor esse impacto. Percebe-se que nos dois cenários (2 e 3) em que os parâmetros STE e STZCR se comportam de forma mais semelhante, o comportamento dos valores MOS em relação as taxas de perdas de pacotes também é mais semelhante.

Deve-se ressaltar que os resultados do P.563 são mais afetados pelos silêncios inseridos do que pelo incremento de taxa de perdas de pacotes. Assim, os resultados do P.563 não foram considerados na análise do impacto dos parâmetros STE e STZCR na qualidade de voz.

TABELA I. MÉDIA DOS ÍNDICES MOS, OBTIDOS COM A REC. P.862, UTILIZANDO O ARQUIVO F_33_EN_C_SE01 DA BASE DE DADOS ANITA

Perda de Pacotes (%)	Cenário 1	Cenário 2	Cenário 3
3	2,936717	3,014630	3,122534
7	2,646216	2,759289	2,778514
11	2,167723	2,453397	2,426066
15	1,865026	2,208807	2,109824
17	1,513427	1,906338	1,882534

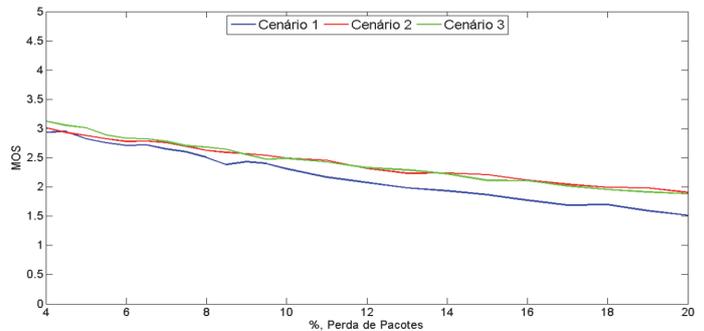


Fig. 4. Média dos índices MOS, obtidos com a Rec. P.862, em relação a taxa de perdas de pacotes.

III. CONCLUSÕES

Como pode-se notar pelos resultados, os parâmetros STE e STZCR influenciam a qualidade final de um sinal de voz, que sofre perdas de pacotes em uma comunicação VoIP. Percebe-se que quanto maior a energia e menor o ZCR de um sinal de voz maior será a degradação da sua qualidade por causa das perdas de pacotes em uma comunicação VoIP. Adicionalmente, como o P.563 mostrou-se ineficiente nos cenários analisados (2 e 3), em trabalhos futuros, pode-se propor alguma melhoria para o P.563 utilizando os parâmetros estudados e outros.

REFERÊNCIAS

- [1] Software wav2rtp. Disponível em <http://wav2rtp.sourceforge.net>.
- [2] ITU-T Rec. P.862, Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.
- [3] ITU-T Rec. P.563, Single-ended method for objective speech quality, assessment in narrow-band telephony applications, May, 2004.
- [4] ITU-T Rec. P.863, Perceptual objective listening quality assessment, Nov. 2011. Disponível em: <http://www.itu.int/rec/T-REC-P.863/en>.
- [5] D. Picovici, A. Raja e C. Flanagan, “Real-Time, Non-intrusive Evaluation of VoIP” in EUROPEAN COM., EUROGP, 10., 2007, Valencia. Anais. Berlin: Springer Berlin Heidelberg, pp. 217-228, 2007.
- [6] T. H. Falk e Wai-Yip Chan, “Enhanced non-intrusive speech quality measurement using degradation models” in Proc. ICASSP, May 2006.
- [7] W. Cherif, A. Ksentini, D. Negru e M. Sidibe, “A_PSQA: PESQ-like non-intrusive tool for QoE prediction in VoIP services” in IEEE International Conference in Communications, pp. 2124-2128, Jun 2012.
- [8] M. Jalil, F. A. Butt e A. Malik, “Short-Time Energy, Magnitude, Zero Crossing Rate and Autocorrelation Measurement for Discriminating Voiced and Unvoiced segments of Speech Signals”, IEEE, 2013.
- [9] N. S. Sharma, “Short-time Energy and Zero Crossing Rate”. Disponível em: <http://www.mathworks.com/matlabcentral/fileexchange/23571-short-time-energy-and-zero-crossing-rate>.
- [10] D. S. Shete, S. B. Patil, “Zero crossing rate and Energy of the Speech Signal of Devanagari Script”, IOSR Journal of VSLI and Signal Processing, Jan 2014.
- [11] Anita Reference Database Description, v.2, EADS Telecom, Abril, 2003.
- [12] Audacity. Disponível em: <http://audacity.sourceforge.net>