

# Avaliação Remota da Soprosidade Vocal em Condições Não-Ideais de Rede

João Pedro Hallack Sansão e Rick Fabian França Oliveira

**Resumo**—Este trabalho apresenta a avaliação do desempenho de medidas acústicas correlatas da soprosidade. O objetivo deste trabalho é avaliar o comportamento de algumas medidas acústicas em amostras de voz transmitidas sob condições não-ideais de rede, com a degradação da qualidade de serviço, como a limitação de banda passante e perdas de pacotes na transmissão. A condição testada neste trabalho é a perda de pacotes durante a transmissão dos dados. Dois bancos de amostras de voz são testados: um banco com amostras de voz sintética e outro com voz natural de pacientes disfônicos. As medidas acústicas testadas são: CPPS, S2NR, SFR e Pitch-Amplitude. Após os experimentos, verificou-se que perdas de até 4% não afetaram significativamente o resultado das medidas nos conjuntos de amostras avaliados.

**Palavras-Chave**—*Soprosidade Vocal, Medidas Acústicas, Transmissão de Dados, Redes, Telecomunicações.*

**Abstract**—This article presents the performance of acoustic measures correlated to vocal breathability. The aim of this article is to evaluate the response of some acoustic measures from speech samples transmitted under non-ideal network conditions. Two sample banks are tested: a bank with synthetic speech samples and another one with natural voice of dysphonic patients. The experiments were run under the condition of packet loss during data transmission. Results show that the measurements are not significantly affected by packet loss up to 4%. The acoustic measures tested are CPPS, S2NR, SFR e Pitch-Amplitude.

**Keywords**—*Vocal Breathability, Acoustic Measures, Data Transmission, Network, Telecommunications.*

## I. INTRODUÇÃO

A análise acústica da voz visa descrever parâmetros objetivos para discriminar falantes disfônicos dos não-disfônicos, correlacionar um determinado parâmetro à avaliação perceptiva da qualidade de voz e permitir o tratamento longitudinal dos pacientes [8].

Um dos parâmetros perceptivos que a literatura correlaciona fortemente a medidas acústicas é a soprosidade [2]. Este fenômeno está relacionado ao ruído glótico, causado pela passagem do fluxo turbulento de ar devido ao fechamento incompleto da glote durante o ciclo fonatório [4]. Diversas medidas estão disponíveis para o estudo da soprosidade [2][5][8].

A análise acústica da voz depende basicamente do registro das elocuições em meio digital para posterior processamento. É um método não-invasivo e de baixo custo devido aos recentes avanços na computação pessoal (incluindo dispositivos móveis com poder computacional), ferramentas para registro e transmissão de multimídia. A tecnologia abriu a possibilidade

de realização de triagens, monitoramento e avaliação da qualidade de voz de forma remota.

Com frequência, devido às limitações de largura de banda para transmissão e de espaço de armazenamento, se faz necessário comprimir os arquivos. Diversos codecs estão disponíveis para esta tarefa, com variadas características. Estudos anteriores indicam que é possível a avaliação da soprosidade usando medidas acústicas em amostras sob compressão [6][7].

Além das limitações de largura de banda, as redes de comunicação estão sujeitas à degradação de qualidade de serviço (QoS). As redes em condições reais estão sujeitas a perda de pacotes, entregas desordenadas, atrasos e erros de transmissão [3]. Este tipo de degradação pode ser notado principalmente em comunicações em tempo real, como em chamadas de voz sobre IP (VoIP), uma potencial ferramenta para avaliação remota de voz.

Desta forma, é interessante conhecer o efeito da degradação da QoS nas medidas acústicas de soprosidade.

O objetivo deste trabalho é avaliar o comportamento de algumas medidas acústicas em amostras de voz transmitidas sob condições não-ideais de rede, com a degradação da qualidade de serviço.

## II. METODOLOGIA

Os experimentos são feitos em duas etapas: com amostras de voz sintética, com parâmetros controlados (frequência fundamental, trato vocal, relação sinal-ruído) e com voz natural, com crescente grau de soprosidade perceptivamente qualificado [8].

As amostras são degradadas utilizando o software wav2rtp, que é capaz de emular perdas em um streaming de dados. Para o uso deste software, é necessária a redução da taxa de amostragem para 8000 Hz.

Estas amostras resultantes são analisadas utilizando as medidas acústicas instrumentais de soprosidade seguintes: CPPS, Pitch-Amplitude, S2NR e SFR [8].

Os valores obtidos nas amostras degradadas são comparados aos valores obtidos nas amostras originais (sem perdas).

As amostras degradadas tem perdas variando de 0 a 20% dos pacotes transmitidos. Verifica-se então a correlação da medida instrumental com a relação sinal-ruído de referência (para as amostras sintéticas) e a correlação da medida com o índice de soprosidade original (para as amostras de voz natural).

Foram utilizados sinais de voz sintetizados com relação sinal-ruído conhecida para as avaliações iniciais. A vogal sintetizada /a/ (amostragem de 22050 Hz, 16 bits por amostra) foi criada convoluindo pulsos glotais com a resposta ao

impulso do trato vocal, a radiação labial sendo modelada por um diferenciador de primeira ordem no fluxo bucal. O modelo de fluxo glótico [1] utilizado nos experimentos tem fator de inclinação (skew) de  $K = 0,65$ . A resposta ao impulso do trato vocal foi estimada utilizando análise com codificação preditiva linear (LPC) com os parâmetros de 95% de pré-ênfase, 22ª ordem, método da covariância, da vogal /a/ de três falantes distintos.

Para estudar a relação entre os valores das medidas citadas e avaliação perceptiva da soproidade, serviu-se de um banco de dados formado pela a vogal sustentada /a/ de 21 indivíduos adultos. As gravações, com três amostras por nível de severidade foram selecionadas e foram perceptualmente avaliados em uma escala de 7 pontos (ausente, pequeno, moderado, extremo com níveis intermediários). A escala perceptiva de soproidade vai de 0 (ausente) a 3 (grau severo), passo de 0,5. As gravações selecionadas (22050 amostras por segundo, 16 bits por amostra) tinham duração superior a 3 segundos, predominantemente soprosas e tinham o nível de soproidade praticamente estável ao longo da elocução [8]. Em falantes disfônicos é comum a ocorrência de instabilidades fonatórias, principalmente nas amostras de soproidade extrema.

### III. RESULTADOS

Após o processamento do lote, num total de 297 amostras (11 valores de SNRref, 3 tratos vocais distintos, 9 níveis de perdas de pacote, 0 a 20%), foram traçados os gráficos da Figura 1, que mostram a correlação do valor obtido pelas medidas e a relação sinal-ruído de referência para cada taxa de perda de pacotes. Na faixa de 0-4%, as medidas mantiveram correlação superior a 0,8. Com perdas de pacotes superiores, a CPPS e S2NR têm suas medidas fortemente degradadas, diminuindo o valor da correlação ou mesmo impedindo seu cálculo. As medidas SFR e pitch-amplitude tem menor degradação na medida, com a correlação caindo para faixa de 0,6 com perdas de 20%.

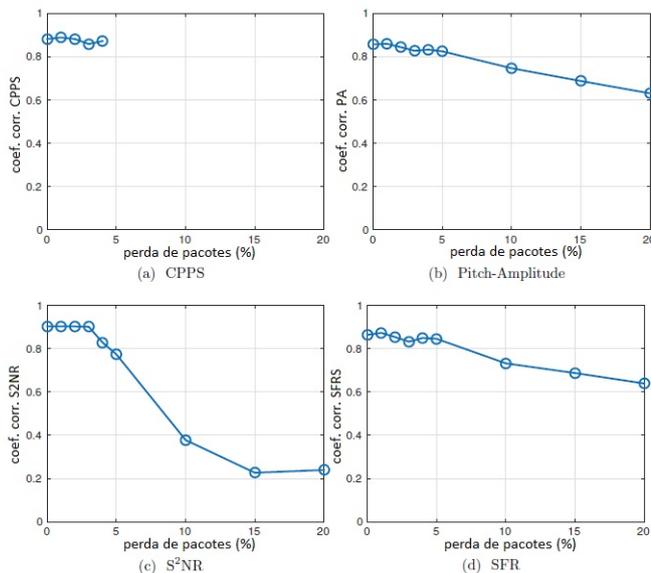


Fig. 1. Correlação da medida com SNR de referência variando o percentual de perda de pacotes com voz sintética.

Com o mesmo procedimento da seção anterior repetido, foram traçados os gráficos da Figura 2. As medidas CPPS e

pitch-amplitude mantêm valores de correlação próximos ao valor de referência na transmissão sem perdas. No caso da S2NR, o valor da correlação aumenta com o aumento da taxa de perda de pacotes.

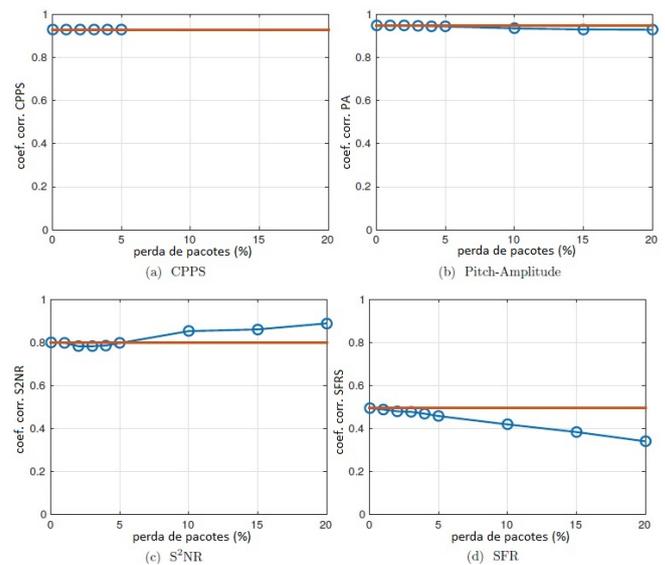


Fig. 2. Correlação com a avaliação perceptiva variando o percentual de perda de pacotes em voz natural.

### IV. CONCLUSÕES

Os resultados apresentados nas figuras mostram que para a voz sintética, na faixa de 0 a 4% de perdas de pacotes, a queda de desempenho das medidas não é significativa. Com taxas superiores de perdas, existe degradação do desempenho. No entanto, para as amostras de voz natural utilizadas, não se detectou a mesma queda da correlação com o aumento das perdas.

Os resultados indicam que, para este conjunto de dados mesmo em condições não ideais de rede, foi possível a avaliação da voz soproosa com as medidas instrumentais citadas.

### REFERÊNCIAS

- [1] Fant, G. (1979). Glottal source and excitation analysis. Speech Transmission Laboratory - Quarterly Progress and Status Report, 1:85-107.
- [2] Hillenbrand, J. e Houde, R. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. Journal of speech and hearing research, 39(2):311.
- [3] Kurose, J. F. e Ross, K. W. (2009). Computer Networking: A Top-Down Approach. Addison Wesley.
- [4] Laver, J., Wirz, S., Mackenzie, J., e Hiller, S. (1981). A perceptual protocol for the analysis of vocal proles. Work in Progress, 14:139-155.
- [5] Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., e Corthals, P. (2009). Acoustic measurement of overall voice quality: A meta-analysis. The Journal of the Acoustical Society of America, 126:2619.
- [6] Sansão, J. P. H., Araújo, L. C., Yehia, H. C., e Vieira, M. N. (2015). Avaliação da soproidade vocal em amostras codicadas pelo codec IETF opus. In XXXIII Simpósio Brasileiro de Telecomunicações 2015 (SBrT2015), Juiz de Fora, Brazil.
- [7] Sansão, J. P. H., Yehia, H. C., e Vieira, M. N. (2013). Evaluation of breathiness acoustic correlates under different compression levels. 10th International Conference Advances in Quantitative Laryngology, page 109.
- [8] Vieira, M. N., Sansão, J. P. H., e Yehia, H. C. (2014). Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. Speech Communication, 61:17-32