Análise dos efeitos dos codecs de áudio na avaliação de desvios vocais

Anselmo de V. Cavalcante, Leonardo W. Lopes, Michel C. Dias, Silvana Cunha Costa, Suzete E. N. Correia

***Resumo*—Este artigo apresenta um estudo sobre as implicações causadas por diferentes codecs de áudio na análise perceptiva da voz. Um cenário de transmissão baseado em VoIP foi criado, empregando o Asterisk e o *softphone* Microsip, para auxiliar no diagnóstico de desvios vocais à distância. Cada sinal selecionado foi submetido a quatro transmissões, cada uma delas utilizando um codec específico. Utilizando a escala EAV, antes e após cada transmissão, um especialista em voz realizou a classificação dos sinais em rugoso, soproso ou normal. Os codecs Speex32 e LPCM16 conseguiram detectar a presença dos desvios com 76% e 88% de Sensibilidade, respectivamente.**

Palavras-Chave*—*Codec de áudio, processamento digital de sinais, telemedicina, voz,VoIP.

***Abstract*—This article presents a study on the implications of different audio codecs on perceptual voice analysis. A VoIP-based transmission scenario was created, using Asterisk and the Microsip softphone, to aid in the diagnosis of distance vocal deviations. Each selected signal was submitted to four transmissions, using a specific codec. Using the EAV scale, before and after each transmission, a speech-language pathologist performed the classification of the signals as rough, breathy or normal. The codecs Speex32 and LPCM16 were able to detect the presence of the deviations with 76% and 88% of Sensitivity, respectively.**

Keywords*—*Audio Codec, digital signal processing, telemedicine, voice, VoIP.

I. INTRODUÇÃO

Com o aumento da expectativa de vida da população, muitas pessoas passaram a apresentar uma série de dificuldades que variam entre a perda de linguagem e distúrbios da fala. Patologias na laringe, sejam de origem neurológica ou orgânica, podem causar desvios vocais, prejudicando a qualidade da comunicação. À medida que a população com distúrbios de comunicação aumenta, a demanda por profissionais e serviços relacionados ao tratamento da voz também cresce.

A avaliação da qualidade vocal, feita por um profissional especialista em voz, emprega, geralmente, uma escala de avaliação analógico-visual (EAV), através da qual é avaliada a presença/ausência de desvio vocal, num processo inicial de triagem, como também o seu grau, que indica a intensidade do desvio. De acordo com a avaliação, que é de caráter subjetivo, será indicada uma análise mais detalhada para determinação da conduta clínica.

Várias barreiras são enfrentadas pelos pacientes para realizar a reabilitação vocal, como incapacidade física para o deslocamento até o local do tratamento, longa distância até os serviços, ausência/indisponibilidade de acompanhantes e dificuldade com transporte e viagem [1].

Nos últimos anos, pesquisadores começaram a investigar a possibilidade de usar a telemedicina para fornecer serviços que auxiliem no tratamento de patologias da fala e linguagem, objetivando alcançar um número maior de pacientes a custos reduzidos [2]. Ela tem sido apontada como uma forma potencial de melhorar os cuidados de saúde nas zonas rurais e de difícil acesso [3].

Dentro da telemedicina, estudos sugerem que o uso da tecnologia Voz sobre IP (*Voice over IP* - VoIP) no tratamento de patologias da voz, é uma alternativa à reabilitação presencial, uma vez que pode mitigar os problemas decorrentes da falta de infraestrutura e dificuldade de locomoção dos pacientes, podendo a reabilitação ser feita na própria residência ou em lugares próximos. Além do mais, o VoIP é acessível a partir de dispositivos semelhantes aos telefones tradicionais e pode ser desenvolvido de forma relativamente barata, usando a Internet [4].

Para o estabelecimento de uma comunicação à distância, utilizando o VoIP, vários parâmetros são levados em consideração, entre eles a escolha dos codificadores e decodificadores de voz, também chamados de codecs. Cada codec possui características específicas, como por exemplo frequência de amostragem, quantidade de bits por amostra e taxa de transmissão. Com o uso dos codecs, o áudio normalmente é comprimido para reduzir a necessidade dea taxa de transmissão, causando a perda de informações [5], podendo levar a um diagnóstico errado da presença/ausência do desvio, bem como de sua severidade.

No trabalho desenvolvido por Ricardo Ferrari *et al* [6] foram analisados os desempenhos de alguns codecs (GSM, G.711, G.722, G.726, G.729, iLBC e Speex) suportados pelo *software* Asterisk, com e sem criptografia, utilizando os protocolos RTP e SRTP, a fim de obter dados importantes para uma tomada de decisão na implementação de um sistema VoIP. Para os testes foi montado um cenário, com o uso de maquinas virtuais, e o auxílio dos *softwares* SIPP, Asterisk, VirtualBox e Wireshark. Para cada codec analisado foram realizadas 16.000 chamadas, sendo que 8.000 com o uso do protocolo RTP e mais 8.000 com o uso do SRTP. Também foram extraídas informações como pico de chamadas simultâneas, chamadas mal sucedidas e bem sucedidas. Dois codecs se destacaram nas comparações, o GSM e o G.726/32, o primeiro com menor tempo de resposta e o segundo com nenhuma chamada mal sucedida para o RTP e apenas uma para o SRTP.

Anselmo de V. Cavalcante, Michel C. Dias, Silvana L. do N. Cunha Costa, Suzete E. N. Correia, Unidade Acadêmica de Indústria, Instituto Federal da Paraíba (IFPB), João Pessoa-PB, Brasil, E-mails: anselmo.cavalante@ifpb.edu.br, michel.dias@ifpb.edu.br, silvana@ifpb.edu.br, suzete@ifpb.edu.br. Leonardo W. Lopes, Laboratório Integrado de Estudos da Voz (LIEV), Universidade Federal da Paraíba (UFPB), João Pessoa-PB, Brasil. E-mail: lwlopes@hotmail.com.

No trabalho de Yanmei Zhu *et al* [7] simulou-se uma comunicação VoIP, com o codec G.729, para determinar os efeitos sobre parâmetros de perturbação acústica de sinais de voz normais e patológicos, utilizando as medidas *jitter* e *shimmer* dos sinais de voz antes e depois da transmissão. Verificou-se que a transmissão VoIP interrompe a forma de onda e aumenta a porcentagem de *jitter* e *shimmer*. No entanto, ainda foi possível a discriminação significativa entre vozes normais e patológicas afetadas pela paralisia laríngea.

O presente artigo descreve um estudo sobre os efeitos dos codecs G.711 Lei A, Speex32, GSM *Full Rate* e LPCM16 na avaliação da qualidade vocal. Um cenário controlado, empregando o Asterisk e o *softphone* Microsip, foi criado para tornar possível a utilização do VoIP e a seleção dos codecs, além da transmissão dos sinais de voz, obtidos de uma base de dados de vozes sintetizadas, do Departamento de Ciência da Computação da Universidade de Brasília [8]. Um avaliador externo (fonoaudiólogo) julgou os níveis dos desvios vocais soprosidade e rugosidade, como também o grau geral desses desvios, antes e depois das transmissões.

Na Seção II deste artigo é realizada uma breve descrição dos codecs utilizados neste trabalho (G.711 Lei A, Speex32, GSM *Full Rate* e LPCM16). Na Seção III é apresentada a utilização da escala EAV como ferramenta de análise perceptivo-auditiva da voz. Na Seção IV a metodologia empregada e na Seção V são apresentados os resultados e, por fim, na Seção VI, a discussão e as considerações finais.

II. CODECS DE ÁUDIO

O áudio normalmente é comprimido para reduzir a necessidade de largura de banda. Todos os sistemas de compressão exigem dois algoritmos: um para comprimir os dados da origem (codificação) e outros para descomprimi-los no destino (decodificação). Na literatura, esse conjunto de algoritmos são conhecidos como codecs [5].

Quando a saída decodificada não é exatamente igual à entrada original, o sistema é considerado com perdas. Se a entrada e a saída forem idênticas, o sistema é sem perdas. Para cada codec pode ser associada uma medida de qualidade, que é uma resposta subjetiva de um ouvinte. Uma medida subjetiva comumente usada para determinar a qualidade do som produzido pelos codecs é o *Mean Opinion Score* (MOS), que pode variar em uma escala de 1 a 5, em que 1 indica voz inaceitável e de baixa qualidade, enquanto um valor - 5 indica alta qualidade de voz, sem problemas perceptíveis.

Alguns dos codecs disponíveis para aplicações VoIP são descritos, de forma sucinta, a seguir.

A. GSM

O *Global System for Mobile communication* (GSM) é um padrão comercial desenvolvido em 1982 pelo *European Telecommunications Standards Institute* (ETSI). Embora originalmente projetado para operação na faixa de 900 MHz, foi logo adaptado para 1800 MHz. Ao longo dos anos, passou a operar em outras bandas de frequência [9].

O padrão GSM apresenta perdas, isto é, alguns dados são perdidos durante a compressão, porém, ele é otimizado para regenerar com precisão a fala na saída de uma ligação [10].

Em um telefone GSM, a voz é convertida em um sinal digital com uma resolução de 13 bits, amostrada a uma taxa de 8 kamostras/s. O GSM analisa a voz e constrói um fluxo de bits composto por uma série de parâmetros que descrevem aspectos da voz. A taxa de saída do codec GSM depende do seu tipo (Tabela I), e varia em um intervalo entre 4,75 kbit/s e 13 kbit/s [10].

TABELA I. TIPOS DE CODEC GSM

Codec	Taxa de bit (kbit/s)
GSM *Full Rate*	13
GSM *Enhanced Full Rate*	12,2
GSM *Half Rate*	5,6
GSM AMR	4,75 - 12,2

B. G.711

O G.711, também conhecido como *Pulse Code Modulation* (PCM), é um codec desenvolvido pela ITU-T em 1988. Originalmente foi criado para telefonia fixa, mas hoje é muito usado em VoIP devido à sua simplicidade e boa qualidade de voz. É uma implementação de quantização logarítmica com 8 bits por amostra, oferecendo assim uma taxa de bits de 64 kbit/s, sendo por isso considerado um codec de alta velocidade [11].

Existem duas versões do codec G.711, que diferem na quantização empregada: Lei A e Lei μ. A Lei μ é usada na América do Norte e Japão e a Lei A é usada no resto do mundo [12].

Usar G.711 para VoIP pode oferecer uma melhor qualidade de voz, uma vez que este mesmo codec é usado pelas redes legadas públicas de telefonia comutada e pelas Redes Digitais com Integração de Serviços (*Integrated Service Digital Network* - ISDN). Ele também tem ótimos níveis de latência (atraso) porque há pouca ou nenhuma necessidade de *buffering*, o que custa aos computadores poder de processamento. A desvantagem em utiliza-lo está no fato de que necessita de mais largura de banda do que outros codecs, podendo chegar até 84 kbit/s. O G.711 é suportado pela maioria das empresas provedoras de VoIP [12].

C. LPCM

A modulação por código de pulso linear (*linear pulse-code modulation* - LPCM) é um tipo específico de codificação PCM onde os níveis de quantificação são linearmente uniformes. Isto contrasta com as codificações PCM em que os níveis de quantização variam em função da amplitude. O LPCM é o padrão utilizado na codificação do áudio presente no DVD (desde 1995), Blu-ray (desde 2006) e HDMI (desde 2002). A quantidade de bits por amostra mais comuns são 8, 16 ou 24. As frequências de amostragem mais utilizadas são 48 kHz (usado em formato de DVD) e 44,1 kHz (usado em formato de CD), podendo chegar até 192 kHz em equipamentos mais novos [13].

D. Speex

Speex é um formato de codificação de áudio, com perda de dados, criado em 2002 e projetado para ser utilizado com a voz. Ele é livre, de código aberto, e baseado na codificação de fala *Code-Excited Linear Prediction* (CELP) [14].

O objetivo dos projetistas foi desenvolver um codec que seria otimizado para alta qualidade de fala e baixa taxa de bits. Para isso, o Speex utiliza taxas de bits múltiplas e suporta taxa de amostragem de banda ultra-larga (taxa de amostragem de 32 kamostras/s), banda larga (taxa de amostragem de 16 kamostras/s) e banda estreita (taxa de amostragem de 8

kamostras/s). Ele também foi desenvolvido para ser robusto em relação a pacotes perdidos, mas não para os corrompidos.

O Speex possui resolução de 16 bits por amostra e latência de 30 milissegundos quando opera em banda estreita e 34 milissegundos quando opera em banda larga ou ultra-larga.

A Tabela II [15] [16] mostra as principais características dos codecs utilizados neste trabalho.

TABELA II. PRINCIPAIS CARACTERÍSTICAS DOS CODECS UTILIZADOS

Codec	Bits por amostra	Taxa de transmissão (kbit/s)	Taxa de amostragem (kamostras/s)	Atraso (ms)	MOS
G.711 Lei A	8	64	8	0,125	4.1
Speex 32	16	44,2	32	34	3.8
GSM Full Rate	13	13	8	20	3.6
LPCM 16	16	256	16	~0,5	4.1

III. ANÁLISE PERCEPTIVO-AUDITIVA DA VOZ

A análise perceptivo-auditiva da voz é um procedimento subjetivo, realizado por profissionais treinados para avaliação de desvios vocais. Nesse método, as percepções do avaliador são subjetivas e individuais, podendo ser influenciadas por sua preferência pessoal, pela experiência e pela cultura [17].

Um procedimento bastante comum na análise perceptivo-auditiva da voz é a quantificação do desvio por meio de escalas, com as quais é possível classificar o grau de severidade dos distúrbios. Uma das principais escalas utilizadas nestas análises é a escala Analógico-Visual (EAV). Essa escala constitui de um intervalo de 100 milímetros, dentro do qual há três pontos de corte definidos a partir de estudos clínicos realizados no Brasil. Os pontos de corte da escala EAV estão em 35,5 mm, 50,5 mm e em 90,5 mm. Caso o avaliador marque na escala um valor entre 0 e 35,5 mm, a voz é considerada normal (grau 1 - G1). Valores entre 35,5 e 50,5 mm indicam uma voz com desvio leve (grau 2 - G2). Valores entre 50,5 e 90,5 mm indicam uma voz com desvio moderado (grau 3 - G3). Se o profissional treinado marcar algum valor entre 90,5 e 100 mm, o desvio vocal é considerado intenso (grau 4 - G4) [18]. A escala EAV está ilustrada na Figura 1 [17].

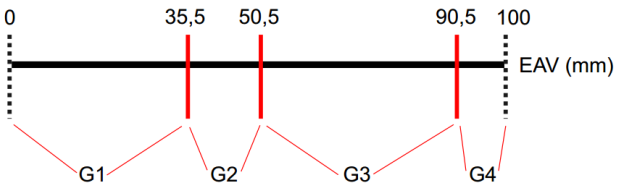

Fig. 1. Escala analógico-visual (EAV).

Entre as características que podem ser observadas em uma análise perceptivo-auditiva estão às relacionadas à qualidade vocal predominante, como a soprosidade e a rugosidade, que são considerados parâmetros de avaliação robustos e universalmente aceitos [19-22], apresentando correlatos nos planos fisiológicos e acústicos. As vozes soprosas são caracterizadas pela percepção de ruído de ar audível nas altas frequências, sobreposto à emissão vocal. Já as vozes rugosas são caracterizadas pela presença de irregularidade na emissão, caracterizado por ruído de baixa frequência [23].

IV. METODOLOGIA

No diagrama de blocos da Figura 2 está ilustrada a metodologia empregada neste trabalho. Os sinais de voz foram selecionados de uma base de dados e armazenados em um servidor Asterisk[1]. Para cada sinal, características perceptivo-auditivas inerentes à voz foram extraídas. Cada um deles foi transmitido quatro vezes através de uma rede de dados, sendo cada transmissão realizada com o uso de um codec específico. Após a transmissão, cada sinal de voz foi recepcionado e armazenado. Uma nova avaliação perceptivo-auditiva foi realizada após a transmissão e os dados obtidos comparados com aqueles extraídos antes da transmissão.

Fig. 2. Modelo da metodologia empregada.

Para a análise comparativa, contou-se com o apoio de um avaliador externo, fonoaudiólogo, com especialização em voz, mestrado em Ciências da Linguagem e doutor em Linguística, com larga experiência em identificação e avaliação de desvios vocais, cujo Coeficiente Kappa de Cohen é 0,79, indicando uma boa confiabilidade [24].

Após a transmissão de cada sinal, foi solicitado ao avaliador que realizasse uma nova análise perceptivo-auditiva, de modo a classificá-lo em normal, rugoso ou soproso, além de inferir o nível de soprosidade, rugosidade e grau geral apresentado por esse sinal, baseado na escala EAV. Os sinais recuperados foram avaliados de maneira aleatória. As informações produzidas por esse avaliador foram comparadas àquelas já dispostas na base de dados.

A avaliação perceptivo-auditiva da base de dados e dos sinais recuperados ocorreu em uma mesma sessão, com duração de 60 minutos, em ambiente silencioso. O avaliador foi treinado com sinais estímulos-âncora, contendo emissões normais e com desvio nos diferentes graus, assim como vozes predominantemente rugosas e soprosas. Além disso, instruiu-se o fonoaudiólogo quanto aos valores de corte [25] que seriam adotados nesta pesquisa para categorização das vozes quanto à ausência e presença de rugosidade e soprosidade. Para avaliação, cada sinal foi apresentado por três vezes através de fone de ouvido simples, em intensidade confortável autorreferida pelo avaliador.

[1] Software livre, de código aberto, desenvolvido para a construção de aplicações de comunicação VoIP.

A. *Base de dados utilizada e seleção dos sinais*

A base de dados foi desenvolvida por um sintetizador (VoiceSim), produzido no Departamento de Ciência da Computação da Universidade de Brasília, em colaboração com os Laboratórios de Imagem, Processamento de Sinal e Acústica da Universidade Livre de Bruxelas. O sintetizador contém um modelo de representação do trato vocal na forma de tubos concatenados através dos quais uma onda acústica se propaga. O material de fala dos estímulos sintetizados foi a vogal do português brasileiro /Ɛ/ ("é"), sustentada por 1 segundo. Esses sinais foram obtidos e armazenados a uma taxa de amostragem de 44100 amostras/s, 16 bits por amostra [8]. A medida de qualidade dos sinais sintetizados produzidos pelo VoiceSim pode ser verificado no trabalho de Englert et al [26].

Foram selecionados inicialmente 36 sinais de voz, divididos em três conjuntos: 11 normais, 14 soprosos e 11 rugosos. Cada um dos 36 sinais selecionados foi transmitido 04 vezes, uma vez para cada tipo de codec (G.711 Lei A, Speex32, GSM *Full Rate* e LPCM16). A seleção foi baseada no grau da escala EAV. Para os sinais com ausência de desvio (normais) foram selecionados da base de dados os que possuíam grau de soprosidade 1 e rugosidade 1. Para os sinais soprosos foram selecionados sinais com grau de soprosidade 2, 3 ou 4 e rugosidade 1. Para os sinais rugosos foram selecionados sinais com grau de rugosidade 2, 3 ou 4 e soprosidade 1.

B. *Cenário de transmissão*

Para tornar possível a transmissão dos sinais de voz pela rede de dados e a seleção dos codecs, foi criado um cenário com o auxílio de dois computadores, conectados através de um equipamento *switch*, conforme ilustrado na Figura 3.

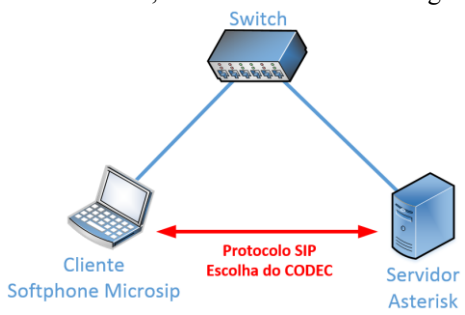

Fig. 3. Cenário utilizado.

Além do *softphone* Microsip, no cliente também foi utilizada a ferramenta de gravação *SoundTap Streaming Audio Recorder*, da NCH *Software*. Esta ferramenta permite a gravação de chamadas VoIP através de um *driver* especial que preserva a qualidade de áudio digital.

O cenário foi criado de maneira controlada, com os elementos conectados via cabos a uma velocidade de 100 Mbit/s, através da tecnologia Ethernet. Parâmetros como o atraso de pacotes e a variação de atraso de pacotes foram desconsiderados neste cenário de estudo, por serem considerados desprezíveis.

O conjunto de 36 sinais originais foi armazenado no servidor Asterisk. Este servidor, por sua vez, foi configurado para atender automaticamente chamadas VoIP, baseadas no Protocolo de Inicialização de Sessão (*Session Initiate Protocol* - SIP), transmitir automaticamente o sinal e em seguida encerrar a chamada. Para cada número de ramal discado (1 a 36) um sinal distinto foi configurado para ser transmitido. Outros parâmetros utilizados nas chamadas VoIP foram o Protocolo de Transporte em Tempo Real (R*eal-time Transport Protocol* - RTP), o Protocolo de Datagrama de Usuário (*User Datagram Protocol* - UDP) e cancelamento de *echo*.

Cada chamada originou-se no cliente, com o auxílio do *software* Microsip, que discava para um número de ramal específico. Logo que a transmissão VoIP iniciava, o *software SoundTrap* começava a gravação da chamada de forma automática e armazenava seu conteúdo em um arquivo específico, em formato WAV.

A seleção do codec foi realizada tanto no servidor Asterisk como também no cliente Microsip. Após todas as transmissões, um conjunto de 144 sinais foi gerado, sendo 36 deles provenientes de cada um dos codecs escolhidos (G.711 Lei A, Speex32, GSM *Full Rate* e LPCM16).

V. RESULTADOS

Na Tabela III estão expostos os valores de Acurácia, Sensibilidade e Especificidade obtidos pelo avaliador após as transmissões, de acordo com codecs estudados. A Tabela IV apresenta os resultados das classificações (Matriz de Confusão) dos 144 sinais analisados pelo avaliador, também após as transmissões. A Acurácia refere-se aos acertos na classificação dos sinais (normal, soproso ou rugoso) após as transmissões. A Especificidade trata da detecção correta da ausência dos desvios em sinais normais e a Sensibilidade é a capacidade de detectar a presença do desvio vocal em sinais desviados.

TABELA III. ACURÁCIA, ESPECIFICIDADE E SENSIBILIDADE DOS CODECS ANALISADOS

Codec	Esp	Sens	Acurácia
G.711 Lei A	100%	52 %	66,7%
Speex32	0	76%	52,8%
GSM *Full Rate*	9,1%	44%	33,3%
LPCM16	18,2%	88%	66,7%

Legenda: Esp: Especificidade; Sens: Sensibilidade

É possível notar que os codecs G.711 Lei A e LPCM16 foram os que tiveram os melhores resultados de Acurácia, ambos com 66,7%, seguidos pelo Speex32, que obteve 52,8%. O pior resultado de Acurácia entre os codecs analisados, ocorreu com o GSM *Full Rate*, que obteve apenas 33,3%.

Observa-se que para a Especificidade, o codec G.711 Lei A obteve resultado de 100% de acerto, enquanto os demais

TABELA IV. RESUMO DA CLASSIFICAÇÃO DOS SINAIS APÓS A TRANSMISSÃO- MATRIZ DE CONFUSÃO

Classificação inicial	G.711 Lei A			Speex32			GSM *Full Rate*			LPCM16		
	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop
Normal (Nor)	11	0	0	0	6	5	1	10	0	2	5	4
Rugoso (Rug)	2	9	0	0	10	1	0	10	1	1	9	1
Soproso (Sop)	1	9	4	0	5	9	0	13	1	0	1	13

apresentaram resultados bem inferiores (0% para o Speex32, 9,1% para o GSM *Full Rate*, 18,2% para LPCM16). Para a Sensibilidade, o codec LPCM16 foi aquele que apresentou melhor resultado, oferecendo 88% de acerto, seguido pelo Speex32 (76%), G.711 Lei A (52%) e GSM *Full Rate* (44%).

VI. DISCUSSÃO E CONSIDERAÇÕES FINAIS

Observa-se que o valor de MOS dos codecs analisados pode ter influenciado na Acurácia, visto que os codecs que possuem os maiores valores de MOS (Tabela II) também apresentaram a melhor Acurácia (Tabela III). Verifica-se ainda que a resolução do codec não influenciou diretamente na Acurácia, uma vez que o LPCM16, mesmo possuindo o dobro de bits por amostra em relação ao G.711 Lei A, obteve o mesmo índice (66,7%).

O valor de atraso do codec pode ter influenciado na identificação da Especificidade, ou seja, na capacidade do avaliador de identificar corretamente as vozes normais entre as que não possuíam desvio, uma vez que quanto maior o atraso (Tabela II), pior foi o índice de Especificidade alcançado (Tabela III).

Em relação à Sensibilidade, observa-se que o número de bits por amostra pode ter influenciado nos resultados, visto que os codecs que possuem o maior número de bits por amostra, apresentaram os melhores índices de Sensibilidade.

Os codecs Speex32 e o LPCM16 foram os que menos prejudicaram a detecção dos desvios vocais no sinal original por parte do avaliador com 76% e 88% de Sensibilidade, respectivamente. O G.711 Lei A foi o único que conseguiu manter a integridade dos sinais normais, sem introduzir desvios significativos, com Especificidade de 100%. No entanto, este codec confundiu o desvio soprosidade com rugosidade. Vale destacar que o desvio rugosidade se encontra numa faixa de frequência até 3 kHz, enquanto que o desvio soprosidade nas faixas de frequências acima de 5 kHz [23]. Dessa forma, o G.711 Lei A, que limita a faixa dos sinais a serem transmitidos em 4 kHz, suprime o desvio soprosidade, não sendo, portanto, adequado para detectar este desvio.

O GSM *Full Rate,* de forma similar ao G.711 Lei A, também eliminou o desvio soprosidade e introduziu o desvio rugosidade em praticamente todos os sinais (Tabela IV).

O Speex32 introduziu ruído nos sinais, tendo transformado os sinais normais em sinais com desvios vocais, tanto rugosidade, quanto soprosidade (Tabela IV). Este codec, no entanto, teve um bom desempenho para detectar os desvios vocais (76%).

O codec LPCM16 apresentou melhor Sensibilidade, detectando a presença do desvio soprosidade, com metade da taxa de amostragem empregada pelo Speex32. Este resultado mostra que o codec LPCM16 é promissor para avaliação da qualidade vocal.

Como trabalhos futuros, sugere-se a extração de métricas acústicas dos sinais de voz, antes e depois das transmissões, a fim de utilizá-las em um classificador com o intuito de se identificar vozes saudáveis e com desvios, bem como o tipo de desvio, sem a necessidade de um avaliador externo. Indica-se ainda, a repetição do mesmo experimento com outros codecs utilizando cenários diversos, por exemplo, uma rede com atrasos de transmissão significativos, perdas de dados, ou ainda uma rede sem fio, a fim de identificar qual codec tem melhor desempenho na classificação dos sinais de voz, nestas configurações.

REFERÊNCIAS

[1] CHERNEY, Leora. VUUREN, Sarel van. *Telerehabilitation, virtual therapists and acquired neurologic speech and language disorders*. Semin Speech Lang. 2012. Disponível em: http://dx.doi.org/10.1055/s-0032-1320044. Acesso em: 12/02/2017.

[2] MASHIMA, P. *et al*. *Telehealth applications in speech-language pathology*. *Journal of Healthcare Information Management*. 1999.

[3] MARTÍNEZ, A. *et al. Analysis of information and communication needs in rural primary health care in developing countries*. IEEE Transactions on Information Technology in Biomedicine 9 (1), 2015.

[4] LAMBRINOS, L. "Deploying open source IP telephony in rural environments". In Proceedings of the International Conference on Next Generation Mobile Applications, Services and Technologies, 2008.

[5] TANENBAUM, Andrew S. WETHERAL, David. *Redes de Computadores.* 5. ed. São Paulo: Pearson, 2011.

[6] FERRARI, Ricardo C. *et al*. *Análise de desempenho dos codecs suportados pelo Astersik com e sem criptografia*. Visão universitária, volume 1, 2014.

[7] ZHU, Yanmei. *et al*. *Effects of the Voice over Internet Protocol on Perturbation Analysis of Normal and Pathological Phonation*. Folia Phoniatrica et Logopaedica, 2010.

[8] BEHLAU, M. MADAZIO, G. LUCERO, J. et al. "Um novo paradigma no ensino da avaliação auditiva de vozes - uso de amostras sintetizadas". XXI Congresso Brasileiro de Fonoaudiologia. Porto de Galinhas, 2013.

[9] ETSI, European Telecommunications Standards Institute. *Mobile technologies GSM*. Acesso em: 23/01/2017. Disponível em: http://www.etsi.org/technologies-clusters/technologies/mobile/gsm.

[10] MESTON, Richard. *Sorting Through GSM Codecs: A Tutorial*. EE Times, 2003.

[11] UNION, *International Telecomunication. ITU-T recommendation G.711*. Geneva, 1988.

[12] RCHANDRA, A. *ITU G.711*. Disponível em: http://www.voip-info.org/wiki/view/ITU+G.711. Acesso em 21/012017.

[13] MONTGOMERY, Christopher. 24/192 Music Downloads. The Xiph.Org Foundation. 2012. Acesso em: 12/03/2017. Disponível em: https://people.xiph.org/~xiphmont/demo/neil-young.html

[14] XIPH. *Speex: A Free Codec For Free Speech*. The Xiph.Org Foundation. Acesso em: 12/03/2017. Disponível em: https://speex.org.

[15] JAMIESON, David. *Speech Codecs and Associated PSQM Values*. *Test set 1*. Vocal Technologies. Acesso em 23/04/2017. Disponível em: https://www.vocal.com/speech-coders/associated-psqm-values

[16] KARAPANTAZIS, Stylianos. PAVLIDOU, Fotini-Niovi. *VoIP: A comprehensive survey on a promising technology*. Computer Networks. Elsevier, 2009.

[17] VIEIRA, Vinícius J. Dias. *Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência*. IFPB. Dissertação de mestrado, 2014.

[18] YAMASAKI, R. *et al*. "Correspondência entre escala analógico-visual e a escala numérica na avaliação perceptivo-auditiva de vozes". 16º Congresso Brasileiro de Fonoaudiologia. Campos do Jordão-SP, 2008.

[19] KEMPSTER, G. B. *et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol*. Americ J of Speech-Lang Pathol, 18: 124-32, 2009.

[20] OATES, J. *Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions*, Folia Phoniatr Logop.,61:49-56, 2009.

[21] PARSA, V., JAMIESON, D. G.. *Identification of pathological voices using glottal noise measures.* J Speech Lang Hear Res, 43: 469-85, 2000.

[22] BHUTA, T., PATRICK, L., GARNETT, J. *Perceptual evaluation of voice quality and its correlation with acoustic measurements*. J Voice, 18: 299-304, 2004.

[23] YANAGIHARA, N. *Significance of harmonic changes and noise components in hoarseness*. Journal of Speech and Hearing Research. 1967; 30: 431-541, 1967.

[24] LOPES, L. W., SILVA, H. F., EVANGELISTA, D. S., SILVA, J. D., SIMÕES, L. B., SILVA, P. O. C., SILVA, M. F. B. L., ALMEIDA, A. A. F. *Relação entre os sintomas vocais, intensidade do desvio vocal e diagnóstico laríngeo em pacientes com distúrbios de voz.* Revista CODAS, vol 10, pp.1782-2317, 2015.

[25] BARAVIEIRA, Paula Belini et al*. Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica*. CoDAS, 2016.

[26] ENGLERT, M. *et al*. "Erro Perceptivo-Auditivo de Vozes Humanas e Sintetizadas". Congresso Brasileiro de Fonoaudiologia, 2016.