



**ANSELMO DE VASCONCELOS CAVALCANTE**

**ANÁLISE DOS EFEITOS DE CODECS DE ÁUDIO NA  
AVALIAÇÃO DE DESVIOS VOCAIS**

João Pessoa - PB  
Março, 2018

**ANSELMO DE VASCONCELOS CAVALCANTE**

**ANÁLISE DOS EFEITOS DE CODECS DE ÁUDIO NA  
AVALIAÇÃO DE DESVIOS VOCAIS**

Dissertação apresentada à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba como requisito necessário à obtenção do grau de Mestre em Engenharia Elétrica.

Orientadora: Prof. Dra. Suzete Élide Nobrega  
Correia

Coorientadora: Prof. Dra. Silvana Luciene do  
Nascimento Cunha Costa

João Pessoa - PB  
Março, 2018

Dados Internacionais de Catalogação na Publicação – CIP  
Biblioteca Nilo Peçanha – IFPB, *Campus* João Pessoa

C376a Cavalcante, Anselmo de Vasconcelos.  
Análise dos efeitos de CODECS de áudio na avaliação de desvios vocais / Anselmo de Vasconcelos Cavalcante. – 2018.  
89 f. : il.

Dissertação (Mestrado em Engenharia Elétrica) – Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB / Coordenação de Pós-Graduação em Engenharia Elétrica, 2018.  
Orientador: Prof<sup>o</sup> Paulo Henrique da Fonseca Silva.  
Coorientador: Prof<sup>o</sup> Elder Eldervitch Carneiro de Oliveira

1. Engenharia de comunicação elétrica. 2. Voip. 3. CODECS. 4. Asterisk. 5. Avaliação da qualidade vocal. 6. Telemedicina. I. Título.

CDU 621.391

Ivanise Andrade M. de Almeida  
Bibliotecária-Documentalista  
CRB-15/0096

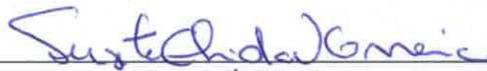
**ANSELMO DE VASCONCELOS CAVALCANTE**

**ANÁLISE DOS EFEITOS DE CODECS DE ÁUDIO NA  
AVALIAÇÃO DE DESVIOS VOCAIS**

Dissertação apresentada à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba como requisito necessário à obtenção do grau de Mestre em Engenharia Elétrica.

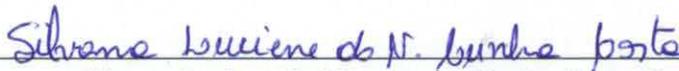
Dissertação de Mestrado defendida e aprovada em 05 / 03 / 2018.

BANCA EXAMINADORA



Suzete Élide Nobrega Correia, Dra. - IFPB

**Orientadora**



Silvana Luciene do Nascimento Cunha Costa, Dra. - IFPB

**Coorientadora**



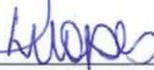
Adaildo Gomes D'Assunção Junior, Dr. - IFPB

**Examinador Interno**



Joabson Nogueira de Carvalho, Dr. - IFPB

**Examinador Interno**



Leonardo Wanderley Lopes, Dr. - UFPB

**Examinador Externo**

João Pessoa - PB

Março, 2018

*Aos meus pais, Maria do Socorro e José Anselmo.  
A minha esposa Carolina.*

## **AGRADECIMENTOS**

- A Deus, que me proporcionou o dom da vida;
- A minha família, especialmente meus pais, que se doaram intensamente para que sempre buscasse meus objetivos;
- A minha esposa, que soube entender com extrema maestria as minhas dificuldades e emoções durante a execução deste trabalho, sem nunca deixar de me incentivar;
- A todos aqueles que um dia foram meus professores, em especial a Suzete Correia, Silvana Costa, Michel Dias e Leonardo Lopes, que me deram total apoio para o desenvolvimento desta pesquisa. Sem a ajuda deles não seria possível sua realização;
- A todos os meus colegas de trabalho e de mestrado do IFPB, que me apoiaram desde o início desta caminhada até aqui.

*“Não sabendo que era impossível,  
ele foi lá e fez.”*  
(Jean Cocteau)

## RESUMO

Este trabalho apresenta um estudo sobre as implicações no uso de diferentes codecs de áudio na análise perceptiva e acústica 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. Para este propósito, foram utilizados 36 sinais de vozes sintetizadas e 36 sinais de vozes reais, classificados como normais, com o desvio rugosidade e com o desvio soproidade. Cada sinal foi submetido a seis transmissões, cada uma delas utilizando um codec específico (G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16). Antes e após cada transmissão, um especialista em voz realizou a classificação dos sinais quanto ao tipo de desvio e, com auxílio do software VoxMetria, extraiu-se as medidas acústicas frequência fundamental, *jitter*, *shimmer*, GNE e desvio padrão da frequência fundamental. Observou-se que, dentre os codecs analisados, o Opus24 foi aquele se mostrou o mais promissor para avaliação da qualidade vocal, por ter sido o único a apresentar taxa de acurácia acima de 70%, tanto para os sinais sintetizados quanto para os sinais reais analisados. Este codec conseguiu manter os desvios vocais inalterados para avaliação em mais de 85% dos sinais reais.

Palavras-chave: codecs, VoIP, avaliação da qualidade vocal, Asterisk, telemedicina.

## ABSTRACT

This research presents a study on the implications of different audio codecs on perceptual and acoustic voice analysis. A VoIP-based transmission scenario was created, using Asterisk and the Microsip softphone, to aid in the diagnosis of vocal deviations in distance. For this purpose, 36 synthesized voices signals and 36 real voice signals were used, classified as normal, with the roughness deviation and with the breathiness deviation. Each signal was submitted to six transmissions, using a specific codec (G.711 A Law, Speex32, GSM Full Rate, LPCM16, Opus24 and SILK16). Before and after each transmission, a voice specialist performed the classification of the signals regarding the type of deviation and, using the VoxMetria software, the fundamental frequency, jitter, shimmer, GNE and fundamental frequency standard deviation acoustic measurements were extracted. It was observed that, among the codecs analyzed, Opus24 was the most promising for voice quality evaluation, since it was the only one to present an accuracy rate above 70% for both, the synthesized signals and the real signals. This codec was able to keep the vocal deviations unaltered for evaluation in more than 85% of the real signals.

Keywords: codec, VoIP, voice quality evaluation, Asterisk, telemedicine.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Processo de conversão analógico/digital e transmissão da voz .....	25
Figura 2 – Filtro passa-baixa.....	26
Figura 3 – Processo de amostragem e reconstrução.....	27
Figura 4 – Processo de quantização de um sinal .....	28
Figura 5 – Comportamento do sinal de voz após cada etapa.....	29
Figura 6 – Campos do cabeçalho do protocolo RTP.....	37
Figura 7 – Processo de encapsulamento do pacote RTP para transmissão .....	38
Figura 8 – Estabelecendo uma chamada SIP .....	39
Figura 9 – Comunicação gerenciada pelo Asterisk utilizando protocolos distintos .....	41
Figura 10 – Arquitetura do Asterisk .....	41
Figura 11 – Escala analógico-visual (EAV) .....	44
Figura 12 – Modelo geral da metodologia empregada .....	47
Figura 13 – Cenário utilizado no estudo .....	51
Figura 14 – Valores de <i>jitter</i> dos sinais sintetizados .....	57
Figura 15 – Valores de <i>shimmer</i> dos sinais sintetizados .....	58
Figura 16 – Valores de GNE dos sinais sintetizados .....	59
Figura 17 – Valores da frequência fundamental dos sinais sintetizados .....	60
Figura 18 – Valores do desvio padrão da frequência fundamental dos sinais sintetizados .....	61
Figura 19 – Valores de <i>jitter</i> dos sinais reais .....	62
Figura 20 – Valores de <i>shimmer</i> dos sinais reais.....	64
Figura 21 – Valores de GNE dos sinais reais.....	65
Figura 22 – Valores da frequência fundamental dos sinais reais.....	66
Figura 23 – Valores do desvio padrão da frequência fundamental dos sinais reais .....	67
Figura 24 – Espectrogramas (tempo x frequência) do sinal real 9 antes da transmissão e após as transmissões .....	68

## LISTA DE TABELAS E QUADROS

Tabela 1 - Escala <i>mean opinion score</i> (MOS) .....	31
Tabela 2 - Tipos de codec GSM .....	32
Tabela 3 - Comparativo entre os codecs utilizados.....	35
Tabela 4 - Métodos utilizados pelo SIP .....	39
Tabela 5 - Características dos elementos empregados no cenário .....	52
Tabela 6 - Classificação dos sinais sintetizados pós-transmissão – matriz de confusão .....	53
Tabela 7 - Acurácia, especificidade e sensibilidade dos codecs analisados (sinais sintetizados) .....	54
Tabela 8 - Classificação dos sinais reais pós-transmissão – matriz de confusão .....	55
Tabela 9 - Acurácia, especificidade e sensibilidade dos codecs analisados (sinais reais) .....	55
.....	
Quadro 1 - Resumo dos resultados .....	73

## LISTA DE SIGLAS

ASHA	<i>American Speech-Language and Hearing Association</i> - Associação Americana de Fonoaudiologia e Audição
LSVT	<i>Lee Silverman Voice Treatment</i> - Tratamento de voz Lee Silverman
VoIP	<i>Voice over IP</i> - Voz sobre IP
IP	<i>Internet protocol</i> – Protocolo da Internet
LSVT-X	<i>Extended Lee Silverman Voice Treatment</i> - Tratamento de voz estendido de Lee Silverman
ADPCM	<i>Adaptive PCM</i> – PCM adaptativo
AMR	<i>Adaptive multi-rate</i> - Multi-taxa adaptativa
BER	<i>Bit Error Rate</i> – Taxa de erro de bit
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CODEC	Codificadores e decodificadores
DTN	<i>Delay-tolerant networking</i> – Rede tolerante a atraso
EAV	Escala analógico-visual
ECG	Eletrocardiograma
EFR	Enhanced full-rate
ESTI	<i>European Telecommunications Standards Institute</i> - Instituto Europeu de Normas de Telecomunicações
FLAC	<i>Free Lossless Audio Codec</i> – Codec de áudio gratuito sem perdas
GNE	<i>Glottal Noise Excitation</i> – Excitação de Ruído Glotal
GSM	<i>Global System for Mobile communication</i> - Sistema global de comunicação móvel
IAX	<i>Inter Asterisk eXchange</i>
IETF	<i>Internet Engineering Task Force</i> – Força tarefa de engenharia da Internet
iLBC	<i>Internet Low Bitrate Codec</i> - Codec de Baixa taxa de bit da Internet
ITU-T	<i>International Telecommunication Union</i> - União Internacional de Telecomunicações
LIEV	Laboratório Integrado de Estudos da Voz
LPC	<i>Linear prediction coefficients</i> - Coeficientes de predição linear
MCTI	Ministério da Ciência, Tecnologia e Inovação
MDVP	<i>Instrumental Multi Dimensional Voice Program</i> - Programa de voz multidimensional instrumental
MOS	<i>Mean opinion score</i> – Pontuação média de opinião

MP3	MPEG-1/2 Audio Layer 3
MPEG	<i>Moving Picture Experts Group</i> - Grupo de especialistas em imagens móveis
PAM	Pulsos modulados em amplitude
PCM	<i>Pulse-code modulation</i> - Modulação por Código de Pulso
QoS	<i>Quality of service</i> – Qualidade de serviço
RFC	<i>Request for comments</i> - Pedido de comentários
RNP	Rede Nacional de Ensino e Pesquisas
RTCP	<i>Real-time transport control protocol</i> – Protocolo de controle de transporte em tempo real
RTP	<i>Real-time transport protocol</i> – Protocolo de transporte em tempo real
RUTE	Rede Universitária de Telemedicina
SIP	<i>Session initiate protocol</i> – Protocolo de inicialização de sessão
SRTP	<i>Secure Real-time Transport Protocol</i> – Protocolo seguro de transporte de tempo real
TI	Tecnologia da informação
UDP	<i>User Datagram Protocol</i> - Protocolo de datagrama de usuário
UFPB	Universidade Federal da Paraíba
VAD	<i>Voice activity detection</i> – Detecção de atividade de voz
WAV	<i>Waveform Audio File Format</i> – Formato de arquivo de áudio em forma de onda
WMA	<i>Windows Media Audio</i> – Arquivo de áudio do Windows

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>15</b>
1.1. Estado da arte .....	17
1.2. Objetivos .....	19
1.2.1. Objetivo geral .....	19
1.2.2. Objetivos específicos .....	19
1.3. Organização do texto .....	20
<b>2. A TELESSAÚDE .....</b>	<b>21</b>
2.1. Vantagens da telessaúde .....	22
2.2. A telessaúde no Brasil .....	22
<b>3. DIGITALIZAÇÃO E TRANSMISSÃO DA VOZ .....</b>	<b>25</b>
3.1. A digitalização da voz .....	25
3.1.1. Filtragem .....	26
3.1.2. Amostragem .....	26
3.1.3. Quantização .....	27
3.1.4. Codificação .....	29
3.2. A compressão de áudio e os codecs .....	30
3.2.1. GSM .....	31
3.2.2. G.711 .....	32
3.2.3. LPCM .....	33
3.2.4. Speex .....	33
3.2.5. SILK .....	34
3.2.6. Opus .....	34
3.2.7. Comparação entre os codecs .....	35
3.3. Transmissão da voz com VoIP .....	35
3.3.1. Qualidade do áudio em VoIP .....	36
3.3.2. Os protocolos RTP e RTCP .....	37
3.3.3. O protocolo SIP .....	38
3.4. Asterisk .....	40
<b>4. AVALIAÇÃO DE DESVIOS VOCAIS .....</b>	<b>43</b>
4.1. Análise perceptivo-auditiva da voz .....	43
4.2. Análise acústica da voz .....	44

4.2.1. Frequência fundamental ( $F_0$ ) .....	45
4.2.2. <i>Jitter</i> .....	45
4.2.3. <i>Shimmer</i> .....	46
4.2.4. <i>Glottal Noise Excitation</i> (GNE) .....	46
<b>5. MATERIAS E MÉTODOS .....</b>	<b>47</b>
5.1. Metodologia .....	47
5.2. Análise perceptivo-auditiva .....	48
5.3. Análise acústica .....	49
5.4. Base de dados utilizadas e seleção dos sinais .....	49
5.4.1. Base de dados de sinais sintetizados .....	49
5.4.2. Base de dados de sinais reais .....	50
5.5. Cenário de transmissão .....	51
<b>6. RESULTADOS E DISCUSSÕES .....</b>	<b>53</b>
6.1. Resultados da análise perceptivo-auditiva .....	53
6.1.1. Sinais sintetizados .....	53
6.1.2. Sinais reais .....	54
6.2. Resultados da análise acústica .....	56
6.2.1. Sinais sintetizados .....	56
6.2.2. Sinais reais .....	62
6.2.3. Análise de espectrogramas .....	68
6.3. Desempenho individual dos codecs .....	71
6.3.1. Codec G.711 Lei A .....	71
6.3.2. Codec Speex32 .....	72
6.3.3. GSM <i>Full Rate</i> .....	72
6.3.4. LPCM16 .....	72
6.3.5. Opus24 .....	72
6.3.6. SILK16 .....	73
<b>7. CONSIDERAÇÕES FINAIS .....</b>	<b>74</b>
7.1. Contribuições da pesquisa .....	75
7.2. Sugestões para trabalhos futuros.....	75
<b>REFERÊNCIAS .....</b>	<b>76</b>
<b>APÊNDICE A - Medidas acústicas sinais sintetizados .....</b>	<b>80</b>
<b>APÊNDICE B - Medidas acústicas sinais reais .....</b>	<b>85</b>

## 1. INTRODUÇÃO

A voz revela informações socioemocionais do indivíduo, sendo considerada muito importante nas relações humanas e profissionais (CIELO *et al*, 2015). Profissionais que utilizam a voz como instrumento de trabalho a exemplo de professores, cantores, operadores de telemarketing, entre outros, bem como pessoas com hábitos sociais como tabagismo, alcoolismo e abuso vocal, podem ter a qualidade vocal alterada.

Patologias na laringe, sejam de origem neurológica ou orgânica, também 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 e seu monitoramento requer sessões periódicas com profissional qualificado (fonoaudiólogo), nem sempre disponível próximo ao local onde o paciente reside. Dessa forma, o tratamento é prejudicado devido às longas distâncias. Como exemplo, pode-se citar a Doença de Parkinson, em que o método mais eficaz (*Lee Silverman Voice Treatment - LSVT*) para tratar esta doença requer uma terapia intensiva, com duração de até dois meses, distribuída em 16 sessões aplicadas de duas a quatro vezes por semana (DIAS *et al.*, 2016). A Doença de Parkinson é classificada como a segunda mais comum dentre as doenças neurodegenerativas (LAU; BRETALAR, 2006). Ela está presente em aproximadamente 1% da população acima de 65 anos e em 0,1% da população geral. No Brasil, estima-se que 0,3% da população em geral seja acometida pela Doença de Parkinson e cerca de 3,3% dos indivíduos com mais de 65 anos de idade (DIAS *et al.*, 2016).

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 (CHERNEY; VUUREN, 2012).

Nos últimos anos, pesquisadores começaram a investigar a possibilidade de usar a telessaúde para fornecer serviços que auxiliem no tratamento de patologias da fala e linguagem, objetivando alcançar um número maior de pacientes a custos reduzidos (MASHIMA *et al*, 1999). Ela tem sido apontada como uma forma potencial de melhorar os cuidados de saúde nas zonas rurais e de difícil acesso (MARTÍNEZ *et al*, 2015).

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 como *smartphones* e *tablets* e pode ser desenvolvido de forma relativamente barata, usando a Internet (LAMBRINOS, 2008).

Embora a ideia de enviar áudio pela Internet já exista pelo menos desde a década de 70, somente por volta do ano 2000 é que o tráfego de áudio em tempo real cresceu mais intensamente. Duas coisas aconteceram para permitir esse crescimento. Primeiro, os computadores tiveram sua capacidade de processamento ampliada e passam a ser equipados com microfones de modo que podem inserir, processar e enviar dados de áudio com facilidade. Segundo, a largura de banda da Internet aumentou significativamente. Devido a essas melhorias as empresas viram a oportunidade de transportar o tráfego de voz usando a largura de banda da Internet e com isso reduzir suas contas telefônicas (TANENBAUM; WETHERAL, 2011). Novas companhias telefônicas surgiram e começaram a oferecer um modo barato de realizar chamadas de voz tradicionais através da Internet, ou seja, chamadas VoIP.

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 de taxa de transmissão, causando a perda de informações (TANENBAUM; WETHERAL, 2011), podendo levar a um diagnóstico errado da presença/ausência do desvio, bem como de sua intensidade. Dessa forma, é relevante avaliar a influência de diferentes codecs de áudio a fim de avaliar quais mudanças ocorrem nos sinais de voz e as consequentes alterações no diagnóstico dos distúrbios vocais.

O presente trabalho descreve um estudo sobre os efeitos dos codecs G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16 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. Os sinais transmitidos e avaliados foram obtidos de duas bases de dados, uma de vozes sintetizadas e outra de vozes reais. Uma avaliação perceptivo-auditiva foi realizada por um fonoaudiólogo que julgou o grau dos desvios vocais sopro e rugosidade, bem como sua intensidade, antes e depois das transmissões. Além disso, utilizou-se o *software* VoxMetria como ferramenta para extração das métricas acústicas *jitter*, *shimmer*, *Glottal Noise Excitation* (GNE), frequência fundamental e desvio padrão da frequência fundamental, antes e após as transmissões.

## 1.1. Estado da arte

Entre os trabalhos pesquisados relacionados ao contexto da telessaúde aplicada a reabilitação vocal, apresenta-se o de Dias *et al.* (2016), que investigou a eficiência da telereabilitação da voz em pacientes com Doença de Parkinson. Vinte pacientes com queixa na voz e acometidos por essa enfermidade participaram da pesquisa. A telereabilitação em tempo real ocorreu a partir de videoconferência e utilizando o método estendido de *Lee Silverman Voice Treatment (LSVT-X)*<sup>1</sup>. Os pacientes foram avaliados antes e depois dessa intervenção por meio de análise perceptual da qualidade vocal. No final da intervenção, todos responderam um questionário sobre a experiência com a telereabilitação. As análises revelaram diminuição na magnitude das alterações da qualidade da voz após a intervenção e indicou melhoria do padrão vocal. Todos os pacientes relataram satisfação e preferência pela telereabilitação em comparação com a reabilitação presencial. Os resultados sugeriram que a telereabilitação é uma intervenção eficiente para o tratamento dos problemas relacionados à qualidade da voz e a Doença de Parkinson e pode ser indicada para pacientes com acesso a tecnologias e dificuldades no alcance de profissionais ou centros especializados.

Outro trabalho que se pode citar é o de Zhu *et al.* (2010). Nele simulou-se uma comunicação VoIP baseada no codec G.729 para determinar os efeitos deste sistema sobre parâmetros de perturbação acústica de sinais de voz normais e patológicos. Cinquenta gravações de vozes normais e 48 gravações de vozes patológicas, afetadas pela paralisia laríngea, foram transmitidas através de um sistema de comunicação VoIP. Os programas de análise acústica de CSpeech e MDVP foram utilizados para determinar a porcentagem de *jitter* e *shimmer* das amostras de voz antes e depois da transmissão. Os efeitos nas medidas de perturbação ocasionados pelos codecs de áudio MP3, WMA e FLAC (utilizados para armazenamento em disco) foram estudados. Como resultado, verificou-se que a transmissão VoIP aumenta a porcentagem de *jitter* e *shimmer* nas amostras de voz. No entanto, após a transmissão VoIP, ainda era possível a discriminação significativa entre vozes normais e patológicas afetadas pela paralisia laríngea, não alterando a classificação inicial. Verificou-se que o método de compressão sem perdas FLAC não exerce qualquer influência sobre as medidas de perturbação. Os métodos de compactação com perdas MP3 e WMA aumentam a porcentagem de *jitter* e *shimmer*. Como conclusão, o estudo valida a viabilidade do VoIP no desenvolvimento de sistemas de coleta e avaliação de dados de sinais de voz remotos.

---

<sup>1</sup> Programa de tratamento para as alterações da voz direcionado aos pacientes com Doença de Parkinson desenvolvido por grupo de fonoaudiólogos americanos.

O estudo de Scholl, Lambrinos e Lindgren (2009) afirma que VoIP é uma solução para apoiar a Telemedicina em clínicas de saúde rural em países em desenvolvimento e apresenta uma abordagem para a criação de redes de Telemedicina escalonáveis e tolerante a atrasos utilizando as redes antigas de telefonia (redes DTN). O trabalho indica que isso possibilitaria a criação de redes de Telemedicina para o compartilhamento de teleconsulta e outras informações médicas entre um grande número de locais em áreas que não podem ser atendidas por soluções existentes. O modelo de rede proposto sugere que comunicações VoIP sejam estabelecidas através de uma rede DTN, através de uma espécie de túnel, utilizando telefones VoIP e *softphones*, se comunicando com um servidor VoIP através do protocolo SIP. Os sistemas que utilizariam essa rede proposta poderiam ser desenvolvidos com base principalmente em *software* de código aberto e poderiam ser executados em *hardware* razoavelmente barato, a fim de manter os custos reduzidos. Afirma ainda que esses sistemas poderiam proporcionar benefícios à saúde rural e que sua abordagem vale a pena ser investigada e desenvolvida.

No trabalho desenvolvido por Ferrari *et al.* (2014) 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 máquinas 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.

Das (2010) desenvolveu uma metodologia para a criação de aplicações de *software* para a comunidade rural utilizando o princípio de pesquisa-ação como melhores práticas que ajudaram a abordar todos os problemas que afetam o sucesso da aplicação. A validação da metodologia é realizada através da utilização de um protótipo de sistema para consultas remotas em saúde, desenvolvido para comunidades rurais de Bangladesh. Este protótipo foi baseado na tecnologia VoIP e permitiu a comunicação entre um pequeno hospital localizado em uma aldeia e um hospital maior localizado na aldeia vizinha. Como conclusão é proposto que a tecnologia VoIP seja utilizada em Bangladesh para promover o desenvolvimento tecnológico em suas áreas rurais sub-atendidas.

Na avaliação da qualidade vocal, o estudo de Lopes, Cavalcante e Costa (2014) buscou analisar a correlação entre medidas acústicas e a intensidade do desvio vocal.

Participaram da pesquisa 186 pacientes disfônicos. O desvio vocal foi avaliado por consenso por três fonoaudiólogos através de uma análise perceptivo-auditiva, utilizando-se uma escala analógico-visual. Foram extraídas as medidas frequência fundamental, desvio padrão da frequência fundamental, *jitter*, *shimmer* e *Glottal to Noise Excitation Ratio* (GNE). Como resultado, a frequência fundamental correlacionou-se negativamente com o grau de rugosidade. O desvio padrão da frequência fundamental correlacionou-se positivamente com o grau de rugosidade. Houve correlação positiva entre o *jitter* e *shimmer* e todos os parâmetros perceptivos. Apenas o GNE diferenciou vozes saudáveis de vozes desviadas e classificou o grau de sopro. O *shimmer* separou vozes rugosas de vozes não rugosas. A frequência fundamental foi útil para diferenciar vozes rugosas das soprosas. Os autores concluíram que: há correlação entre as medidas acústicas e perceptivo-auditivas; o *shimmer* e o GNE podem ser utilizados na detecção de rugosidade e sopro, respectivamente; O GNE é útil para classificar o grau de sopro e a frequência fundamental diferenciou vozes rugosas e soprosas.

Em toda a literatura pesquisada, não foi encontrado nenhum trabalho que medisse a influência de diferentes codecs no processo de análise perceptivo-auditiva, além dos efeitos produzidos por eles em medidas acústicas. É exatamente esta lacuna que esta pesquisa pretende ajudar a preencher.

## **1.2. Objetivos**

### **1.2.1. Objetivo geral**

Analisar os efeitos produzidos por codecs de áudio na avaliação perceptivo-auditiva e acústica de vozes normais, soprosas e rugosas.

### **1.2.2. Objetivos específicos**

- Avaliar o desempenho de codecs de áudio na transmissão de sinais de vozes sintetizados;
- Avaliar o desempenho de codecs de áudio na transmissão de sinais de vozes reais;
- Avaliar o desempenho dos codecs em manter a qualidade vocal com o auxílio de um especialista em voz e também de um software computacional;
- Identificar o codec mais adequado para transmissão dos sinais de voz.

### 1.3. Organização do texto

O texto deste trabalho encontra-se organizado da seguinte forma:

- No Capítulo 2 é realizada uma explanação sobre a telemedicina, suas vantagens e sua situação no Brasil;
- No Capítulo 3 é realizada uma descrição do processo de digitalização e transmissão da voz, incluindo a exposição dos codecs utilizados (G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16). Ainda no Capítulo 3 é apresentada a tecnologia Voz sobre IP e seus principais elementos;
- No Capítulo 4 são apresentados os métodos utilizados na avaliação de desvios vocais, incluindo as principais métricas acústicas;
- No Capítulo 5 a metodologia empregada neste estudo é descrita, destacando a base de dados utilizada e o cenário desenvolvido;
- No Capítulo 6 são apresentados os resultados deste trabalho;
- Por fim, no Capítulo 7 são feitas as considerações finais, incluindo sugestões de ideias para pesquisas futuras.

## 2. A TELESSAÚDE

Neste capítulo é apresentado o conceito de telemedicina ou telessaúde (WEN, 2008), seus tipos de modalidade, as vantagens do seu uso e como ela está sendo utilizada no Brasil.

As tecnologias de comunicação vêm sendo empregadas na medicina há vários anos. Uma das primeiras referências à telemedicina na literatura médica foi em um artigo publicado em 1950, que descrevia a transmissão de imagens radiológicas por telefone (DEMIRIS, 2003). A partir da década de 1990, houve um crescimento no emprego da telemedicina, motivado principalmente pelo desenvolvimento de novas tecnologias de comunicação e pela diminuição no custo dos equipamentos utilizados na transmissão da informação (WOOTTON, 2001).

A telemedicina consiste no uso das tecnologias de telecomunicações e informática para a interação entre profissionais de saúde e pacientes, com a finalidade de realizar atos médicos, estando o médico e o paciente geograficamente distantes. As informações médicas de naturezas diversas são transmitidas através de meios eletrônicos de telecomunicações como: cabos, fibras óticas, satélites, etc., para a análise dessas informações médicas a longa distância pelos profissionais de saúde com a finalidade de dar suporte no atendimento ao paciente (COSTA; SANTOS, 2001). Ela também pode ser entendida como o emprego de sinais eletrônicos para transferir informações médicas (fotografias, imagens em radiologia, áudio, dados de pacientes, videoconferência) de um local a outro através da Internet, de computadores, de satélites ou de equipamentos de videoconferência, com a finalidade de melhorar o acesso à saúde (KIM, 2004).

Existem diversos projetos de Telemedicina desenvolvidos ou em desenvolvimento em vários países. Embora todos estes projetos tenham como objetivo comum diminuir o custo e facilitar o acesso de pacientes em ambientes distantes, ou mesmo de especialistas médicos de grandes centros, eles diferem em termos de implementação e funcionalidade. Na grande maioria dos trabalhos, inclusive no Brasil, a Telemedicina tem o foco voltado para a teleconsulta entre um médico e um paciente, para geração de laudos virtuais ou a realização de teleconferências entre dois ou mais médicos com o objetivo de analisar um caso médico complicado (MACEDO, 2008).

Além da teleconsulta, a telemedicina pode ser classificada em mais cinco modalidades (COSTA; SANTOS, 2001):

- Teleatendimento: pacientes são atendidos à distância para utilização de serviços de saúde como marcação de consultas, de procedimentos e informações gerais, através de telefone ou Internet.

- Telediagnóstico: consultas remotas são realizadas com troca de imagens, áudio, vídeo e texto para fins de diagnóstico, geralmente ocorrendo em tempo real.
- Telemonitoração: acompanhamento do estado de saúde de pacientes e seus sinais biológicos à distância.
- Telecirurgia: alguns procedimentos cirúrgicos podem ser realizados ou assessorados remotamente, através de sistemas de realidade virtual com sinais visuais, auditivos e tácteis entre o local onde está o médico e o local onde está o paciente, utilizando-se equipamento de manipulação remota de instrumentos.
- Teledidática: é a educação a distância associada ao uso da videoconferência e ao acesso a bancos de informações em saúde para o ensino e a educação continuada.

## **2.1. Vantagens da telessaúde**

A telessaúde é um recurso que contribui significativamente para a melhoria da qualidade da assistência médica, para a redução do tempo gasto entre o diagnóstico e a terapia e para a extensão dos serviços médicos especializados e de qualidade aos locais que não os possuem. Além dessas vantagens, podem ser citadas citar (MACEDO, 2008):

- Acesso rápido a especialistas em locais distantes nos casos de desastres e emergências;
- Uso mais efetivo de recursos, através da centralização de especialistas e a descentralização da assistência primária em saúde, alcançando um número maior de pessoas;
- Melhoria nas condições de diagnóstico através da cooperação de profissionais multidisciplinares com compartilhamento de informações médicas;
- Acesso remoto aos serviços de saúde nas várias modalidades, especialmente em áreas rurais;
- Redução no número de consultas presenciais, através da coleta eletrônica dos dados clínicos do paciente;
- Possibilidade de acesso ao prontuário eletrônico do paciente.

## **2.2. A telessaúde no Brasil**

O tipo de serviço ofertado relacionado à telemedicina varia de estado para estado. Em Minas Gerais o foco está na implantação do serviço de eletrocardiograma (ECG) à distância, em virtude do número de ocorrências de doenças cardiovasculares. No Rio de Janeiro, exames radiológicos de tórax são realizados em serviços de atenção primária para

em seguida serem digitalizados e enviados para avaliação de remota de radiologistas. Em Santa Catarina criou-se uma ampla rede de transmissão de exames, tornando possível que exames obtidos em hospitais do interior do estado sejam laudados por médicos especialistas em centros colaboradores. No âmbito da teleeducação é o estado de São Paulo que vem se destacando, através de iniciativas como o Projeto Homem Virtual, o Projeto Jovem Doutor, o Projeto Cybertutor e o Projeto Sala de Aula Interativa do Futuro. No Rio Grande do Sul, devido à grande ocorrência de doenças respiratórias, foi implantado o primeiro serviço no Brasil de telediagnóstico para doenças respiratórias crônicas, com o uso da telemedicina. Além disso, destaca-se ainda, neste estado, o serviço de ultrassonografia em unidade móvel, onde gestantes podem fazer ultrassonografia obstétrica monitorada em tempo real (MALDONADO, 2016).

Diversas iniciativas governamentais com o objetivo de desenvolver a telemedicina vêm ocorrendo no Brasil. Entre as principais delas estão às relacionadas ao Programa Telessaúde Brasil Redes, a Rede Nacional de Ensino e Pesquisas (RNP) e a Rede Universitária de Telemedicina (RUTE).

Criado em 2007 pelo Ministério da Saúde e posteriormente ampliado em 2011, o Programa Nacional Telessaúde Brasil Redes tem foco na atenção primária, integrando ensino e serviço por meio de ferramentas de tecnologias da informação, que oferecem condições para promover a Teleassistência e a Teleeducação. Ele e está presente em 11 estados, oferecendo teleconsultorias a aproximadamente 1.500 Unidades Básicas de Saúde. Entre os principais serviços oferecidos pelo Programa Telessaúde Brasil Redes, estão (SAÚDE, 2017):

- **Teleconsultoria:** Consulta com o objetivo de esclarecer dúvidas sobre procedimentos clínicos, ações de saúde e questões relativas ao processo de trabalho. Pode ser realizada em tempo real (síncrona) ou por meio de mensagens off-line (assíncrona).
- **Telediagnóstico:** Serviços de apoio ao diagnóstico onde os exames são realizados em uma localidade remota e enviados para emissão de laudo em outro local.
- **Segunda Opinião Formativa:** Resposta sistematizada a perguntas originadas pelas teleconsultorias. É construída com base em revisão bibliográfica, evidências científicas e clínicas.
- **Teleeducação:** Conferências, aulas e cursos ministrados por meio da utilização das tecnologias de informação e comunicação.

Visando à construção de uma infraestrutura nacional acadêmica de rede de Internet, foi criada em 1989 pelo Ministério da Ciência, Tecnologia e Inovação (MCTI) a RNP. Em

2006, a RNP lançou o projeto RUTE, para implantar infraestrutura de interconexão nos hospitais universitários e unidades de ensino de saúde. Atualmente, a RNP está presente em 27 Unidades da Federação e a RUTE, por sua vez, em 124 núcleos de telemedicina e telessaúde no Brasil (RNP, 2017) (RUTE, 2017).

Na iniciativa privada, observa-se um interesse crescente em telemedicina impulsionado pela perspectiva de ampliação do número de serviços prestados e redução de atendimentos e de internações. Verifica-se ainda o aumento do envolvimento de fornecedores de serviços e de produtos customizados, tais como aparelhos de comunicação, computadores, equipamentos de transmissão/recepção, além de *softwares*, para atender a uma demanda cada vez maior de novas e diversificadas aplicações voltadas para a telemedicina (MALDONADO, 2016).

Características do mercado brasileiro fazem do país um polo de atração de investimentos externos, oferecendo oportunidades significativas para o desenvolvimento de aplicações voltadas para a telemedicina, indicando uma boa perspectiva de crescimento para este mercado no futuro próximo (MALDONADO; MARQUES; CRUZ, 2016).

Diante do cenário favorável para o desenvolvimento de soluções em telemedicina no Brasil, acredita-se que o desenvolvimento de aplicações voltadas para o processo de triagem inicial de pacientes, quanto a presença ou ausência de desvios vocais, como também para o processo de reabilitação vocal, seja uma grande tendência dentro da modalidade de teleatendimento. Essas aplicações, quando executadas de maneira remota entre pacientes e profissionais da saúde, podem promover a descentralização da assistência, alcançando um número maior de pessoas e facilitando o atendimento, sobretudo em áreas rurais distantes e de difícil acesso.

### 3. DIGITALIZAÇÃO E TRANSMISSÃO DA VOZ

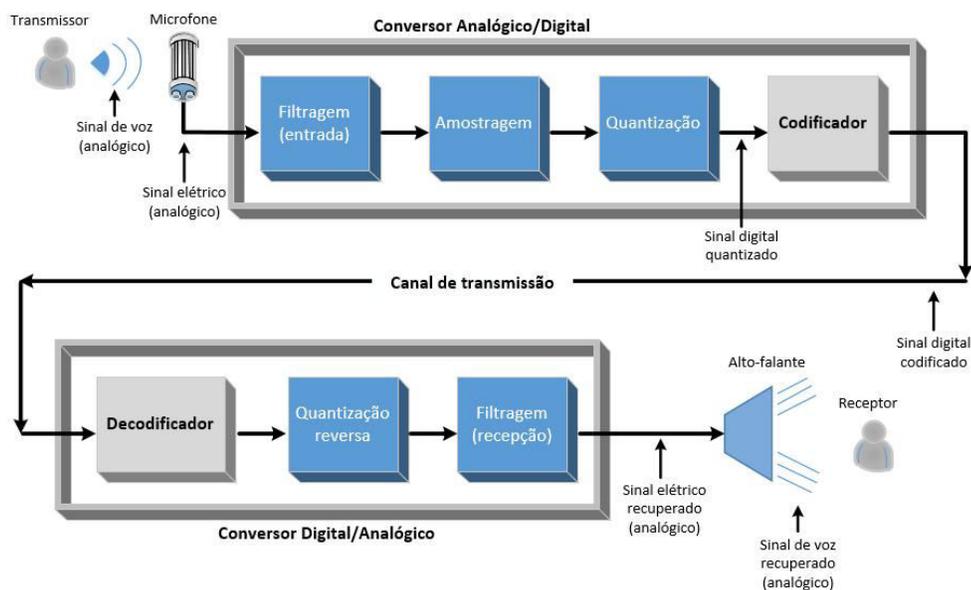
Neste capítulo são apresentadas as etapas percorridas por um sinal de voz analógico até que ele se torne um sinal de voz digital, estando apto para a transmissão através de uma rede de dados. Também é apresentado o conceito de codecs de áudio e as principais características daqueles utilizados neste estudo. Além disso, é abordado o processo de transmissão da voz utilizando-se a tecnologia Voz sobre IP (VoIP) e os principais protocolos envolvidos nesta transmissão.

#### 3.1. A digitalização da voz

A voz é a produção de sons que o ser humano faz usando as pregas vocais (COSTA, 2008). Ela se manifesta na natureza como uma forma de onda contínua e analógica. O áudio digital é uma representação binária da onda de áudio analógica que pode ser usada para recriá-lo (TANENBAUM; WETHERAL, 2011).

A Figura 1 ilustra o processo de codificação e decodificação da voz. O sinal de voz capturado pelo microfone é digitalizado (conversão analógico – digital) através das etapas de **filtragem, amostragem, quantização e codificação**. A conversão analógico-digital é quem permite que sistemas de comunicação digital transportem sinais analógicos. Após a transmissão, ocorre o processo inverso, em que o conversor digital – analógico, recebe o sinal digital e o converte para a forma analógica, enviando o sinal para os alto-falantes.

Figura 1 – Codificação e Decodificação da voz.



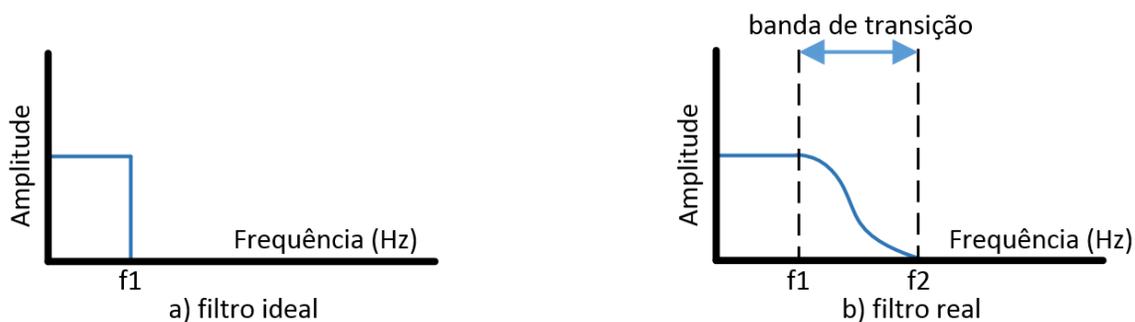
Fonte: Autoria própria.

### 3.1.1. Filtragem

Os sinais originais reais geralmente não são limitados em frequência, entretanto para que se possa recuperá-los completamente a partir da amostragem, é necessário empregar um processo de filtragem passa-baixas. Os filtros anti-*aliasing* tem como objetivo atenuar as componentes de alta frequência do sinal que não são essenciais para a informação contida nele (ROGER, 2012). A frequência de corte do filtro real é quem estabelece o valor da máxima frequência que o sinal deve possuir, limitando-o dessa forma em frequência.

A Figura 2 ilustra o processo de filtragem passa-baixas ideal (Figura 2a) e real (Figura 2b). O filtro ideal permite a passagem de todas as frequências do sinal desejáveis (abaixo de  $f_1$ ) e rejeita todas as frequências acima desse valor. No entanto, um filtro passa-baixas ideal não é fisicamente realizável. Na prática, os filtros reais deixam passar todas as frequências menores que  $f_1$ , e corta as frequências acima de  $f_2$ . A região entre  $f_1$  e  $f_2$ , conhecida como a banda de transição, contém uma atenuação gradual das frequências de entrada. Nesse caso, para que se possa recuperar o sinal sem perdas, é necessário considerar que a máxima frequência do sinal corresponde à frequência mais elevada na banda de transição.

Figura 2 – Filtro passa-baixas.



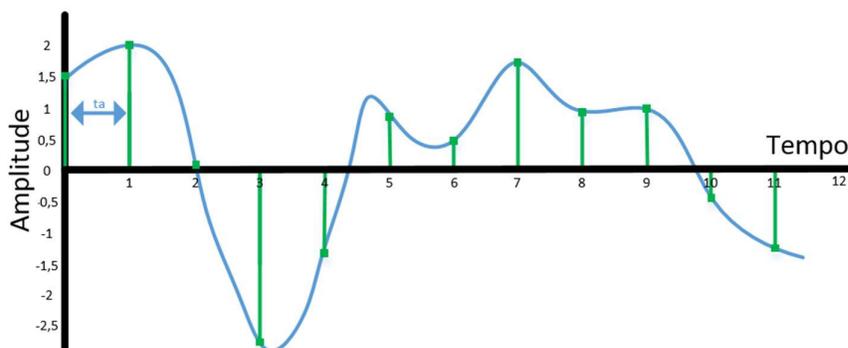
Fonte: Autoria própria.

### 3.1.2. Amostragem

Amostragem é o processo no qual se converte um sinal, representado por uma função contínua no tempo, em uma sequência numérica representada por uma função discreta no tempo. Pode ser definida como o processo de medição instantânea de valores de um sinal contínuo em intervalos regulares. O intervalo entre as amostras é determinado por um pulso de sincronismo e a sua frequência é chamada de taxa de amostragem ( $Ta$ ). A

amostragem constitui uma etapa primordial na geração de sinais digitais e está representada na Figura 3.

Figura 3 – Processo de amostragem.



Fonte: Autoria própria.

O Teorema de Nyquist ou Teorema da Amostragem (LATHI; DING, 2012) estabelece que um sinal  $x(t)$ , cujo espectro é limitado em banda a  $B$  Hz, ou seja,

$$X(f) = 0 \quad \text{para } |f| > B \quad (1)$$

pode ser reconstruído exatamente (sem qualquer erro) a partir de suas amostras em tempo discreto tomadas uniformemente a uma taxa  $fa$  amostras por segundo. A condição é que a frequência de amostragem para recuperação perfeita do sinal seja no mínimo igual a duas vezes a máxima frequência do sinal.

$$fa \geq 2B. \quad (2)$$

Caso a frequência de amostragem seja inferior a  $2B$ , haverá uma sobreposição no espectro, fenômeno conhecido como *aliasing*, que inviabiliza a correta recuperação do sinal. Para prevenir o *aliasing* deve-se remover ou filtrar as frequências acima da frequência mais alta desejada através de um filtro passa-baixas, como explicado na seção anterior.

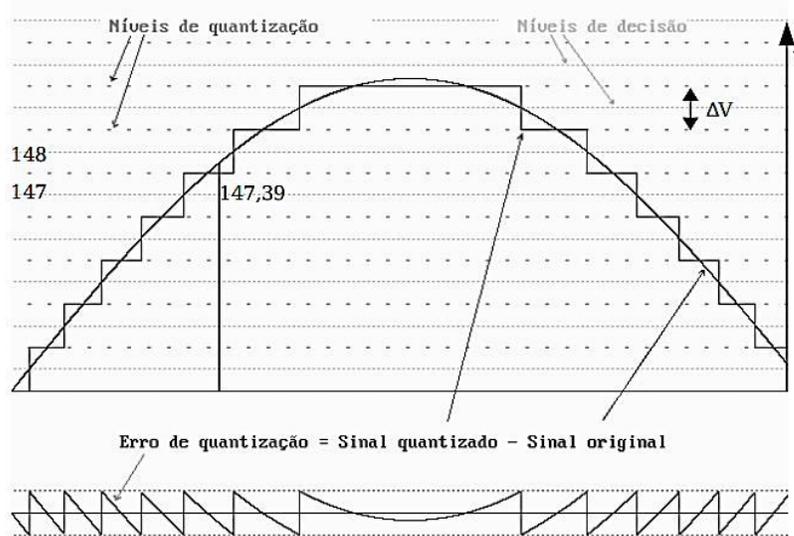
### 3.1.3. Quantização

Quantização é um processo no qual amostras individuais do sinal são arredondadas para o nível de tensão de referência mais próximo, quantizando a infinidade de valores

possíveis em uma quantidade finita que possa ser representada em bits, obtendo-se um sinal digital. O erro introduzido neste processo é conhecido como erro de quantização. A quantidade de níveis a ser utilizada depende do número de bits utilizado e o intervalo entre os níveis pode ser uniforme ou variar conforme a amplitude (MOECKE, 2006).

A Figura 4 mostra um exemplo de processo de quantização de um sinal. O  $\Delta V$  apresentado nesta imagem significa o saldo de amplitude (PINHEIRO, 2008).

Figura 4 – Processo de quantização de um sinal.



FONTE: PINHEIRO, 2008.

A audição humana, por exemplo, não é igualmente sensível às baixas e as altas amplitudes. Quanto maior for a amplitude do sinal, menos detalhada é a nossa percepção. Neste caso, é mais coerente quantizar com mais bits as baixas amplitudes em detrimento das maiores, afinal, o ruído de quantização seria mais inconveniente quando objetiva-se escutar com clareza alguma informação de pequena amplitude. Para grandes amplitudes o ruído poderia até passar despercebido (MARTINS *et al*, 2010).

Objetivando atender a esta característica da audição humana, foram propostas duas maneiras de se alocar bits de maneira não uniforme para representar sinais digitais: a Lei  $\mu$  (américa do norte e Japão) e a Lei A (Europa e américa do sul). A formulação destas leis se dá conforme abaixo:

$$\text{Lei A: } \Delta v = \frac{1 + \log(A\Delta s)}{1 + \log(A)}, \text{ sendo } 1/A \leq \Delta s \leq 1 \quad (3)$$

$$\text{Lei } \mu: \Delta v = \frac{\log(1 + \mu \Delta s)}{\log(1 + \mu)} \quad (4)$$

Em que:

- $\mu$  e  $A$  são constantes;
- $\Delta s$  é o intervalo de amplitude do sinal de entrada;
- $\Delta v$  é o comprimento do intervalo de quantização correspondente.

É comum as implementações em que  $\mu = 255$  e  $A=87,6$ , o que permite manter o ruído de quantização constante em toda a faixa dinâmica utilizando-se apenas 256 intervalos (8 bits).

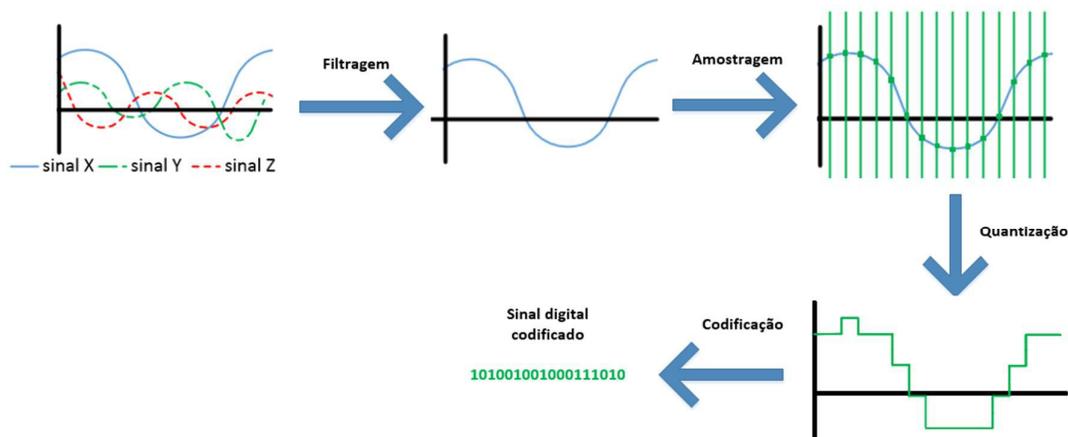
### 3.1.4. Codificação

A codificação é a operação que associa um determinado código binário para cada nível de quantização. A necessidade da codificação vem do fato de que caso os sinais quantizados fossem transmitidos diretamente, as amplitudes seriam facilmente distorcidas pelo meio de transmissão. Outra justificativa para o uso da codificação é que sem ela os circuitos de identificação seriam extremamente complexos, visto que teríamos pelo menos cerca de 100 níveis para transmitir um sinal de voz, por exemplo (MARTINS *et al*, 2010).

Utilizando o código binário, por exemplo, os pulsos são codificados por dois níveis de amplitude possíveis, expresso por 1 ou 0 o que simplifica muito os circuitos de reconhecimento destes sinais.

A Figura 5 mostra o comportamento do sinal (amplitude X tempo) ao longo do processo de digitalização.

Figura 5 – Comportamento do sinal de voz após cada etapa.



Fonte: Autoria própria.

Observa-se num primeiro instante que após o processo de filtragem os vários sinais existentes são resumidos a apenas um único sinal (sinal X). Na amostragem são obtidas amostras do sinal resultante da filtragem. Observa-se nesta fase que para cada instante de tempo indicado pelas linhas verdes uma amostra do sinal é obtida. No processo de quantização os valores de amostra obtidos do sinal são redefinidos de acordo com o algoritmo de quantização escolhido. Na codificação os valores quantizados são convertidos em linguagem binária (zero ou um).

De acordo com o número de amostras geradas por segundo e a quantidade de bits usada para representar cada uma das amostras, pode-se calcular a taxa de bits (*bit rate*) usada pelo codec. A taxa de bits é geralmente medida utilizando a unidade Mbit/s (Megabits por segundo), que representa o número de bits utilizado em 1 segundo de áudio. Nesse caso, a taxa de bits equivale apenas à codificação da voz, e não inclui o tráfego gerado por informações de controle utilizadas pelos demais protocolos VoIP (SOUZA *et al*, 2007).

### 3.2. A compressão de áudio e os codecs

O áudio normalmente é comprimido para reduzir a taxa de transmissão. 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 (TANENBAUM; WETHERAL, 2011).

Os codecs possuem duas assimetrias importantes. A primeira é que para muitas aplicações um arquivo de áudio só será codificado uma vez (quando for armazenado), mas poderá ser decodificado milhares de vezes (quando for reproduzido pelos ouvintes), isso significa que neste caso é aceitável que o algoritmo de codificação seja lento e exija um *hardware* robusto, desde que o algoritmo de decodificação seja rápido e não exija um *hardware* com alto poder de processamento. No caso de aplicações de tempo real, como é o caso do VoIP<sup>2</sup>, a codificação lenta é inaceitável, uma vez que ela precisa acontecer em tempo real. Conseqüentemente, tráfegos multimídia em tempo real utilizam codecs diferentes dos que são utilizados no armazenamento de áudio em disco.

A segunda assimetria é que o processo de codificação/decodificação não precisa ser reversível, ou seja, ao comprimir um arquivo de áudio, transmiti-lo e depois descomprimi-lo, o usuário não espera receber de volta o arquivo original, exatamente como ele era antes da transmissão. Para aplicações em tempo real, esse requisito não existe. Em geral, é aceitável ter um sinal de áudio após a codificação e depois decodificá-lo de forma ligeiramente

---

<sup>2</sup> Conjunto de padrões que permite que chamadas telefônicas sejam feitas por meio de uma conexão de banda larga com a Internet.

diferente do original, desde que a informação seja a mesma e percebida de forma inteligível pelo ouvinte (TANENBAUM; WETHERAL, 2011).

Quando a saída decodificada não é exatamente igual a 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 comumente usada para determinar a qualidade do som produzido pelos codecs é o *Mean Opinion Score* (MOS), que é um padrão definido pela *International Telecommunication Union* (ITU-T). Esta medida pode variar em uma escala de 1 a 5, onde 1 indica voz de baixa qualidade, enquanto um valor de 5 indica alta qualidade de voz, sem problemas perceptíveis. Maiores informações sobre os níveis de qualidade do MOS encontram-se na Tabela 1.

Tabela 1 - Escala *mean opinion score* (MOS).

Score	Definição	Descrição
5	Excelente	Um sinal de voz perfeito gravado em um local silencioso
4	Bom	Qualidade de uma chamada telefônica convencional
3	Razoável	Requer algum esforço na escuta
2	Pobre	Fala de baixa qualidade e difícil de entender
1	Ruim	Fala não clara, quebrada

Fonte: HOWARD; ANGUS, 2010 (adaptado).

Alguns dos codecs disponíveis tanto para aplicações de multimídia em tempo real, como por exemplo no VoIP, quanto para armazenamento de áudio em disco, são: o GSM, G.711, LPCM, Speex, SILK e Opus. Esses codecs são descritos a seguir.

### 3.2.1. GSM

O *Global System for Mobile communication* (GSM) é um padrão comercial que teve seu desenvolvimento iniciado em 1982 pelo *European Telecommunications Standards Institute* (ETSI) e em 1992 teve sua primeira versão comercial disponibilizada. Embora originalmente projetado para operação na faixa de 900 MHz, foi logo adaptado para 1800 MHz. Ao longo dos anos, a versatilidade do GSM resultou na adaptação das especificações a muitas mais bandas de frequência para atender a nichos de mercado (ETSI, 2017).

O padrão GSM suporta quatro tecnologias de compressão diferentes para analisar e compactar a fala: *full-rate*, *enhanced full-rate* (EFR), *adaptive multi-rate* (AMR) e *half-rate*. Apesar de todas as quatro tecnologias apresentarem perda, isto é, alguns dados são

perdidos durante a compressão, estes codecs foram otimizados para regenerar com precisão a fala na saída de uma ligação (MESTON, 2003).

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 bits de saída do codec depende do seu tipo (Tabela 2), com um intervalo entre 4,75 kbit/s e 13 kbit/s (MESTON, 2003). O codec GSM utilizado neste trabalho foi o GSM *Full Rate*.

Tabela 2 – Tipos de codec GSM.

Codec	Taxa de bit (bit/s)
GSM <i>Full rate</i>	13
GSM <i>Enhanced Full rate</i>	12,2
GSM <i>Half rate</i>	5,6
GSM AMR	4,75 - 12,2

Fonte: Baseado em MESTON, 2003.

### 3.2.2. G.711

O G.711, também é conhecido como *Pulse Code Modulation* (PCM), é um codec que foi 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 (UNION, 1988).

Existem duas versões do codec G.711: Lei A e Lei  $\mu$ . A Lei  $\mu$  é usada na América do Norte e Japão e a Lei A é usada no resto do mundo. A diferença entre elas está no método utilizado na amostragem do sinal analógico. Em ambas, o sinal não é amostrado linearmente, mas de forma logarítmica. A Lei A fornece uma faixa mais dinâmica, ao contrário da Lei  $\mu$  (RCHANDRA, 2013).

Usar G.711 para VoIP pode oferecer uma melhor qualidade de voz, uma vez que este mesmo codec é usado pelas redes públicas de telefonia mais antigas. 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 utilizá-lo está no fato de necessitar de mais taxa de transmissão do que outros codecs, podendo chegar até 84 kbit/s. O G.711 é suportado pela maioria dos provedores de VoIP (RCHANDRA, 2013). Neste trabalho utilizou-se o G.711 Lei A.

### 3.2.3. 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 (algoritmo de Lei A ou o algoritmo de Lei  $\mu$ ). O LPCM é o padrão utilizado na codificação do áudio presente no DVD (desde 1995), Blu-ray (desde 2006) e HDMI (desde 2002). As suas taxas de transmissão mais comuns são 8, 16, 20 ou 24 bits por segundo, podendo chegar até a 6.144 Mbit/s. As taxas de amostragem mais comuns são 48 kamostras/s, usada em formato de DVD, ou 44,1 kamostras/s, usada em CD. As frequências de amostragem de 96 kamostras/s ou 192 kamostras/s podem ser utilizadas em alguns equipamentos mais novos (MONTGOMERY, 2012). Neste trabalho foi utilizado o codec LPCM16, que possui taxa de amostragem de 16 kamostras/s.

### 3.2.4. Speex

Speex é um formato de codificação de áudio, onde ocorre perda de dados, criado em 13 de fevereiro de 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) (XIPH, 2017).

O objetivo dos projetistas foi desenvolver um codec 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 (32 kamostras/s), banda larga (16 kamostras/s) e banda estreita (8 kamostras/s). Ele também foi desenvolvido para ser robusto em relação a pacotes perdidos, mas não para os corrompidos. As principais características do Speex podem ser resumidas da seguinte forma (XIPH, 2017):

- Banda estreita (8 kamostras/s), banda larga (16 kamostras/s) e ultra-banda larga (32 kamostras/s) no mesmo fluxo de bits;
- Ocultação de perda de pacotes;
- Operação de taxa de bits variável (VBR), de 2 kbit/s a 44 kbit/s;
- Detecção de atividade de voz (VAD);
- Cancelamento de eco;
- Supressão de ruído.

Neste estudo foi utilizado o codec Speex32 (32 kamostras/s).

### 3.2.5. SILK

O SILK é um codec desenvolvido pela Skype para ser utilizado em VoIP como substituto do codec SVOPC. Apesar de patenteado, seu código foi disponibilizado para uso público em 2009. Ele opera em quatro taxas de amostragem diferentes: banda estreita de 8 kamostras/s, banda média de 12 kamostras/s, banda larga de 16 kamostras/s e super banda larga de 24 kamostras/s. O codec SILK possui detecção automática de áudio em banda larga e pode diminuir a largura de banda se não houver benefícios para a utilização de frequências mais altas. Possui taxa de bits variável que pode ser configurada entre 6 e 40 kbps. Quando opera a com taxas de bits maiores, proporciona uma melhor qualidade de som. Seu algoritmo causa atraso de 25 ms durante a compressão do áudio. Este codec possui um mecanismo de correção de erro que torna robusto a perda de pacotes (JAMIELSON, 2016). Este trabalho utilizou o SILK16 em seu estudo (banda larga, com 16 kamostras/s).

### 3.2.6. Opus

O Opus é um codec livre e de código aberto. Ele é muito utilizado em transmissões em tempo real da voz e de música através da Internet, mas também se destina ao armazenamento de áudio. Em seu desenvolvimento, foi projetado para codificar eficientemente a voz e o áudio em geral, mantendo a baixa latência e baixa complexidade, de tal modo que possa ser executado por processadores de baixo custo. Além disso, nele foi incorporada a tecnologia do codec SILK do Skype<sup>3</sup> e o codec CELT da Xiph.Org. Foi padronizado pela Força Tarefa de Engenharia da Internet (*Internet Engineering Task Force - IETF*) em 2012 através da RFC<sup>4</sup> 6716. Os principais recursos do Opus são (VOS, 2017):

- Taxa de bits de 6 kbit/s até 510 kbit/s;
- Taxa de amostragem de 8 kamostras/s até 48 kamostras/s;
- Tamanhos de quadro de 2,5 ms a 60 ms;
- Suporte para taxa de bits constante e taxa de bits variável;
- Suporte para mono e estéreo;
- Ajustes automáticos de taxa de bits, largura de banda e tamanho do quadro.

Neste trabalho foi utilizado o codec Opus24, que possui taxa de amostragem de 24 kamostras/s.

---

<sup>3</sup> Software que permite a comunicação de áudio e vídeo através da Internet.

<sup>4</sup> *Request for Comments* (RFC) são documentos técnicos desenvolvidos e mantidos pelo IETF (*Internet Engineering Task Force*), instituição que especifica os padrões que serão implementados e utilizados em toda a internet.

### 3.2.7. Comparação entre os codecs

A Tabela 3 mostra as principais características dos codecs utilizados neste trabalho. Eles foram escolhidos por apresentarem características bastante distintas e também por questão de compatibilidade técnica em relação aos componentes de *software* pertencentes ao cenário de transmissão utilizado neste trabalho (Seção 4.4).

Tabela 3 – Comparativo entre os codecs utilizados.

Codec	Padronização	Descrição	Bits por amostra	Taxa de transmissão (kbit/s)	Taxa de amostragem (kamostras/s)	Atraso (ms)	MOS
G.711 Lei A	ITU-T	<i>Pulse code modulation</i> (PCM)	8	64	8	0,125	4.1
Speex32	Xiph.org	<i>Code-excited linear prediction</i> (CELP)	16	44,2	32	34	3.8
GSM Full Rate	ETSI	<i>RegularPulse Excitation Long-Term Predictor</i> (RPE-LTP)	13	13	8	20	3.6
LPCM16	ITU-T	<i>Linear pulse-code modulation</i> (LPCM)	16	256	16	~0,5	4.1
SILK16	Skype	<i>SILK Codec</i>	16	30	16	25	4.5
Opus24	Xiph.org	<i>Opus codec</i>	16	20	24	26,5	4.7

Fonte: JAMIESON (2017), KARAPANTAZIS e PAVLIDOU (2009), SCHULZRINNE (2008), JOKISCH e MARUSCHKE (2016) e GOUDARZI, SUN e IFEACHOR (2011).

### 3.3. Transmissão da voz com VoIP

A crescente utilização da Internet instigou o surgimento de uma série de novas tecnologias, muitas vezes substituindo algumas já existentes, como é o caso do VoIP, que corresponde a um conjunto de padrões (*framework*) que permite que chamadas telefônicas sejam feitas por meio de uma conexão de banda larga com a Internet, diferente do que acontece no serviço de telefonia convencional, que necessita de um meio físico (cabearamento) específico para a transmissão da voz. Na prática, o VoIP permite que a voz saia de uma origem, seja dividida em pacotes, trafegue por redes de dados (redes locais e Internet) e chegue ao destino onde os pacotes são reunidos e reorganizados, reconstruindo assim a voz para que esta seja reproduzida para o destinatário (KELLER, 2011).

Com a ampliação das velocidades de acesso à Internet, o VoIP passou a fazer parte das grandes corporações, com o objetivo principal de reduzir o valor das contas telefônicas. A principal diferença entre ele e a telefonia tradicional é a forma como a voz é transportada.

O protocolo IP, usado normalmente para enviar e receber e-mails e navegar em websites, no VoIP é usado para transportar a voz (KELLER, 2011).

Alguns benefícios do VoIP quando comparados a telefonia tradicional, são:

- **Redução de custos:** utilização do link de Internet para realização de chamadas externas. Implementação de toda a estrutura de controle através de *software* livre;
- **Infraestrutura única:** a convergência das redes de dados e voz faz com que a rede física se torne única;
- **Mobilidade:** possibilidade de utilizar telefones IP conectados a redes sem fio;
- **Novas funcionalidades:** criptografia, comunicação em grupo e filtragem de chamadas.

### 3.3.1. Qualidade do áudio em VoIP

O bom funcionamento do VoIP, isto é, chamadas com áudio de qualidade, é dependente de alguns elementos, como:

- **Perda de pacotes:** é causada devido a características inerentes à própria rede de dados. Está relacionada a erros no encaminhamento dos pacotes realizado pelos equipamentos da rede e a falhas durante a transmissão. O limite de perda de pacote tolerável em uma chamada VoIP é de 5%, a fim de não prejudicar a qualidade da chamada;
- **Atraso (*delay*):** é a diferença de tempo entre o envio e o recebimento de um sinal. Contempla o tempo de processamento dos pacotes nos equipamentos que compõem a rede. Quanto maior o atraso pior a qualidade da chamada de voz;
- **Jitter:** é a variação do atraso ocasionada pelo excesso de tráfego ou baixa taxa de transmissão. O excesso de *Jitter* gera distorção no áudio.
- **ECO:** é o retorno do áudio enviado. Em geral ele é rápido e imperceptível ao cérebro, porém o aumento do atraso reduz a velocidade do retorno podendo fazer com que passemos a ouvir a nossa própria voz no telefone, ou seja, o eco.
- **Supressão de silêncio (*voice activity detection – VAD*):** é a detecção da ausência de som em uma chamada de voz, utilizada para evitar o envio de pacotes de áudio com a ausência de dados de voz, tornando o uso da rede mais eficiente.
- **Qualidade de serviço (*Quality of service – QoS*):** mecanismos de controle utilizados para priorizar fluxos de dados (como pacotes de voz, por exemplo) dentro da rede, visando minimizar perdas e garantir um melhor desempenho.

### 3.3.2. Os protocolos RTP e RTCP

O RTP (*Real-time transport protocol* – Protocolo de transporte em tempo real), descrito na RFC<sup>5</sup> 1889, é um protocolo amplamente utilizado nas aplicações multimídia. Sua função é multiplexar diversos fluxos de dados em tempo real sobre um único fluxo de pacotes UDP, que por sua vez pode ser enviado a um ou vários destinos. Como o RTP utiliza simplesmente o UDP, seus pacotes não são tratados de maneira especial pelos roteadores<sup>6</sup>, a menos que algum recurso de qualidade de serviço do IP esteja ativo. Em outras palavras, não há nenhuma garantia especial sobre a entrega, e pacotes podem ser perdidos, atrasados ou adulterados (TANENBAUM; WETHERAL, 2011).

O RTP pode ser utilizado para transportar formatos comuns de áudio, como PCM, GSM e MP3, como também de vídeo, como o MPEG e H.263. Também é usado para complementar outros importantes protocolos interativos de tempo real, como o SIP e H.323.

O funcionamento do RTP se inicia do lado remetente que recebe um sinal digital codificado do codificador e o encapsula dentro de um pacote RTP. Em seguida, este pacote é encapsulado em um segmento UDP, passa para o IP e em seguida para o quadro Ethernet que será devidamente transmitido pela rede. O lado receptor faz o processo inverso até extrair o pacote RTP do segmento UDP. Por fim, o sinal de áudio codificado contido no pacote RTP é extraído e transferido para o decodificador (KUROSE; ROSS, 2010).

Outros parâmetros pertencentes ao pacote RTP são o tipo de codificação de áudio utilizado, número de sequência, marca de tempo (*timestamp*) e o identificador de sincronização da fonte, que é responsável por identificar a qual fluxo de dados o pacote pertence.

A Figura 6 demonstra os campos de cabeçalho RTP e a Figura 7 como se dá o processo de encapsulamento do pacote RTP em um quadro Ethernet.

Figura 6 – Campos do cabeçalho do pacote RTP.

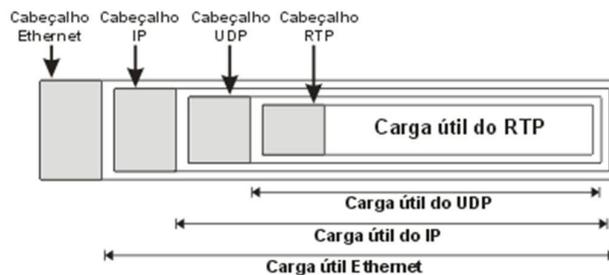


Fonte: baseado em KUROSE; ROSS, 2010.

<sup>5</sup> *Request for Comments* são documentos técnicos desenvolvidos e mantidos pelo IETF (*Internet Engineering Task Force*), instituição que especifica os padrões que serão implementados e utilizados em toda a internet.

<sup>6</sup> É um dispositivo conectado a duas ou mais linhas de dados que encaminha pacotes de dados entre redes de computadores.

Figura 7 – Processo de encapsulamento do pacote RTP para transmissão.



Fonte: baseado em TANENBAUM; WETHERAL, 2011.

O RTCP (*Real-time Transport Control Protocol* – protocolo de controle de transporte em tempo real) é um protocolo definido na RFC 3550 e cuida do feedback, da sincronização e da interface do usuário, mas não transporta nenhuma amostra de mídia (TANENBAUM; WETHERAL, 2011). Eles são enviados periodicamente e contêm relatórios de remetente e/ou receptor com dados estatísticos que podem ser úteis para a aplicação. Esses dados contêm o número de pacotes enviados, número de pacotes perdidos e variação de atraso entre chegadas. Essas informações podem ser utilizadas pelo remetente, por exemplo, para modificar a taxa de transmissão ou alterar o codec a fim de oferecer melhor qualidade nas chamadas de voz.

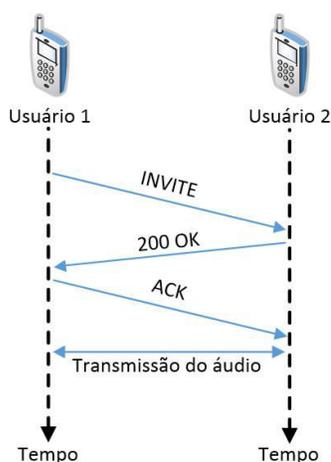
### 3.3.3. O protocolo SIP

O SIP (*Session Initiate Protocol* – protocolo de inicialização de sessão) é um protocolo criado pela IETF e definido nas RFCs 3261 e 5411. Ele provê mecanismos para estabelecer chamadas entre dois interlocutores através de uma rede IP. Permite ao originador da chamada avisar ao destinatário que deseja iniciar uma chamada. Possui ainda o recurso de negociação entre os participantes da chamada quanto a escolha do codec a ser utilizado. Outra característica do SIP é que ele provê mecanismos para gerenciamento de chamadas, tais como adicionar novos fluxos de mídia, mudar a codificação, convidar outros participantes, além de transferir ligações (KUROSE; ROSS, 2010).

O SIP pode estabelecer chamadas entre duas partes (ligações telefônicas comuns), sessões de várias partes (em que todos podem ouvir e falar) e sessões de *multicast* (com um transmissor e muitos receptores). As sessões podem conter áudio, vídeo ou dados. O SIP cuida apenas da configuração do gerenciamento e do encerramento de sessões. Outros protocolos, como o RTP/RTCP, são usados para transporte de dados. O SIP é um protocolo de camada de aplicação e pode funcionar sobre UDP ou TCP, conforme necessidade (TANENBAUM; WETHERAL, 2011).

A Figura 8 mostra um exemplo onde dois usuários desejam estabelecer uma chamada VoIP baseada no protocolo SIP. A sessão se inicia quando o usuário 1 envia uma mensagem do tipo INVITE para o usuário 2 utilizando a porta UDP 5060, valor padrão do SIP. Entre outros parâmetros, a mensagem INVITE contém um identificador do usuário 1, o endereço IP do usuário 2 e a escolha do codec a ser utilizado na transmissão do áudio. Ao receber a mensagem INVITE, o usuário 2 envia uma resposta SIP do tipo 200 OK confirmando sua disponibilidade em estabelecer a chamada utilizando o codec sugerido. Após receber a mensagem 200 OK, o usuário 1 envia uma mensagem SIP de reconhecimento (ACK) para o usuário 2. Após essa transação SIP, os usuários 1 e 2 podem iniciar a transmissão do conteúdo do áudio através do protocolo RTP (FONSECA, 2007).

Figura 8 – Estabelecendo uma chamada SIP.



Fonte: baseado em FONSECA, 2007.

Além das mensagens já mencionadas, o SIP utiliza outros métodos, conforme apresentado na Tabela 4.

Tabela 4 – Métodos utilizados pelo SIP.

Método	Descrição
INVITE	Solicita o início de uma sessão
ACK	Confirma que uma sessão foi iniciada
BYE	Solicita o término de uma sessão
OPTIONS	Consulta um dispositivo sobre seus recursos
CANCEL	Cancela uma solicitação pendente
REGISTER	Informa um servidor de redirecionamento sobre a localização atual do usuário

Fonte: TANENBAUM; WETHERAL, 2011.

### 3.4. Asterisk

O Asterisk é um software livre, de código aberto, desenvolvido para a construção de aplicações de comunicação. Pode prover serviços tradicionais presentes em uma central telefônica e, ao mesmo tempo, ser totalmente compatível com VoIP. Atualmente é utilizado por pequenas e grandes empresas, como *call centers*, agências governamentais e diversas outras ao redor do mundo. Hoje, existem mais de um milhão de sistemas de comunicação baseados em Asterisk em uso, em mais de 170 países. O Asterisk pode se tornar a base de um sistema telefônico comercial completo, ou ser usado para aprimorar ou estender um sistema de telefonia já existente (DIGIUM, 2017).

Algumas das principais funcionalidades que se pode obter com o Asterisk, são:

- Central telefônica tradicional
- Comunicação com VoIP
- Servidor de Voicemail
- Chamadas em conferência
- Call Center
- Atendente virtual

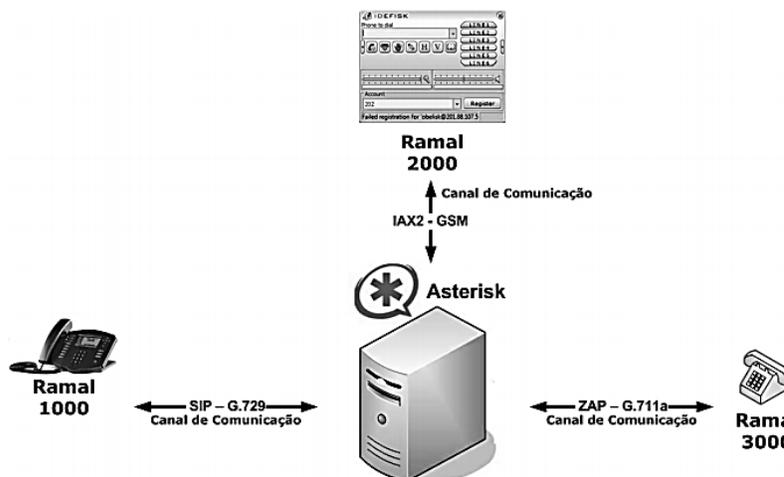
O funcionamento e a operação do Asterisk têm como base quatro componentes (KELLER, 2011):

- **Protocolo:** forma como os clientes se comunicam com o servidor Asterisk, como exemplo existe o SIP, IAX, H.323, RTP, RTCP, entre outros.
- **Canal de comunicação:** todo cliente possui uma identificação para o sistema (ramal) que associado ao protocolo utilizado na comunicação forma-se o canal de comunicação, por exemplo, SIP/4000 (protocolo SIP, ramal 4000).
- **Codec:** modo como o áudio foi digitalizado para o transporte na rede. Como exemplo temos G.711, G.276, GSM, G.729 e Speex, entre outros.
- **Aplicação:** todo o processamento do Asterisk tem como base o seu plano de discagem, ou seja, a organização das regras de discagem. Como exemplo, tem-se *Dial* (inicia uma chamada), *Answer* (atende uma chamada), *Wait* (aguarda um tempo), etc.

Toda chamada processada por um servidor Asterisk segue o mesmo procedimento: um cliente envia uma sequência de caracteres para o servidor, o qual autentica o cliente e então busca por uma regra equivalente aos caracteres recebidos dentro dos grupos de regras associados a esse cliente; somente assim é executada a aplicação especificada na

regra e a chamada é completada. A Figura 9 exemplifica uma comunicação entre três clientes gerenciada pelo Asterisk, onde cada cliente utiliza um protocolo distinto.

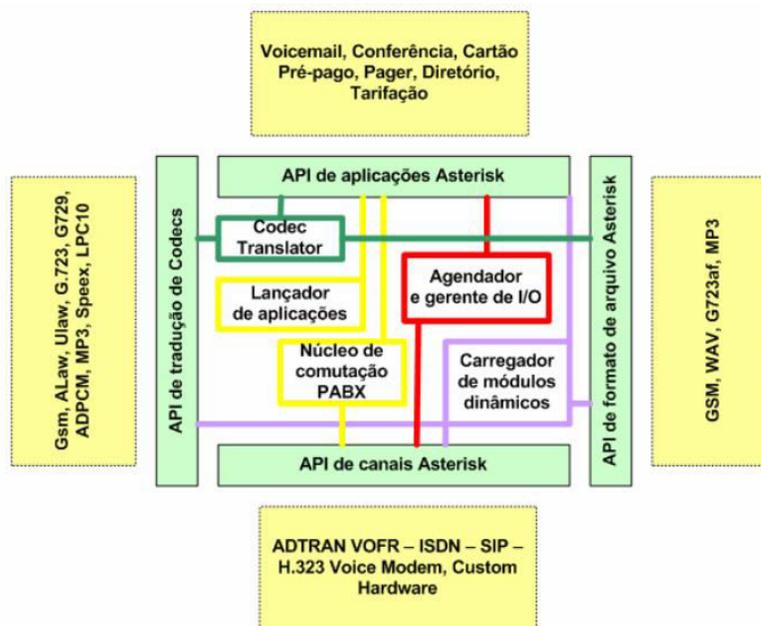
Figura 9 – Comunicação gerenciada pelo Asterisk utilizando protocolos distintos.



Fonte: KELLER, 2011.

A Figura 10 mostra a arquitetura de funcionamento do Asterisk. Observa-se a existência de quatro APIs principais, entre elas a de aplicações, a de canais, a de tradução de codecs e a de formato de arquivos.

Figura 10 – Arquitetura do Asterisk.



Fonte: GONÇALVES, 2005.

A API de aplicações do Asterisk define várias ações que podem ser aplicadas a chamadas. Ela é responsável por realizar a comunicação com dispositivos externos. A API de canais tem a função de redirecionar as chamadas entrantes para um canal específico, de acordo com o protocolo utilizado, antes que elas cheguem ao núcleo de comutação. Tem também o papel de encaminhar as chamadas de saída para um driver de canal antes que elas possam ser direcionadas a um dispositivo externo. A API de tradução de codecs proporciona que o Asterisk possa fazer a conversão de fluxos de áudio codificados com diferentes formatos de codecs, permitindo que chamadas iniciadas com um dispositivo que opera com o codec GSM possam ser convertidas, por exemplo, em outro fluxo de chamada para alcançar um receptor que reconhece apenas o formato de codec G.711. A API de formato de arquivo tem seu funcionamento semelhante à API de tradução de codecs, mas opera com formato de arquivos. Caso exista uma gravação armazenada em formato GSM, a API de formato de arquivo precisa ser utilizada para tocar esta gravação em canais que não utilizam codecs GSM, por exemplo (BRYANT; MADSEN; MEGGELEN, 2013).

No capítulo seguinte são apresentadas as análises realizadas presencialmente pelo fonoaudiólogo durante o processo de detecção de desvios vocais, bem como a descrição das medidas mais utilizadas na avaliação da qualidade vocal.

## 4. AVALIAÇÃO DE DESVIOS VOCAIS

O processo de avaliação de desvios vocais precisa ser realizado sob diferentes aspectos para a obtenção de um diagnóstico mais preciso sobre os prováveis fatores causadores desses distúrbios. Entre os métodos mais empregados na avaliação de desvios vocais estão a análise perceptivo-auditiva e a análise acústica da voz (BEHLAU *et al*, 2013; COLTON; CASPER, LEONARD, 2011). Estes dois métodos são descritos neste capítulo.

### 4.1. Análise perceptivo-auditiva da voz

Desvios vocais podem ser provocados por abuso vocal ou por lesões laringeas, que podem ser origem orgânica, a exemplo das lesões de massa nas pregas vocais como nódulos, cistos, edemas e pólipos ou por patologias de origem neurológica como a paralisia nas pregas vocais, mal de Parkinson, entre outras. A qualidade vocal pode ser restabelecida, em alguns casos, por meio de terapias fonoaudiológicas, podendo ser empregadas, também, após processos cirúrgicos e/ou tratamentos medicamentosos.

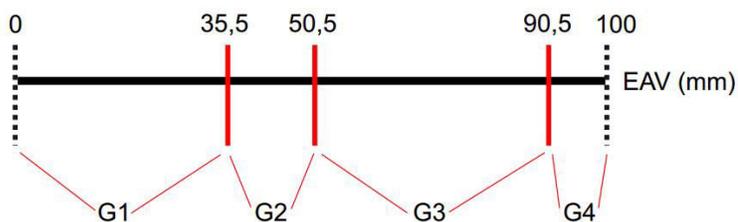
Entre os desvios vocais que podem ser observados em uma análise perceptivo-auditiva estão os relacionados à qualidade vocal, como a soproidade e a rugosidade. As vozes soprosas são caracterizadas pela emissão de ar, acima do normal, durante a produção da voz. Neste tipo de desvio os pacientes geralmente se queixam de que não são capazes de pronunciar sentenças completas sem que fiquem sem ar. Já as vozes rugosas são caracterizadas pela presença de rouquidão ou aspereza na voz. A rugosidade reflete a vibração aperiódica das pregas vocais (COLTON; CASPER, LEONARD, 2011).

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 (COLTON; CASPER, LEONARD, 2011).

Um procedimento bastante comum na análise perceptivo-auditiva da voz é a quantificação da avaliação por meio de escalas, com as quais é possível classificar o grau de intensidade dos distúrbios. Uma das principais escalas utilizadas nestas análises é a Escala Analógico-Visual (EAV). Essa escala se constitui em 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) (YAMASAKI, 2008). A escala EAV está ilustrada na Figura 11.

Figura 11 – Escala analógico-visual (EAV).



Fonte: VIEIRA, 2014.

A avaliação da qualidade vocal por meio de escalas apresenta caráter subjetivo, o que tem resultado dependente da experiência do(s) avaliador(es). Técnicas objetivas de análise podem ser acrescentadas à análise subjetiva, aumentando a confiabilidade do diagnóstico e a eficácia do tratamento. A análise acústica tem sido empregada, nos últimos anos, proporcionando caráter objetivo ao diagnóstico.

## 4.2. Análise acústica da voz

A análise acústica se refere à aplicação de métodos computacionais não invasivos (*softwares*), que realizam a extração de características do sinal de voz inerentes ao sistema de produção vocal, com o objetivo de quantificar as alterações vocais (COLTON; CASPER, LEONARD, 2011). A utilização de técnicas de análise acústica não tem o objetivo de eliminar o uso dos exames perceptivos ou de outros métodos, como também não questiona a eficiência dos mesmos. Esta análise busca auxiliar os profissionais da voz como ferramenta adicional para a obtenção de um diagnóstico mais preciso, além de poder ser utilizada para terapia vocal em pessoas com problemas na fala (COSTA, 2008).

Diversas características são extraídas do sinal de voz para análise e classificação dos sinais, de forma que representem aquilo que se deseja observar, a exemplo dos desvios vocais, sejam eles provocados ou não por lesões laríngeas. As características representativas do sinal de voz podem ser obtidas tanto no domínio do tempo (energia, autocorrelação, *shimmer*, frequência fundamental e seus derivados como *jitter*, *shimmer*, entre outros) quanto no domínio da frequência (Espectrograma<sup>7</sup>, densidade espectral de potência, frequências formantes, entre outras) (COLTON; CASPER, LEONARD, 2011).

<sup>7</sup> Gráficos que representam a densidade espectral de energia do sinal.

As medidas citadas são tradicionalmente empregadas e obtidas a partir do modelo linear de produção da fala. Outras medidas obtidas a partir da análise dinâmica não linear também vêm sendo empregadas a exemplo da entropia, dimensão de correlação, medidas de quantificação de recorrência, entre outras (ZHANG *et al*, 2004).

Algumas das medidas extraídas dos sinais de voz que são tradicionalmente utilizadas no processo de análise acústica por profissionais que realizam a terapia fonoaudiológica são a frequência fundamental, o *jitter*, o *shimmer* e a excitação do ruído glotal (*Glottal Noise Excitation* - GNE), descritas a seguir.

#### 4.2.1. Frequência fundamental ( $F_0$ )

A frequência fundamental, também conhecida como  $F_0$ , é determinada como sendo a menor frequência de uma onda periódica. No estudo da voz ela é o resultado de características biodinâmicas das pregas vocais, tais como a sua vibração e as variações de intensidade que ocorrem entre sons agudos e graves (FRAZÃO; COSTA, CORREIA, 2013).

Os fatores que determinam a frequência fundamental são o comprimento da prega vocal, o alongamento, a massa em vibração e a tensão envolvida. O valor médio referente à frequência fundamental varia entre 80 e 400Hz. Eles são afetados pelo sexo e pela idade. Nos adultos jovens a frequência fundamental varia de 80 a 250 Hz, sendo que nos homens a faixa é de 80 a 150 Hz, nas mulheres de 150 a 250 Hz e nas crianças está entre 250 a 400 Hz (BEHLAU *et al*, 2013).

A frequência fundamental pode diferenciar vozes rugosas e soprosas (LOPES; CAVALCANTE; COSTA, 2014).

#### 4.2.2. *Jitter*

Como explicado na seção 3.3.1, em um processo de transmissão da voz, o *jitter* refere-se à variação do atraso ocorrido entre a origem e o destino, porém no contexto da análise acústica o *jitter* é definido como sendo o índice de perturbação da frequência fundamental em curto intervalo de tempo. Expressa o quanto o valor da frequência fundamental em um período é diferente do anterior ou de seu sucessor imediato. Altera-se principalmente com a falta de controle da vibração das pregas vocais. Pode ser expresso através de medidas absolutas ou medidas relativas. As medidas absolutas do *jitter* são expressas em segundos, milissegundos ou microssegundos. Já as medidas relativas consideram a frequência fundamental média e são expressas em porcentagem (BEHLAU *et al*, 2013). O *software* utilizado nesta pesquisa para a extração das medidas acústicas

(VoxMetria) utiliza medidas relativas (porcentagem). Esta medida apresenta correlação positiva com a soproiedade (LOPES; CAVALCANTE; COSTA, 2014).

#### **4.2.3. Shimmer**

O *shimmer* é definido como sendo o índice de perturbação da amplitude em curto intervalo de tempo. Indica a variação irregular da amplitude da onda sonora, de forma consecutiva, nos intervalos de observação do sinal. Altera-se principalmente nas situações de redução de resistência glótica e correlaciona-se com a presença de ruído à emissão (rouquidão) e com a soproiedade. Suas medidas são expressas em porcentagem. O cálculo do *shimmer* é baseado na diferença entre os picos mais altos de cada impulso, em que é calculada a diferença entre as amplitudes sucessivas do sinal. Altos níveis de *shimmer* normalmente estão associados às vozes patológicas (TEIXEIRA; FERREIRA; CARNEIRO, 2011). A medida de shimmer pode também ser empregada na detecção de rugosidade (LOPES; CAVALCANTE, COSTA, 2014).

#### **4.2.4. Glottal Noise Excitation (GNE)**

O *Glottal Noise Excitation* (GNE) é a medida acústica que calcula o ruído produzido pela oscilação das pregas vocais. Está intimamente relacionado com a soproiedade. É aplicável mesmo para oscilações glotais altamente irregulares. Se o valor de GNE for próximo a um significa excitação normal da glote, enquanto que os valores próximos a zero resultam de uma excitação do tipo ruído. O GNE não necessita do cálculo da frequência fundamental e gera informações mais independentes de *jitter* e *shimmer*. Seus valores são considerados normais quando maiores ou iguais a 0,5 dB (CARRASCO; OLIVEIRA; BEHLAU, 2010). O GNE pode ser empregado para diferenciar vozes saudáveis de vozes desviadas e classificar o grau de soproiedade (LOPES; CAVALCANTE; COSTA, 2014).

A seguir, é descrita a metodologia empregada na pesquisa incluindo o cenário empregando para transmissão, a forma de obtenção e análise dos sinais.

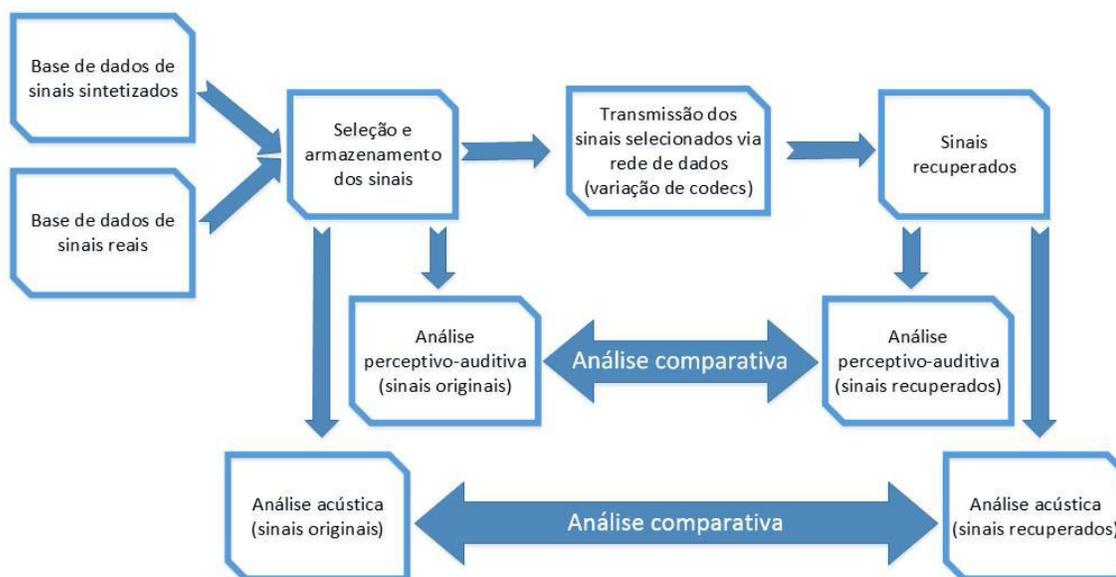
## 5. MATERIAIS E MÉTODOS

Neste capítulo, é descrita a metodologia empregada neste trabalho incluindo a descrição das bases de dados utilizadas, a maneira como as análises perceptivo-auditiva e acústica foram realizadas e o cenário de transmissão.

### 5.1. Metodologia

Os sinais de voz utilizados na pesquisa foram selecionados de duas bases de dados, sendo uma de sinais sintetizados e outra de sinais reais, ambas armazenadas em um servidor Asterisk. Para cada sinal, foi realizada uma avaliação perceptivo-auditiva e foram extraídas as características acústicas inerentes aos sinais de voz. Cada um dos sinais selecionados foi transmitido seis 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 novas medidas acústicas foram extraídas. Os dados obtidos foram comparados com aqueles extraídos antes da transmissão. Os passos descritos estão apresentados no diagrama em blocos da Figura 12.

Figura 12 – Modelo geral da metodologia empregada.



Fonte: Autoria própria.

## 5.2. Análise perceptivo-auditiva

Para a avaliação perceptivo-auditiva contou-se com o apoio de um avaliador externo, fonoaudiólogo, com especialização em voz, mestrado em Ciências da Linguagem e doutor em Linguística. Possui larga experiência em identificação e avaliação de desvios vocais, atuando hoje como líder do Laboratório Integrado de Estudos da Voz (LIEV) da Universidade Federal da Paraíba (UFPB). O Coeficiente Kappa de Cohen<sup>8</sup> do avaliador é 0,79, indicando uma boa confiabilidade (LOPES *et al*, 2015).

Após a transmissão de cada sinal, foi solicitado ao avaliador que realizasse uma análise perceptivo-auditiva, de modo a classificá-lo em normal, rugoso ou soproso, além de inferir o grau de intensidade de intensidade dos desvios vocais soproso e rugosidade, 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 foi realizada em duas sessões distintas, sendo uma para a os sinais sintetizados e outra para os sinais reais. Em cada sessão, ocorreu a avaliação tanto dos sinais originais como dos sinais recuperados. Cada sessão teve duração de 60 minutos e foi realizada em ambiente silencioso. Antes de cada sessão, o avaliador foi treinado com sinais estímulos-âncora, contendo emissões normais e com desvio nos diferentes graus, assim como vozes predominantemente rugosas e soproso. Além disso, instruiu-se o fonoaudiólogo quanto aos valores de corte (BARAVIEIRA *et al*, 2016), que seriam adotados nesta pesquisa, para categorização das vozes quanto à ausência e presença de rugosidade e soproso. Para a avaliação, cada sinal foi apresentado por três vezes através de fone de ouvido simples, em intensidade confortável autorreferida pelo avaliador.

Os resultados da avaliação perceptivo-auditiva, tanto dos sinais sintetizados como dos reais, foram dispostos em uma matriz de confusão e posteriormente organizados em uma tabela de acordo com as taxas de Acurácia, Sensibilidade e Especificidade obtidas pelo avaliador após as transmissões, onde:

- **Acurácia:** refere-se aos acertos na classificação dos sinais (normal, soproso ou rugoso) após as transmissões, obtida por:

$$\text{Acurácia} = \frac{\text{Quantidade de sinais classificados corretamente}}{\text{Número total de sinais}} \quad (7)$$

---

<sup>8</sup> Medida estatística que indica a confiabilidade na classificação de um determinado objeto. Varia de -1 a 1, onde -1 indica uma concordância pobre e 1 uma concordância perfeita.

- **Especificidade:** trata da detecção correta da ausência dos desvios em sinais normais, dado por:

$$\text{Especificidade} = \frac{\text{Quantidade de sinais normais classificados como normais}}{\text{Número total de sinais normais}} \quad (8)$$

- **Sensibilidade:** é a capacidade de detectar a presença do desvio vocal em sinais desviados, obtida por:

$$\text{Sensibilidade} = \frac{\text{Quantida de sinais desviados classificados como desviados}}{\text{Número total de sinais desviados}} \quad (9)$$

### 5.3. Análise acústica

Para a realização da análise acústica dos sinais, foi realizada a extração de características acústicas através do *software* VoxMetria, versão 4, da CTS Informática<sup>9</sup>. A extração aconteceu antes das transmissões em todos os sinais originais selecionados e depois das transmissões em todos os sinais recuperados. Foram extraídos valores dos parâmetros frequência fundamental ( $F_0$ ), desvio padrão da frequência fundamental, *jitter*, *shimmer* e *Glottal Noise Excitation* (GNE).

### 5.4. Bases de dados utilizadas e seleção dos sinais

#### 5.4.1. Base de dados de sinais sintetizados

A base de dados de sinais sintetizados utilizada foi desenvolvida pelo 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 (BEHLAU *et al*, 2013). O uso de vozes sintetizadas pode proporcionar maior controle sobre os desvios vocais presentes no sinal de voz.

Os desvios vocais foram produzidos utilizando três parâmetros de manipulação. Para a rugosidade, o comprimento do ciclo de excitação glótica, foi gerado pela introdução de uma perturbação na rigidez do tecido da prega vocal, de acordo com a equação:

<sup>9</sup> Maiores informações em: <http://www.ctsinformatica.com.br/voxmetria-analise-de-voz-e-qualidade-vocal>

$$\Delta K = a\varepsilon K \quad (5)$$

Onde “a” é um parâmetro de escala, “ $\varepsilon$ ” é uma variável aleatória, e “K” é um coeficiente de rigidez da prega vocal. Para a respiração, adicionou-se ruído, sob a fórmula:

$$\Delta u = b\varepsilon u \quad (6)$$

Onde “u” é a taxa de fluxo glotal, “b” é um parâmetro de escala, e “ $\varepsilon$ ” é uma variável aleatória semelhante ao *jitter*. Para as vozes tensas, foi realizado o aumento de tensão, “K”, da pressão subglotal e a diminuição da pressão vocal (BEHLAU *et al*, 2013).

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. Além disso, todos os sinais desta base já haviam sido avaliados previamente, usando análise perceptivo-auditiva, escala EAV. A medida de qualidade dos sinais sintetizados produzidos pelo VoiceSim pode ser verificada em Englert *et al* (2016).

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 seis vezes, uma vez para cada tipo de codec (G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16). 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 sopro 1 e rugosidade 1. Para os sinais soprosos foram selecionados sinais com grau de sopro 2, 3 ou 4 e rugosidade 1. Para os sinais rugosos foram selecionados sinais com grau de rugosidade 2, 3 ou 4 e sopro 1.

#### **5.4.2. Base de dados de sinais reais**

A base de dados de sinais reais utilizada foi desenvolvida e disponibilizada pelo Laboratório Integrado de Estudos da Voz (LIEV) da Universidade Federal da Paraíba, Campus 1, situada em João Pessoa, Paraíba. A base faz parte de um projeto intitulado “Integração das medidas acústicas, perceptivas, fisiológicas e de autoavaliação em pacientes com e sem alteração laringea”. Os pacientes do estudo possuíam faixa etária entre 18 e 65 anos, sendo tanto do sexo feminino como do masculino, e já apresentavam queixa vocal. Esta base, até a data da pesquisa, possuía cerca de 800 amostras. Além disso, todos os sinais desta base já haviam sido avaliados previamente, empregando análise perceptivo-auditiva, escala EAV.

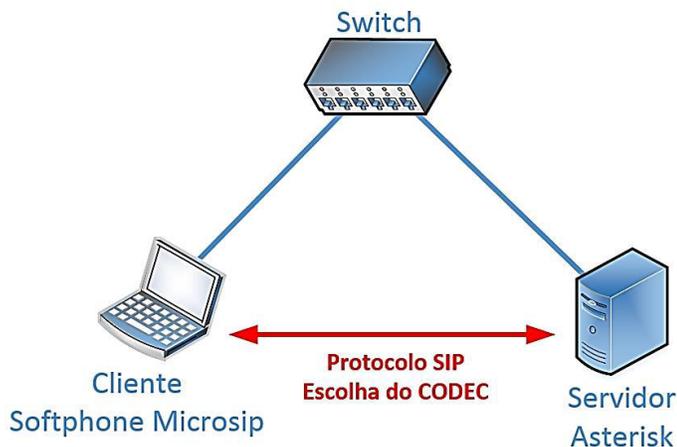
Foram extraídos dos pacientes sinais de vozes para vogal sustentada /E/ (“é”). A coleta dos dados foi realizada em um ambiente tratado acusticamente (cabine de gravação), com ruído ambiental inferior a 50 dB. A taxa de amostragem foi de 44100 amostras/s e 16 bits por amostra, preservando-se, assim, a maior parte das informações do sinal vocal.

Para este estudo foram selecionados da base de dados 36 sinais, dos quais 7 são de vozes saudáveis (normais), 11 de vozes rugosas e 18 sinais de vozes soprosas. Cada um dos 36 sinais selecionados foi transmitido seis vezes, uma vez para cada tipo de codec (G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16). 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.

### 5.5. 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 ao auxílio de dois computadores, conectados através de um equipamento *switch*, conforme ilustrado na Figura 13.

Figura 13 – Cenário utilizado no estudo.



Fonte: Autoria própria.

Além do *softphone* Microsip<sup>10</sup>, no cliente também foi utilizada a ferramenta de gravação *SoundTap Streaming Audio Recorder*, da NCH Software<sup>11</sup>. Esta ferramenta permite a gravação de chamadas VoIP através de um driver especial que preserva a qualidade de áudio digital. A gravação utilizou uma taxa de amostragem de 44100 amostras/s e 16 bits/amostra.

A Tabela 5 apresenta as características de cada elemento utilizado no cenário.

Tabela 5 – Características dos elementos empregados no cenário.

Equipamento	Hardware	Sistema operacional	Software(s)
<b>Servidor</b>	Qbex Core i3 4GB	Debian 8.7.1 x64	Asterisk 14.2.1
<b>Cliente</b>	DELL Vostro 5470 Core i7 8GB	Windows 7 Professional x64	Microsip 3.14.4 SoundTap 3.0.0
<b>Switch</b>	Linksys WRT54G	Firmware v8.00.8	N/A

Fonte: próprio autor.

O cenário foi criado de maneira controlada, com os elementos conectados via cabos a uma velocidade de 100Mbps, através da tecnologia Ethernet. Parâmetros como atraso de pacotes e variação de atraso de pacotes foram desconsiderados neste estudo, por ser considerados desprezíveis para a aplicação em foco.

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

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

Cada sinal foi transmitido seis vezes, uma vez para cada codec escolhido (Speex32, G.711 Lei A, GSM *Full Rate*, LPCM16, Opus24 e SILK16). A seleção do codec foi realizada tanto no servidor Asterisk como também no cliente Microsip. Durante cada transmissão, apenas um codec ficava habilitado, garantido que apenas o codec escolhido ficasse em operação. Após todas as transmissões, um conjunto de 432 sinais foi gerado, sendo 216 deles provenientes dos sinais sintetizados e 216 dos sinais reais.

<sup>10</sup> Maiores informações em [www.microsip.org](http://www.microsip.org)

<sup>11</sup> Maiores informações em [www.nch.com.au/soundtap/index.html](http://www.nch.com.au/soundtap/index.html)

## 6. RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados e discutidos os resultados referentes à análise dos sinais após a transmissão pelos codecs. Ele está dividido em três seções, a primeira abordando os resultados da análise perceptivo-auditiva, a segunda discutindo os resultados da análise acústica e a terceira sintetizando os principais efeitos obtidos por cada um dos codecs estudados.

### 6.1. Resultados da análise perceptivo-auditiva

#### 6.1.1. Sinais sintetizados

A Tabela 6 apresenta os resultados das classificações (matriz de confusão) dos 216 sinais sintetizados analisados pelo avaliador, após as transmissões, em relação às características de normalidade, rugosidade e sopro. Os 216 sinais foram decorrentes da transmissão dos 36 sinais originais selecionados (11 normais, 14 soprosos e 11 rugosos) com a utilização dos seis codecs avaliados.

Tabela 6 - Classificação dos sinais sintetizados pós-transmissão - matriz de confusão

Classificação inicial	G.711 Lei A			Speex32			GSM Full Rate			LPCM16			Opus24			SILK16		
	Nor	Rug	Sop	Nor	Rug	Sop	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	10	1	0	7	3	1
Rugoso (Rug)	2	9	0	0	10	1	0	10	1	1	9	1	2	9	0	0	9	2
Soproso (Sop)	1	9	4	0	5	9	0	13	1	0	1	13	0	6	8	0	2	12

Os codecs G.711 Lei A e o codec GSM *Full Rate* confundiram o desvio sopro com rugosidade. Vale destacar que o desvio rugosidade se encontra numa faixa de frequência até 3 kHz, enquanto que o desvio sopro nas faixas de frequências acima de 3 kHz (YANAGIHARA, 1967). Dessa forma, o G.711 Lei A e o GSM *Full Rate*, que limitam a faixa dos sinais a serem transmitidos em 4 kHz, suprimem o desvio sopro, não se mostrando, portanto, adequados para detectar este desvio (Tabela 6).

Os codecs que melhor preservaram o desvio sopro foram o Speex32, o LPCM16, o SILK16 e o Opus24 os quais tem taxas de amostragem de 32, 16 e 24 kamostras/s, respectivamente. Os demais fornecem taxas de 8 kamostras/s, suprimindo esse desvio na maior parte dos sinais sintetizados.

O Speex32 parece ter introduzido ruído nos sinais, tendo transformado os sinais normais em sinais com desvios vocais, tanto rugosidade quanto sopro. Este codec, no entanto, teve maior facilidade para detectar os desvios vocais (Tabela 6).

Tabela 7 – Acurácia, especificidade e sensibilidade dos codecs analisados (sinais sintetizados).

Codec	Especificidade	Sensibilidade	Acurácia
G.711 Lei A	100%	52,00%	66,67%
Speex32	0,00%	76,00%	52,78%
GSM <i>Full Rate</i>	9,09%	44,00%	33,33%
LPCM16	18,18%	88,00%	66,67%
Opus24	90,91%	68,00%	75,00%
SILK16	63,64%	84,00%	77,78%

Os codecs SILK16 e Opus24 proporcionaram os melhores resultados de Acurácia. Já o pior resultado ocorreu com o GSM *Full Rate* (Tabela 7).

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 3) também apresentaram a melhor Acurácia (Tabela 7). Verifica-se, ainda, que a resolução do codec parece não ter influência direta na Acurácia, uma vez que o LPCM16, mesmo possuindo o mesmo número de bits por amostra em relação aos codecs SILK16 e Opus24, obteve resultado bem inferior. Além disso, o LPCM16, mesmo possuindo o dobro de bits por amostra em relação ao codec G.711 Lei A, obteve a mesma taxa de Acurácia.

Para a Sensibilidade, o codec LPCM16 foi aquele que apresentou melhor resultado, oferecendo 88% de acerto, seguido pelo SILK16 e pelo Speex32, que obtiveram taxas acima de 70%. Esses três codecs foram os que menos prejudicaram a detecção dos desvios vocais nos sinais originais sintetizados, por parte do avaliador.

Em relação à Sensibilidade, observa-se também que o número de bits por amostra pode ter causado influência, visto que os codecs que possuem o maior número de bits por amostra, apresentaram as melhores taxas de Sensibilidade.

Para a Especificidade, o codec G.711 Lei A foi o único que conseguiu manter a integridade dos sinais sintetizados normais, sem introduzir desvios significativos, obtendo resultado de 100%. Outro codec que obteve taxa acima de 90% foi o Opus24, enquanto os demais apresentaram resultados bem inferiores.

### 6.1.2. Sinais reais

A Tabela 8 apresenta os resultados das classificações (matriz de confusão) dos 216 sinais reais analisados pelo avaliador, após as transmissões, em relação às características

de normalidade, rugosidade e sopro. Salienta-se que os 216 sinais foram decorrentes da transmissão dos 36 sinais originais seleccionados (7 normais, 18 soprosos e 11 rugosos) com a utilização dos seis codecs avaliados.

Tabela 8 - Classificação dos sinais reais pós-transmissão - matriz de confusão

Classificação inicial	G.711 Lei A			Speex32			GSM Full Rate			LPCM16			Opus24			SILK16		
	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop	Nor	Rug	Sop
Normal (Nor)	4	2	1	1	5	1	3	3	1	5	1	1	6	1	0	4	3	0
Rugoso (Rug)	2	8	1	0	10	1	1	7	3	1	10	0	2	9	0	0	11	0
Soproso (Sop)	6	3	9	0	5	13	3	7	8	5	3	10	3	4	11	1	8	9

Novamente verifica-se que os codecs com maior taxa de amostragem preservaram melhor os sinais desviados, enquanto que o G.711 Lei A e o GSM Full Rate que operam a 8 kamostras/s apresentaram maior dificuldade.

O Speex32 parece ter introduzido ruído nos sinais, tendo transformado boa parte dos sinais normais em sinais com desvios vocais, tanto rugosidade quanto sopro (Tabela 8).

Na Tabela 9 estão apresentadas as taxas de Acurácia, Sensibilidade e Especificidade obtidas na avaliação dos sinais reais após os codecs.

Tabela 9 – Acurácia, especificidade e sensibilidade dos codecs analisados (sinais reais)

Codec	Especificidade	Sensibilidade	Acurácia
G.711 Lei A	57,14%	58,62%	58,33%
Speex32	14,29%	79,31%	66,67%
GSM Full Rate	42,86%	51,72%	50,00%
LPCM16	71,43%	68,97%	69,44%
Opus24	85,71%	68,97%	72,22%
SILK16	57,14%	68,97%	66,67%

Os codecs LPCM16 e Opus24 foram os que tiveram os melhores resultados de Acurácia. Já o pior resultado ocorreu com o GSM Full Rate (Tabela 9).

Para a Sensibilidade, o codec Speex32 obteve o melhor resultado, seguido pelos codecs LPCM16, Opus24 e SILK16, enquanto os demais apresentaram resultados bem inferiores.

Observa-se que a resolução dos codecs analisados pode ter influenciado na Sensibilidade e na Acurácia, visto que os codecs que possuem a maior quantidade de bits

por amostra (Tabela 3) também apresentaram os melhores resultados para essas duas características (Tabela 9).

Para a Especificidade, o codec Opus24 foi aquele que apresentou melhor resultado, indicando possuir boa capacidade para perceber a ausência de desvios vocais em sinais normais. Ele foi seguido pelo LPCM16, que obteve índice acima de 70%.

A taxa de amostragem parece não ter influência na Especificidade, uma vez que, dentre os codecs estudados, o Speex32 possui a maior taxa de amostragem (Tabela 3) e mesmo assim obteve o pior índice de Especificidade, enquanto o G.711 Lei A, mesmo possuindo a metade da taxa de amostragem em relação ao SILK16 (Tabela 3), obteve o mesmo índice de Especificidade (Tabela 9).

## 6.2. Resultados da análise acústica

### 6.2.1. Sinais sintetizados

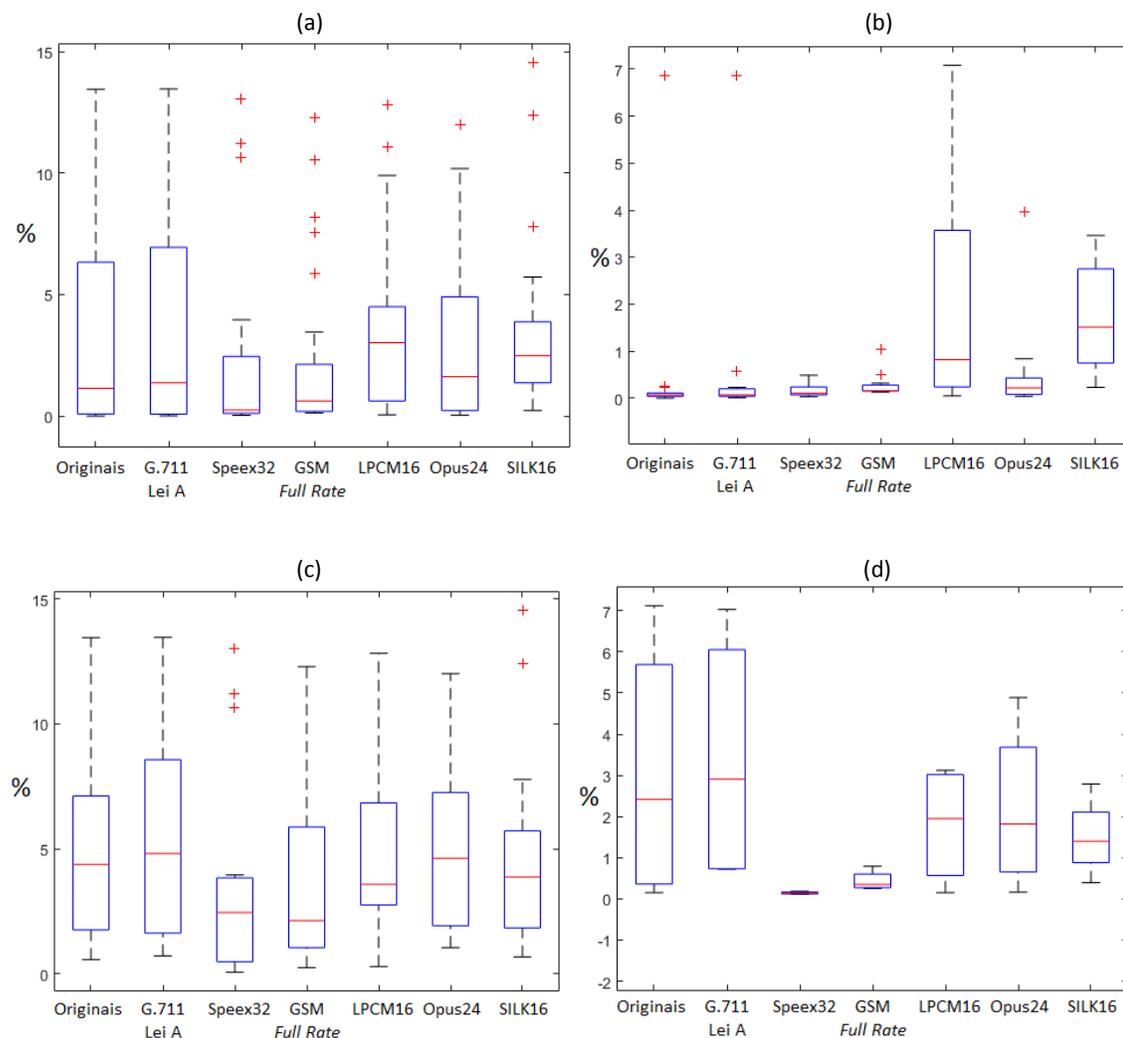
As Figuras 14, 15, 16, 17 e 18 apresentam, respectivamente, gráficos em formato *boxplot* das medidas acústicas *jitter*, *shimmer*, GNE, frequência fundamental e desvio padrão da frequência fundamental, extraídas dos 36 sinais originais sintetizados e dos seus respectivos sinais recuperados, após as transmissões, utilizando os seis codecs estudados. Esses gráficos foram gerados a partir das tabelas presentes no APÊNDICE A.

Através dos *boxplots* é possível identificar 5 estatísticas:

- Mínimo: linha preta contínua na horizontal, localizada na parte inferior, ao final da linha preta vertical pontilhada;
- Primeiro quartil: limite inferior horizontal do retângulo azul
- Segundo quartil ou mediana: linha vermelha na horizontal
- Terceiro quartil: limite superior horizontal do retângulo azul
- Máximo: linha preta contínua na horizontal, localizada na parte superior, ao final da linha preta vertical pontilhada.

O retângulo azul localizado entre o primeiro quartil e o terceiro quartil contém 50% dos valores do conjunto de dados. A posição da linha mediana no retângulo informa sobre a assimetria da distribuição. Pontos fora destes limites são considerados valores atípicos (*outliers*) e são denotados por uma cruz vermelha.

Figura 14 – Valores de *jitter* (em porcentagem) dos sinais sintetizados, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.



De forma geral, percebe-se que, para todos os tipos de sinais sintetizados, os codecs G.711 Lei A e Opus24 tiveram maior facilidade em manter o parâmetro de *jitter*. Destaca-se os codecs Speex32 e GSM *Full Rate*, que apresentaram maior distorção de valores para o *jitter* (Figura 14a).

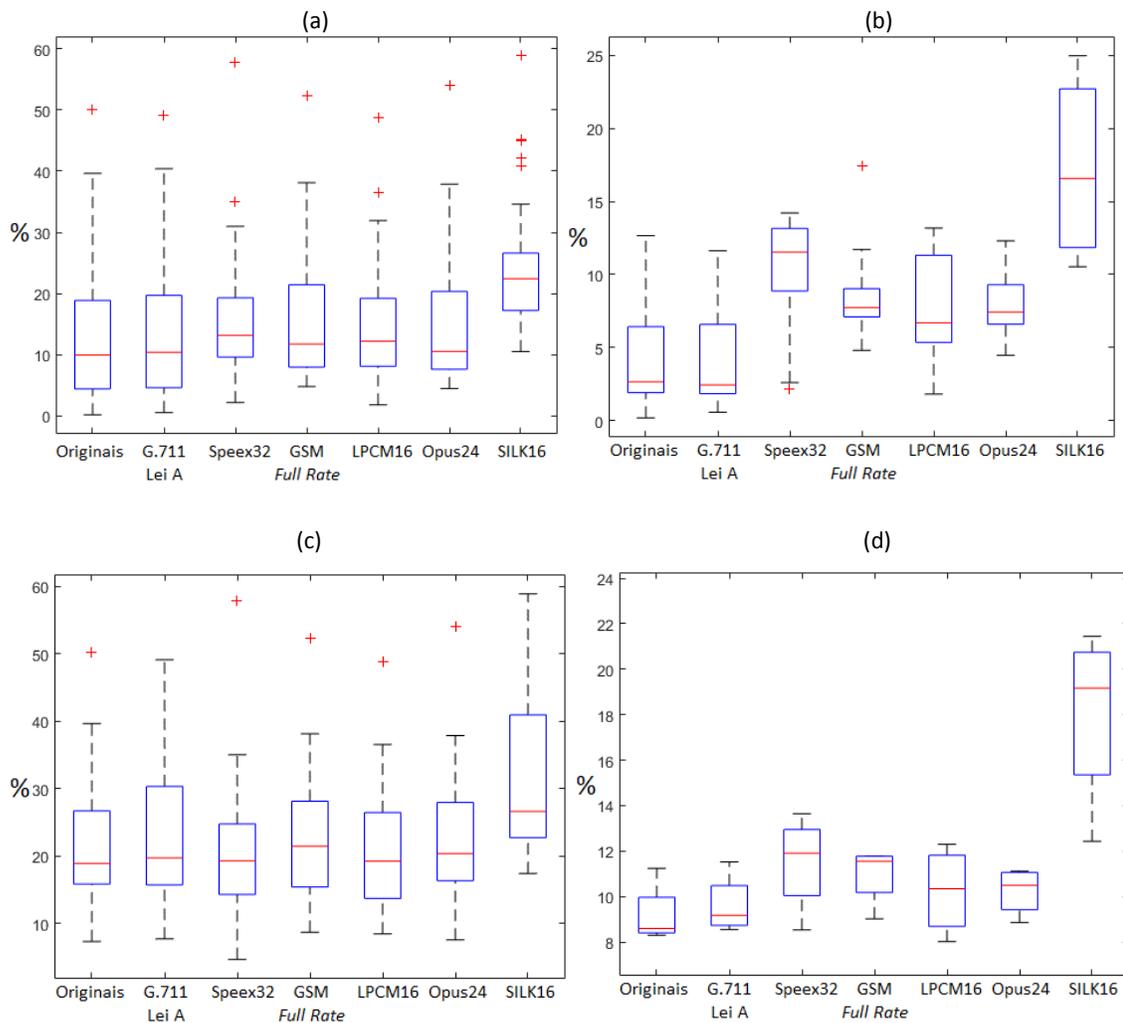
O parâmetro *jitter* pode ter relação com a Especificidade, uma vez que o G.711 Lei A e o Opus24 obtiveram os melhores resultados para o *jitter* (Figura 14a) e também para a Especificidade (Tabela 7), enquanto os codecs Speex32 e GSM *Full Rate* obtiveram os piores resultados para o *jitter* e também para a Especificidade. Este resultado pode estar relacionado ao descrito por Lopes, Cavalcante e Costa (2014), que afirma que o *jitter* possui correlação positiva com os desvios de rugosidade e sopro.

Em relação apenas aos sinais normais sintetizados, todos os codecs tiveram certa facilidade em manter o parâmetro *jitter*, com exceção do LPCM16 e do SILK16 que apresentaram maior dificuldade (Figura 14b).

Considerando apenas os sinais rugosos sintetizados, o codec Opus24 foi aquele que apresentou melhor resultado para o *jitter*, enquanto o pior resultado foi apresentado pelo Speex32 (Figura 14c).

Quando considerados apenas os sinais soprosos sintetizados, verifica-se que o codec G.711 Lei A foi o que apresentou maior facilidade em manter o *jitter*. Percebe-se ainda que para este tipo de sinal os codecs Speex32 e GSM *Full Rate* apresentaram os piores resultados, destacando-se o Speex32 que reduziu o parâmetro *jitter* (Figura 14d).

Figura 15 – Valores de *shimmer* (em porcentagem) dos sinais sintetizados, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.



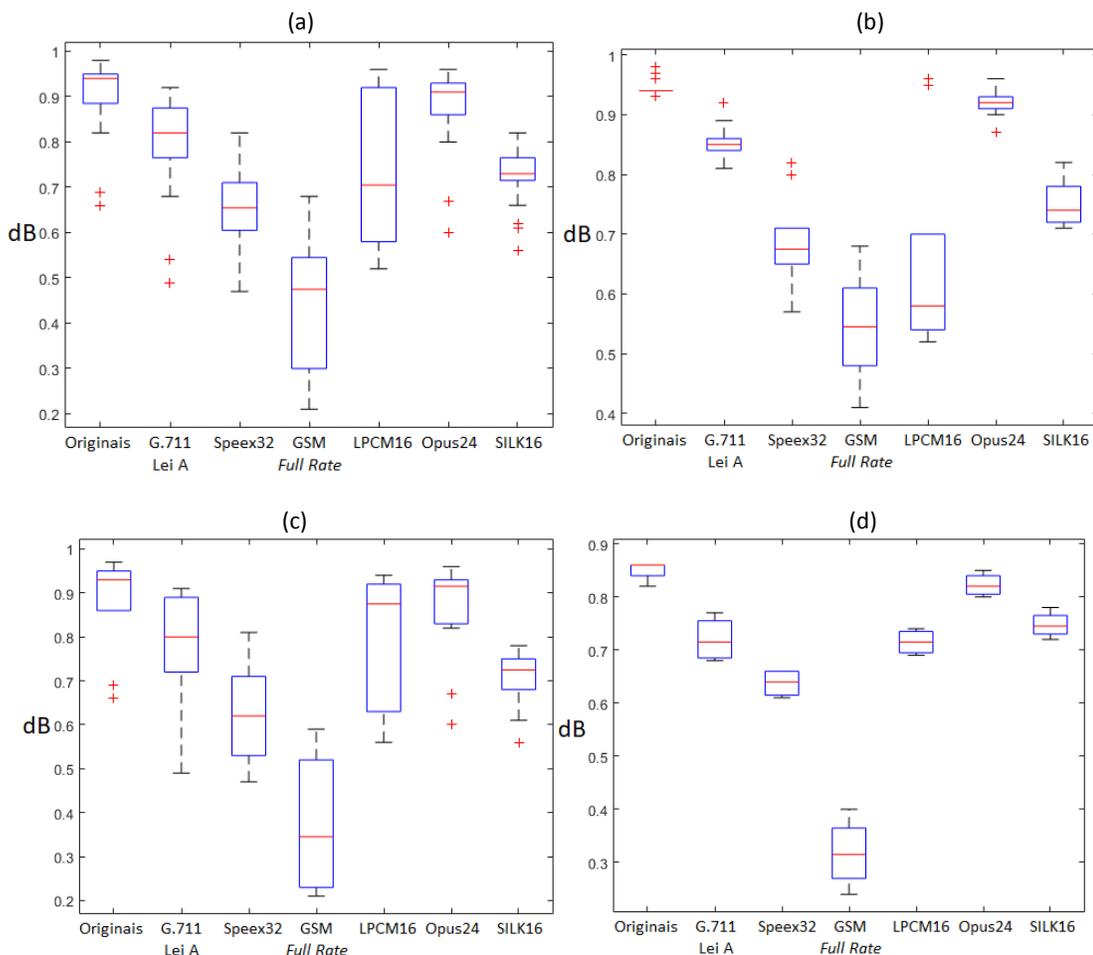
Observa-se que, o codec que apresentou maior facilidade em manter o parâmetro *shimmer* foi o G.711 Lei A e, de forma geral, todos os demais tiveram maior dificuldade, com destaque para o codec SILK16 que obteve pior resultado para o *shimmer* (Figura 15a).

Em relação apenas aos sinais normais sintetizados, identifica-se que o codec G.711 Lei A foi aquele que melhor conseguiu manter o *shimmer*. Para este tipo de sinal, o SILK16 foi o codec que apresentou maior dificuldade em conservar o *shimmer* (Figura 15b).

Em relação apenas aos sinais rugosos sintetizados, nota-se que o codec Opus24 apresentou melhor resultado em conservar o *shimmer*, enquanto todos os demais apresentaram dificuldade, destacando-se o SILK16 apresentou maior dificuldade (Figura 15c).

Em relação apenas aos sinais soprosoos sintetizados, percebe-se que o codec G.711 Lei A foi aquele que apresentou melhor índice do *shimmer*, seguido pelo Opus24, enquanto o SILK16 apresentou maior inconformidade (Figura 15d).

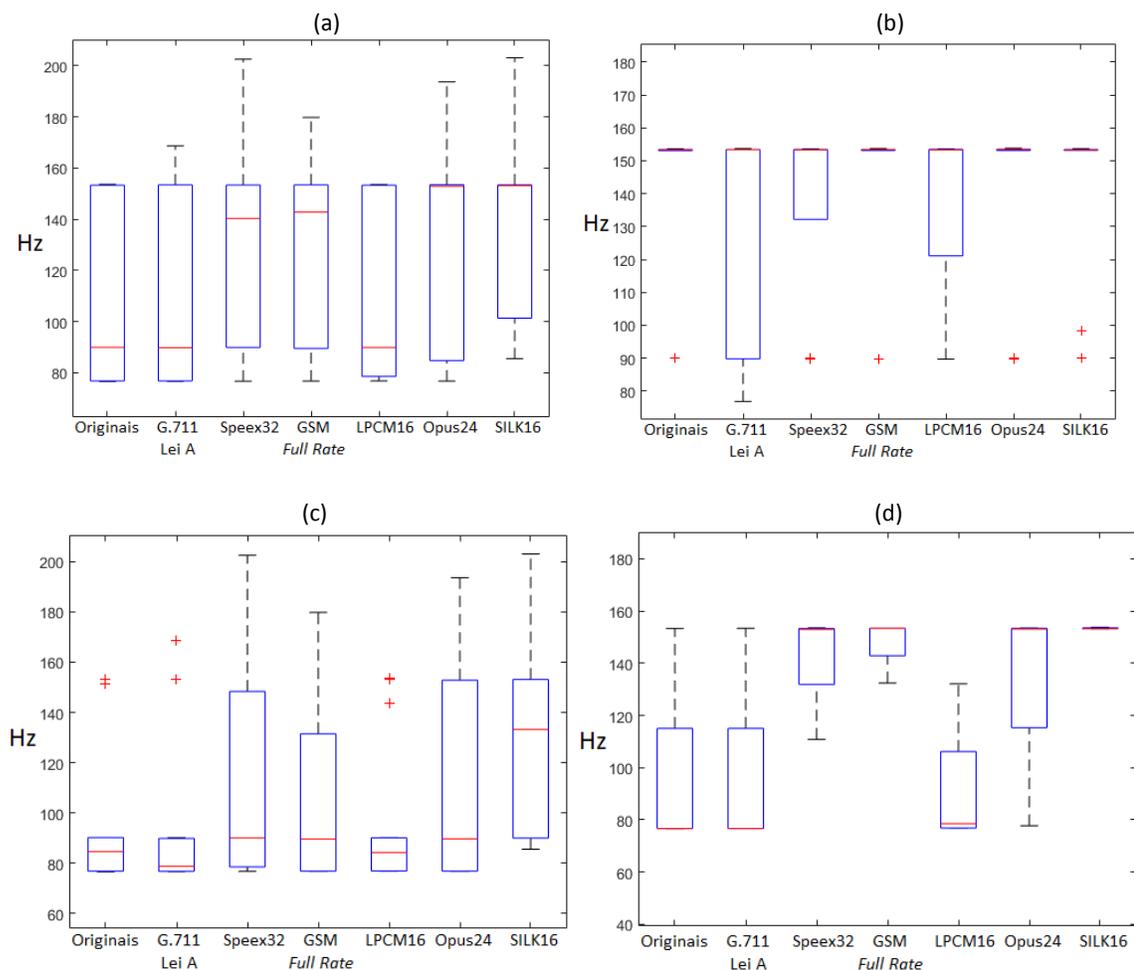
Figura 16 – Valores de GNE (em decibel) dos sinais sintetizados, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosoos.



Considerando o GNE para todos os sinais sintetizados, o codec que melhor manteve a integridade deste parâmetro foi o Opus24, os demais apresentaram maior dificuldade, com destaque para o GSM *Full Rate*, que obteve expressivamente o pior resultado (Figura 16a). A dificuldade do GSM *Full Rate* em manter o GNE, associado ao resultado obtido por ele na Tabela 6, onde conseguiu detectar a presença de sopro em apenas um dos sinais soprosos após as transmissões, podem estar relacionados ao descrito por Lopes, Cavalcante e Costa (2014), onde afirma que o GNE possui correlação negativa em relação ao desvio de sopro.

Quando considerado apenas os sinais normais sintetizados, ou apenas os sinais rugosos ou ainda apenas os sinais soprosos, os resultados apresentados em relação ao conjunto de sinais sintetizados para o parâmetro GNE se repetiram, destacando-se novamente o codec Opus24 com os melhores resultados e o GSM *Full Rate* com os piores (Figuras 16b, 16c e 16d).

Figura 17 – Valores da frequência fundamental (em hertz) dos sinais sintetizados, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.

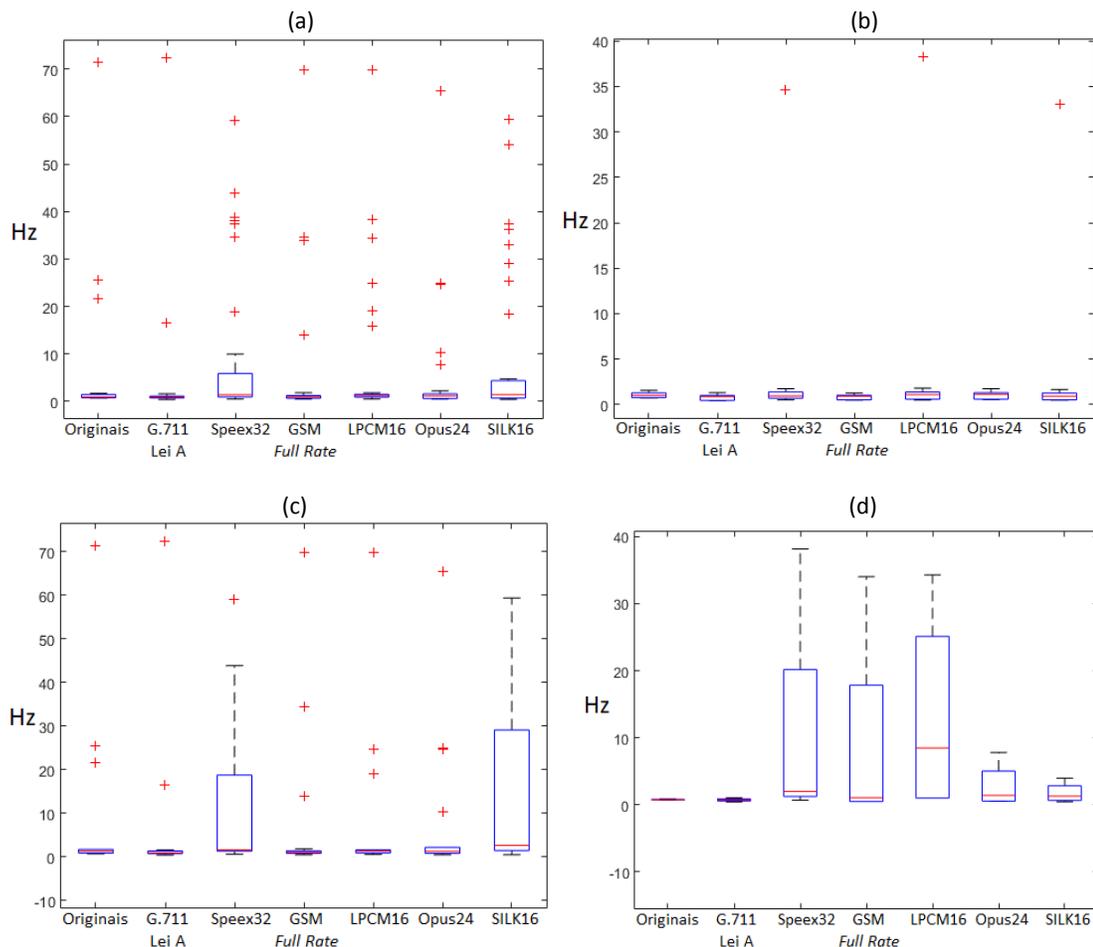


Considerando todos os tipos de sinais sintetizados, percebe-se que, para a frequência fundamental, os codecs G.711 Lei A e LPCM16 foram os que obtiveram maior facilidade em manter este parâmetro (Figura 17a), porém quando considerado apenas os sinais normais, esses dois codecs foram os que apresentaram maior dificuldade (Figura 17b). Ainda em relação a todos os tipos de sinais sintetizados, o SILK16 foi que apresentou maior dificuldade em manter a integridade da frequência fundamental (Figura 17a).

Em relação apenas aos sinais rugosos sintetizados ou apenas aos soprosos sintetizados, percebe-se novamente que os codecs G.711 Lei A e LPCM16 foram aqueles que apresentaram maior facilidade em manter os valores da frequência fundamental (Figuras 17c e 17d).

Destaca-se ainda que, em relação apenas aos sinais soprosos sintetizados, o codec SILK16 resultou em uma frequência fundamental de aproximadamente 153 Hz em praticamente todos os sinais (Figura 17d).

Figura 18 – Valores do desvio padrão da frequência fundamental (em hertz) dos sinais sintetizados, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.



Para o desvio padrão da frequência fundamental, percebe-se que para todos os tipos de sinais sintetizados, houve o surgimento de diversos valores atípicos (*outliers*) em praticamente todos os codecs. Identifica-se ainda que, de forma geral, todos os codecs obtiveram desempenhos parecidos, com exceção do Speex32 e do SILK26 que apresentaram os piores resultados (Figura 18a).

Em relação aos sinais normais sintetizados, verifica-se que houve pouca variação do desvio padrão da frequência fundamental em relação a todos os codecs estudados, porém, observa-se o surgimento de valores atípicos (*outliers*) nos codecs Speex32, LPCM16 e SILK16 (Figura 18b).

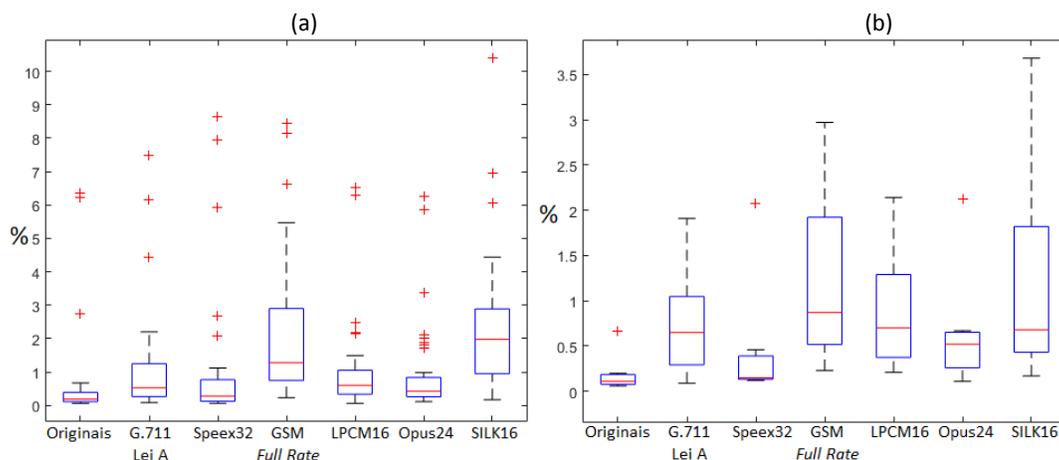
Considerando apenas os sinais rugosos sintetizados, observa-se que todos os codecs obtiveram desempenhos semelhantes para o desvio padrão da frequência fundamental, com exceção do Speex32 e SILK16 que apresentaram os piores resultados (Figura 18c).

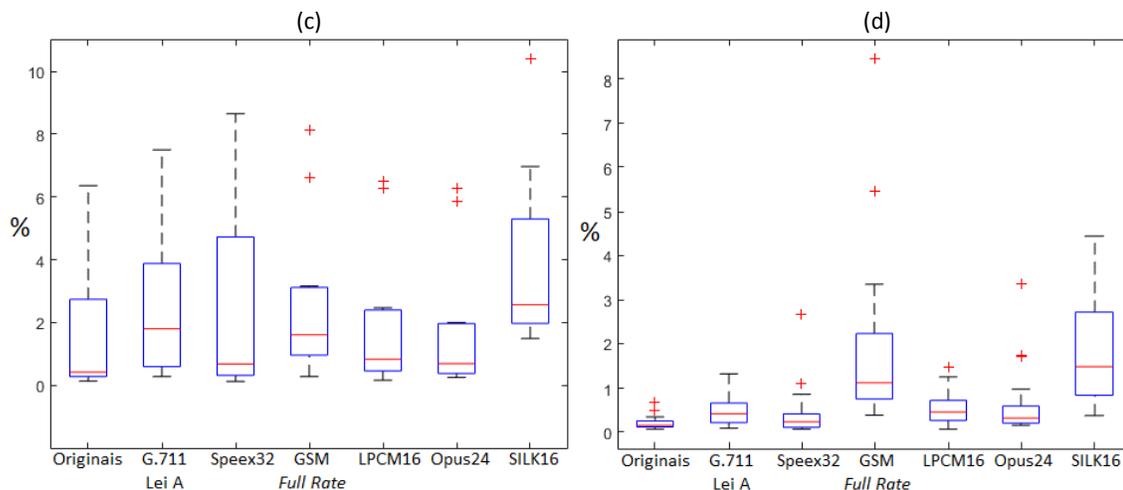
Para os sinais soprosos sintetizados, percebe-se que o codec G.711 Lei A foi aquele que conseguiu o melhor desempenho em relação a manter o desvio padrão da frequência fundamental (Figura 18d).

## 6.2.2. Sinais reais

As Figuras 19, 20, 21, 22 e 23 apresentam, respectivamente, gráficos em formato *boxplot* das medidas acústicas *jitter*, *shimmer*, GNE, frequência fundamental e desvio padrão da frequência fundamental, extraídas dos 36 sinais originais reais e dos seus respectivos sinais recuperados, após as transmissões utilizando os seis codecs estudados. Esses gráficos foram gerados a partir das tabelas presentes no APÊNDICE B.

Figura 19 – Valores do *jitter* (em porcentagem) dos sinais reais, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.





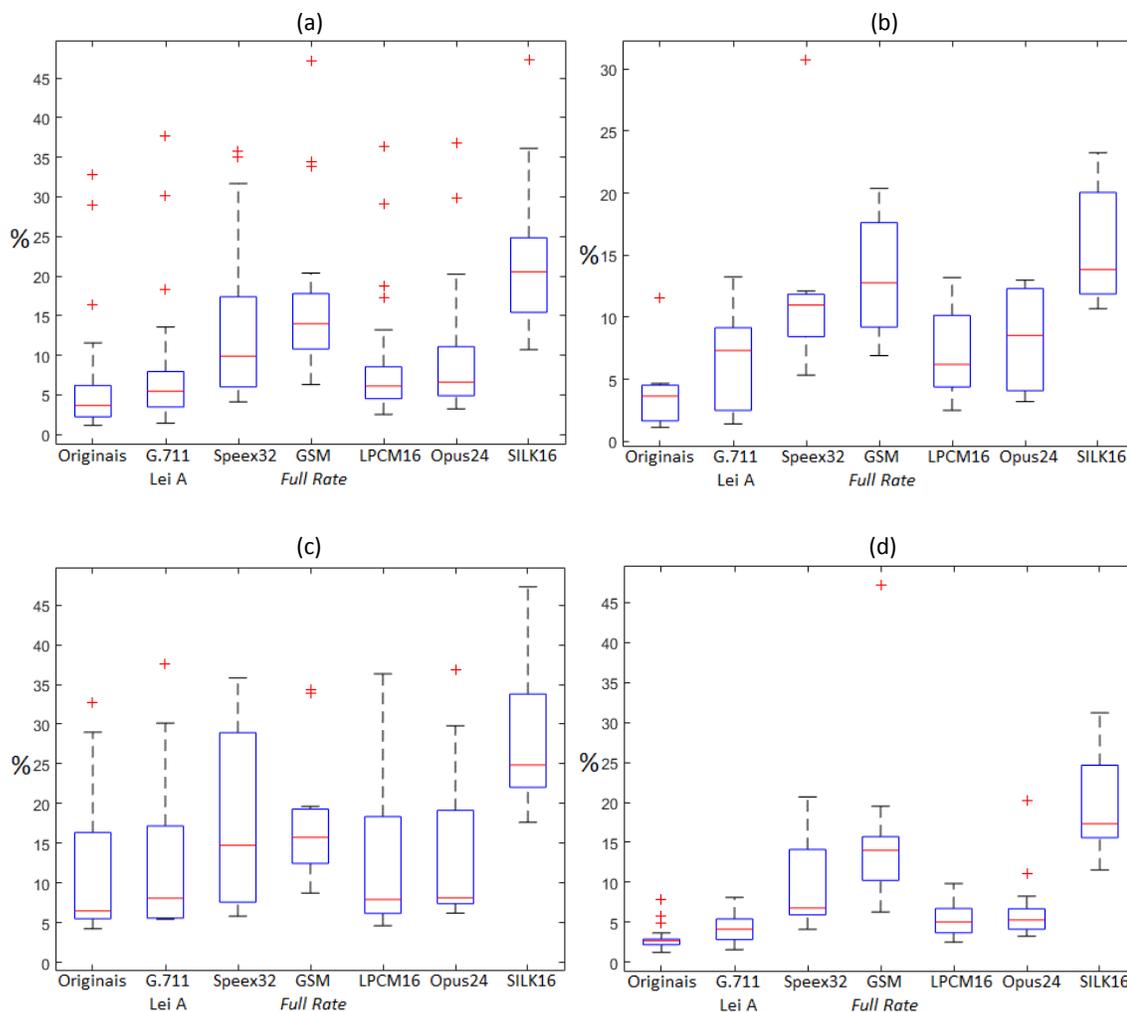
Observa-se que, para todos os tipos de sinais reais analisados, o parâmetro *jitter* sofreu bastante alteração após as transmissões em relação a todos os codecs estudados, inclusive com a adição de valores atípicos (*outliers*). Aqueles que melhor conseguiram manter a integridade deste parâmetro foram os codecs Speex32 e Opus24. Ambos possuem as maiores taxas de amostragem (32 e 24 kamostras/s, respectivamente). Já os codecs GSM *Full Rate* e SILK16 obtiveram maior dificuldade em manter este parâmetro inalterado (Figura 19a).

Analisando-se apenas os sinais normais, identifica ainda maior alteração do parâmetro *jitter* quando comparado ao conjunto de sinais reais. Para este tipo de sinal, o GSM *Full Rate* e SILK16 obtiveram maior dificuldade em manter a integridade do *jitter* (Figura 19b).

Em relação apenas aos sinais rugosos, verifica-se novamente uma dificuldade por parte de todos os codecs em manter a integridade do *jitter* (Figura 19c).

Quando são analisados apenas os sinais soprosos, é possível perceber que para o parâmetro *jitter* novamente todos os codecs tiveram dificuldade em manter sua integridade, destacando-se os codecs Speex32 e Opus24 com os melhores resultados e o GSM *Full Rate* e SILK16 com os piores (Figura 19d).

Figura 20 – Valores do *shimmer* (em porcentagem) dos sinais reais, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.



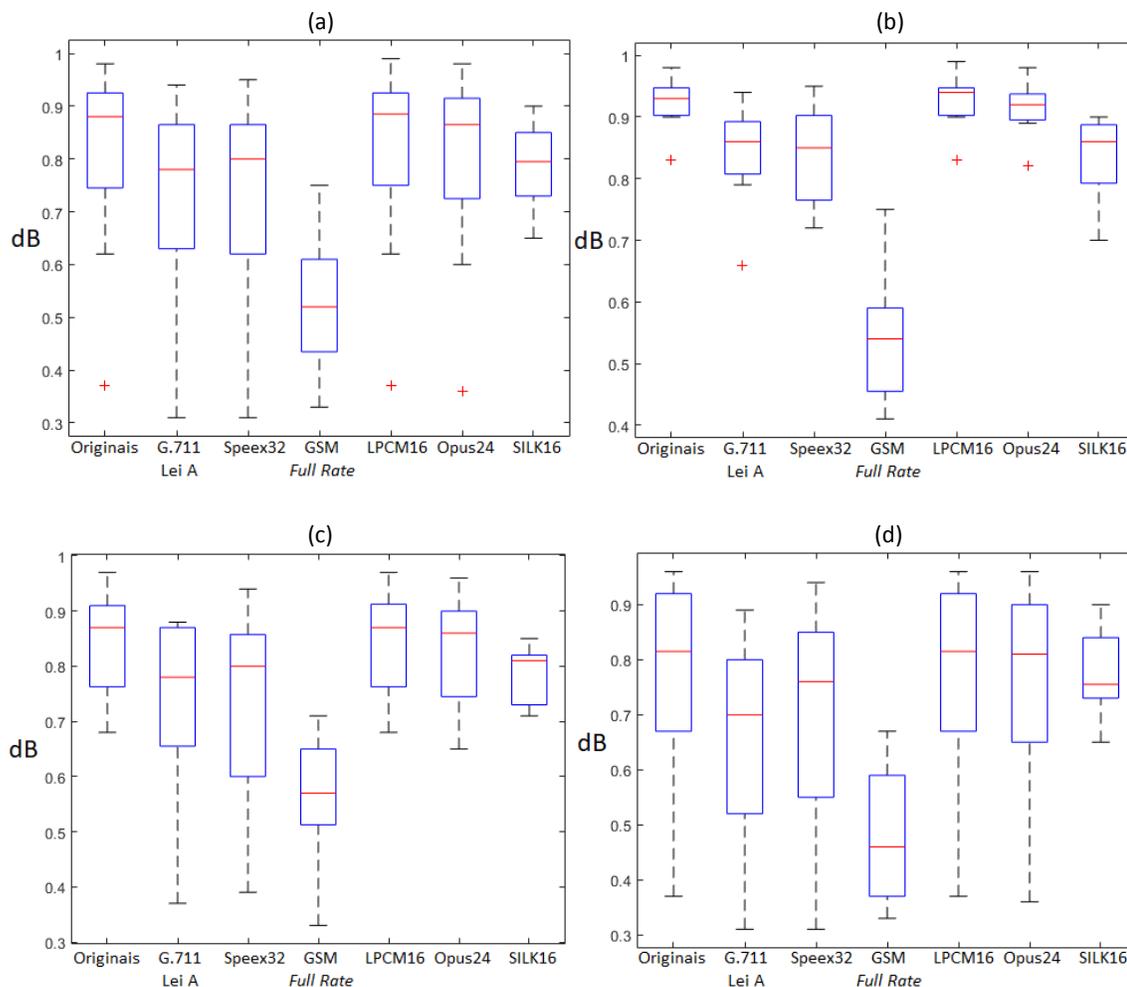
De forma geral, observa-se que, para todos os tipos de sinais reais, os codecs G.711 Lei A e LPCM16 foram aqueles que melhor mantiveram o *shimmer*. O codec que teve maior dificuldade em manter este parâmetro foi o SILK16 (Figura 20a).

Analisando-se os resultados apenas dos sinais normais, percebe-se que todos os codecs tiveram dificuldades em manter a integridade do *shimmer* (Figura 20b).

Quando são verificados apenas os sinais rugosos, é possível perceber que os codecs G.711 Lei A, LPCM16 e Opus24 foram aqueles que apresentaram melhores resultados. Para este tipo de sinal, o codec que apresentou pior desempenho foi o SILK16 (Figura 20c).

Em relação apenas aos sinais soprosos, percebe-se que todos os codecs utilizados neste estudo tiveram dificuldades em manter os valores do *shimmer*. Destaca-se o SILK16 que apresentou pior resultado (Figura 20d).

Figura 21 – Valores do GNE (em decibel) dos sinais reais, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.

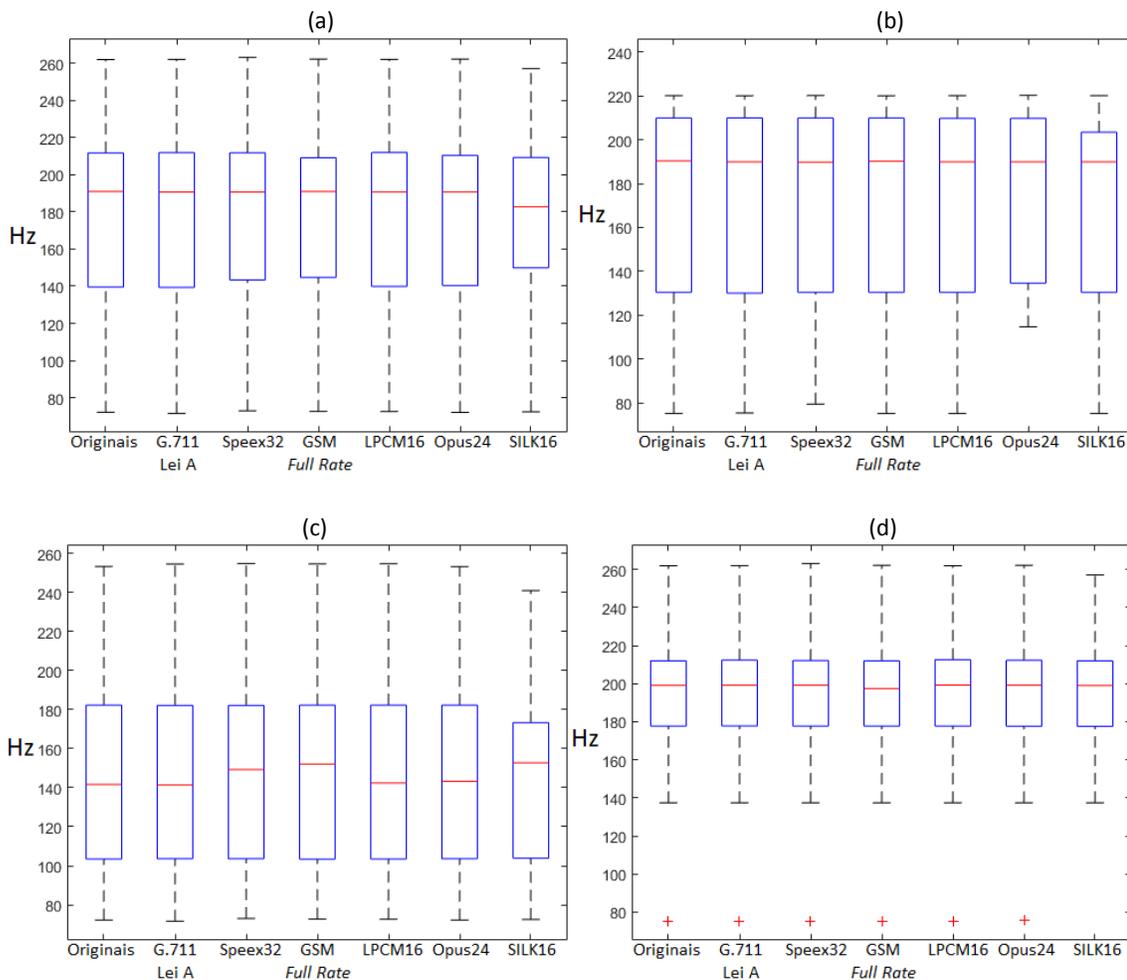


De forma geral, para todos os tipos de sinais, observa-se que os codecs LPCM16 e Opus24 foram aqueles que obtiveram os melhores resultados em manter o parâmetro GNE. Já o GSM *Full Rate* foi aquele que teve maior dificuldade (Figura 21a). A dificuldade do GSM *Full Rate* em manter o GNE, associado ao resultado obtido por ele na Tabela 8, onde conseguiu detectar a presença de sopro em apenas oito sinais soprosos após as transmissões, podem estar relacionados ao descrito por Lopes, Cavalcante e Costa (2014), onde afirma que o GNE possui correlação negativa em relação ao desvio de sopro.

Considerando-se isoladamente apenas os sinais normais ou apenas os rugosos ou ainda apenas os soprosos, percebe-se as mesmas condições anteriores, ou seja, o melhor

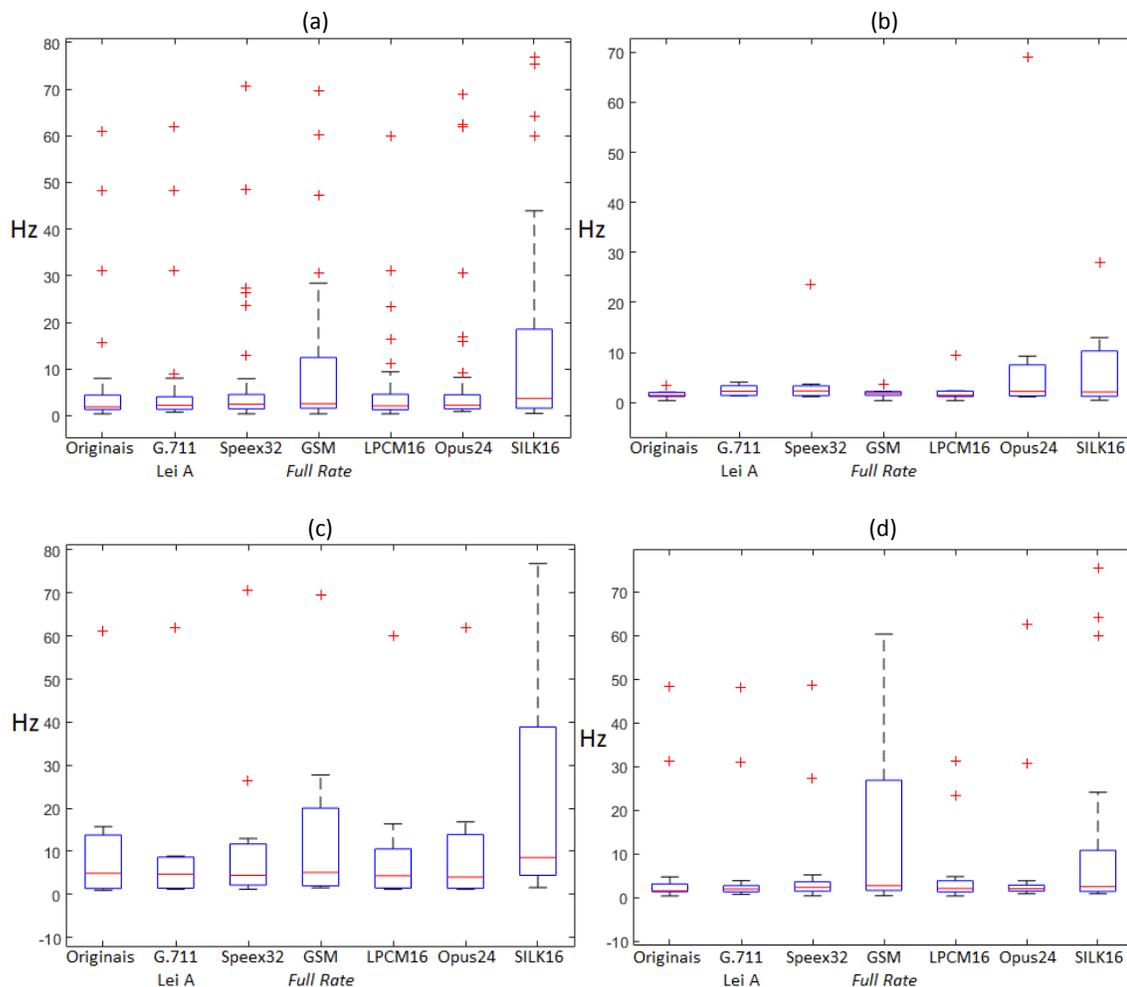
resultado se deu para o codec LPCM16, seguido pelo Opus24 e o pior resultado foi apresentado pelo codec GSM *Full Rate* (Figuras 21b, 21c e 21d).

Figura 22 – Valores da frequência fundamental (em hertz) dos sinais reais, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais sprosos.



Quanto ao parâmetro frequência fundamental, identifica-se que dentre as métricas estudadas para os sinais reais, foi aquela que apresentou melhor resultado para todos os casos analisados. Destaca-se o codec SILK16 que, entre todos os codecs avaliados, foi aquele que apresentou o pior resultado para a frequência fundamental, especialmente em relação aos sinais rugosos (Figuras 22a, 22b, 22c e 22d).

Figura 23 – Valores do desvio padrão da frequência fundamental (em hertz) dos sinais reais, obtidos antes (valores originais) e após as transmissões, distribuídos de acordo com sua classificação inicial, sendo: (a) todos os tipos de sinal, (b) sinais normais, (c) sinais rugosos e (d) sinais soprosos.



Considerando a métrica de desvio padrão da frequência fundamental percebe-se que, para todos os tipos de sinais reais, os codecs analisados tiveram desempenho semelhante, obtendo pouca variação em relação aos sinais originais, com exceção dos codecs GSM *Full Rate* e SILK16, que tiveram resultados inferiores aos demais. Observa-se também o surgimento de diversos valores atípicos (*outliers*) em praticamente todos os codecs (Figura 23a).

Quando considerados apenas os sinais normais reais, os codecs Opus24 e SILK16 encontraram maior dificuldade em manter o desvio padrão da frequência fundamental (Figura 23b).

Analisando-se apenas os sinais rugosos reais, é possível perceber que o codec Opus24 foi aquele que melhor manteve a integridade do parâmetro desvio padrão da

frequência fundamental, destacando-se novamente o codecs GSM *Full Rate* e o SILK16 com os piores resultados (Figura 23c).

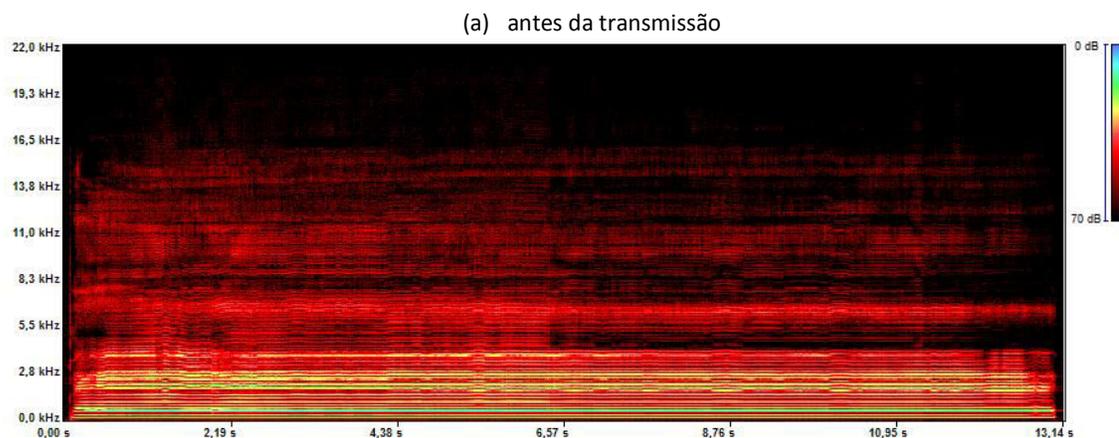
Para os sinais soprosos reais, identifica-se que todos os codecs analisados tiveram desempenho semelhante em manter a integridade, com exceção dos codecs GSM *Full Rate* e SILK16, que tiveram maior dificuldade (Figura 23d).

### 6.2.3. Análise de espectrogramas

Nesta seção serão apresentados alguns espectrogramas (gráficos no plano tempo x frequência), gerados a partir dos sinais, antes e após as transmissões, com o intuito de observar as faixas de frequência ocupadas pelos sinais. Estes gráficos foram obtidos com o auxílio do *software* Fonoview, versão 4<sup>12</sup>.

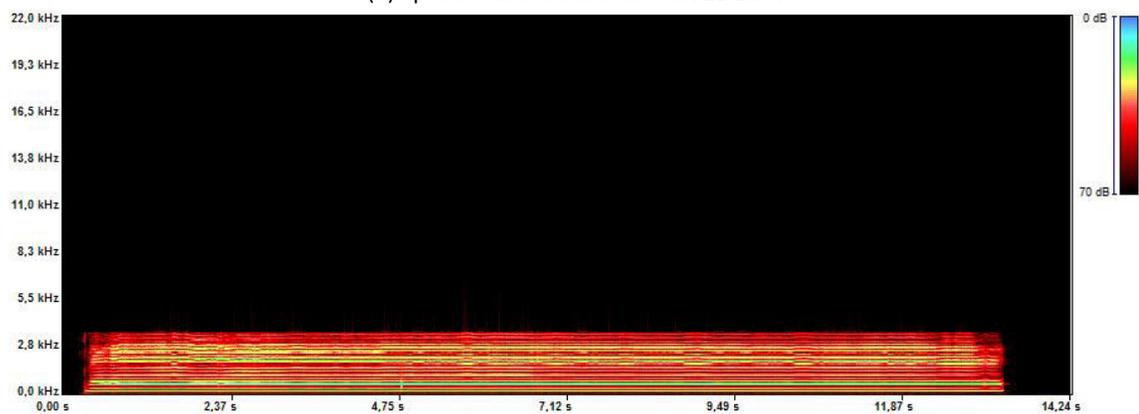
A Figura 24 apresenta os espectrogramas do sinal real número 9, antes e após as transmissões, com o uso de cada um dos codecs estudados, o qual foi avaliado antes da transmissão como sendo soproso. Após as transmissões, com os codecs G.711 Lei A, GSM *Full Rate* e LPCM16, ele foi considerado normal, enquanto que após as transmissões com os codecs Speex32, Opus24 e SILK16 foi considerado rugoso.

Figura24 – Espectrogramas (tempo x frequência) do sinal real 9 antes da transmissão e após as transmissões.

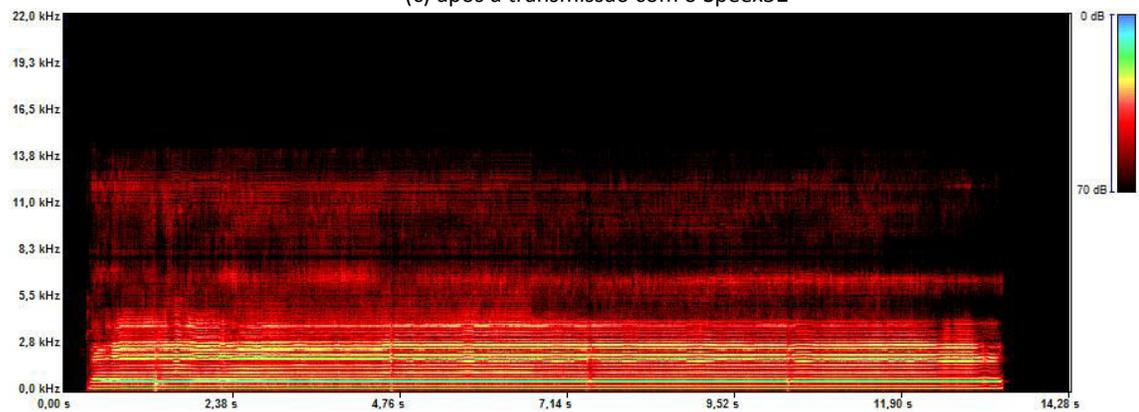


<sup>12</sup> Maiores informações em: <http://www.ctsinformatica.com.br/fonoaudiologia/voz/fonoview-analise-da-comunicacao-oral-em-tempo-real>

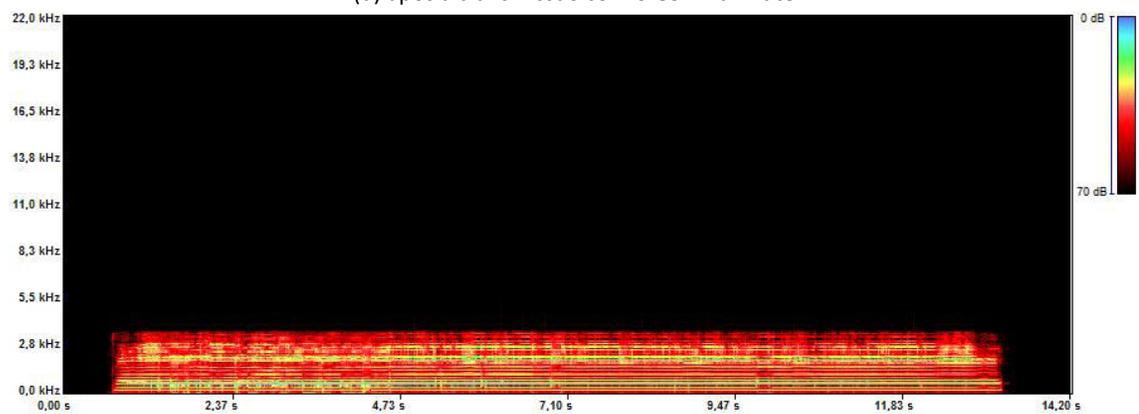
(b) após a transmissão com o G.711 Lei A



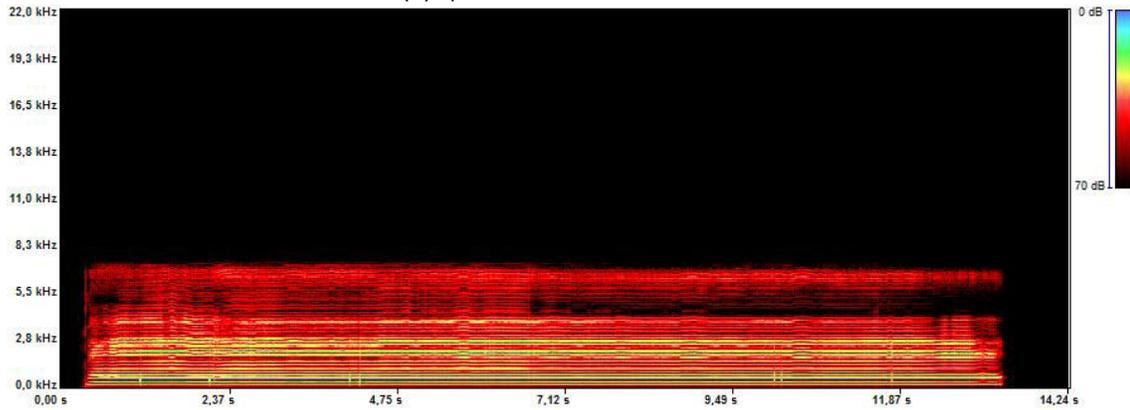
(c) após a transmissão com o Speex32



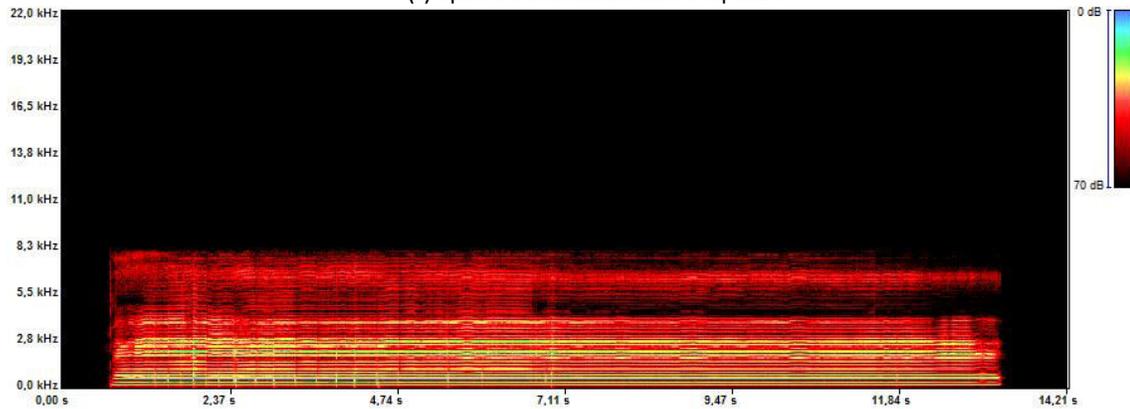
(d) após a transmissão com o GSM Full Rate



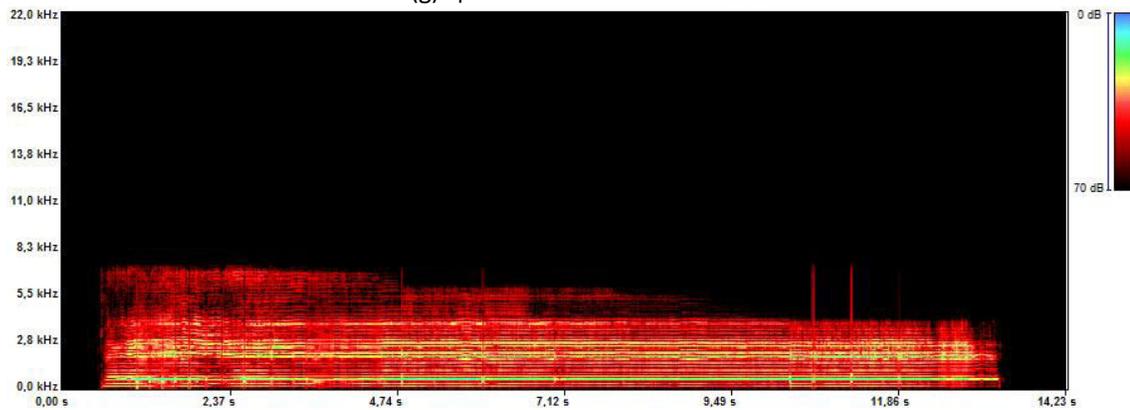
(e) após a transmissão com o LPCM16



(f) após a transmissão com o Opus24



(g) após a transmissão com o SILK16



Na Figura 24a verifica-se a existência de energia na faixa de frequência de 0 até 22kHz, visto que a taxa de amostragem dos sinais originais extraídos da base de dados de sinais reais é de 44100 amostras/s. Percebe-se também um maior nível de energia (em amarelo) na faixa de frequência de 0 até 4kHz, uma vez que este é neste intervalo onde a maior parte da informação da voz está concentrada.

Nas Figuras 24b e 24d a faixa de frequência onde existe energia é apenas de 0 a 4kHz, isto ocorre devido aos codecs G.711 Lei A (Figura 24b) e GSM *Full Rate* (Figura 24d) possuírem taxa de amostragem de 8 kamostras/s, fazendo com que apenas a energia presente nesta faixa de frequência seja preservada (Teorema de Nyquist – seção 3.1.2).

Na Figura 24c a energia preservada após a transmissão ficou em torno da faixa de frequência de 0 a 16kHz, visto que o codec Speex32 possui taxa de amostragem de 32 kamostras/s.

Já nas Figuras 24e e 24g a energia preservada se estendeu na faixa de frequência de 0 a 8kHz, uma vez que os codecs LPCM16 (Figura 24e) e SILK16 (Figura 24g) possuem taxa de amostragem de 16 kamostras/s.

A Figura 24f apresenta a preservação da energia do sinal original apenas na faixa de frequência de 0 a 8,3kHz. Esperava-se que esta faixa de frequência se estendesse de 0 a 12kHz, uma vez que o codec Opus24 possui taxa de amostragem de 24 kamostras/s. Acredita-se que isto pode ter ocorrido devido a forma de implementação deste codec empregada no servidor Asterisk ou no *softphone* Microsip.

Na Figura 24g percebe-se ainda a diminuição da intensidade da energia a partir dos 7 segundos e o surgimento de ruídos entre os tempos 10 e 12 segundos. Acredita-se que esses dois fatores sejam ocasionados devido em virtude algoritmo de compressão utilizado pelo SILK16, uma vez que este comportamento não foi evidenciado em nenhum outro codec.

### **6.3. Desempenho individual dos codecs**

Esta seção apresenta uma síntese do desempenho produzido por cada codec estudado, em relação aos resultados obtidos das análises perceptivo-auditiva e acústica.

#### **6.3.1. Codec G.711 Lei A**

- Tanto para os sinais sintetizados quanto para os reais não obteve índices acima de 67% para a Sensibilidade nem para a Acurácia. Este resultado pode estar relacionado à sua taxa de amostragem de apenas 8 kamostras/s, fazendo com que informações importantes da voz que estejam relacionadas aos desvios vocais não sejam consideradas por este codec;
- Tanto para o conjunto de sinais sintetizados quanto para o conjunto de sinais reais apresentou os melhores resultados para os parâmetros de *shimmer*, frequência fundamental e desvio padrão da frequência fundamental, indicando ser promissor para

a utilização dessas métricas no procedimento de avaliação da qualidade vocal à distância.

### **6.3.2. Codec Speex32**

- Tanto para os sinais sintetizados quanto para os reais apresentou os piores resultados para os índices de Especificidade, parecendo ter introduzido ruído no sinal, confundindo o avaliador quanto à ausência de desvios vocais em sinais normais;
- Obteve resultados acima de 70% em relação à Sensibilidade, sendo o codec que apresentou o melhor resultado para esta característica quando comparado os sinais sintetizados e reais.

### **6.3.3. Codec GSM Full Rate**

- Tanto para os sinais sintetizados quanto para os reais apresentou os piores resultados para os índices de Sensibilidade e Acurácia, além do segundo pior resultado para a Especificidade, indicando ser um codec pouco adequado para realização da análise perceptivo-auditiva;
- Tanto para o conjunto de sinais sintetizados quanto para o conjunto de sinais reais apresentou os piores resultados para as métricas de *jitter* e GNE, indicando ser um codec precário para a avaliação dessas métricas, especialmente o GNE, obtendo um desempenho bastante ruim.

### **6.3.4. Codec LPCM16**

- Apresentou resultados bastante distintos entre as avaliações perceptivo-auditivas dos sinais sintetizados e reais, como por exemplo, em relação a Especificidade, onde obteve um dos melhores índices para os sinais reais e um dos piores para os sinais sintetizados;
- Em relação à análise acústica, considerando as duas bases de dados, não houve coincidências de resultados em nenhum dos casos onde o LPCM16 se destacou.

### **6.3.5. Codec Opus24**

- Tanto para os sinais sintetizados quanto para os reais, obteve índices para a Especificidade e para a Acurácia acima de 85%, se mostrando como um codec promissor para a análise perceptivo-auditiva no que se refere à detecção correta da

ausência de desvios vocais e à identificação do tipo de sinal (normal, rugoso ou sopro);

- Tanto para o conjunto de sinais sintetizados quanto para o conjunto de sinais reais apresentou os melhores resultados para as métricas de *jitter* e GNE, indicando ser um codec promissor para a avaliação dessas métricas.

### 6.3.6. Codec SILK16

- Apesar de obter índices de Acurácia e Sensibilidade acima de 75% para os sinais sintetizados, o mesmo não aconteceu para os sinais reais;
- Apresentou os piores resultados para o *shimmer*, frequência fundamental e desvio padrão da frequência fundamental em relação aos sinais sintetizados e também aos sinais reais, se mostrando um codec precário para a avaliação dessas métricas.

O Quadro 1 apresenta o resumo dos resultados obtidos por cada um dos codecs estudados em relação aos índices da análise perceptivo auditiva e as métricas da análise acústica, quando consideradas as duas bases de dados.

Quadro 1 – Resumo dos resultados.

Resultado	Especificidade	Sensibilidade	Acurácia	Jitter	Shimmer	GNE	F0	Desvio padrão F0
<b>Melhor(es)</b>	Opus24	Speex32	Opus24	Opus24	G.711 Lei A	Opus24	G.711 Lei A, Speex32, GSM Full Rate, LPCM16, Opus24	G.711 Lei A, Speex32, LPCM16, Opus24
<b>Pior(es)</b>	Speex32	GSM Full Rate	GSM Full Rate	GSM Full Rate	SILK16	GSM Full Rate	SILK16	GSM Full Rate, SILK16

## 7. CONSIDERAÇÕES FINAIS

Esta pesquisa buscou investigar os efeitos produzidos pelos codecs G.711 Lei A, Speex32, GSM *Full Rate*, LPCM16, Opus24 e SILK16 em um processo de avaliação da qualidade vocal à distância. Para isso, construiu-se um cenário de transmissão, baseado na tecnologia VoIP e transmitiu-se sinais de vozes sintetizados e reais entre um computador cliente e um servidor, utilizando-se dos codecs mencionados. Antes e após as transmissões, realizou-se a análise perceptivo-auditiva e acústica dos sinais com o objetivo de avaliar o desempenho de cada codec em manter a qualidade vocal.

Os resultados da análise perceptivo-auditiva foram apresentados através das características de Especificidade, Sensibilidade e Acurácia e mostraram que, para os conjuntos de sinais sintetizados e reais, o codec Opus24 foi o que obteve o melhor índice de Especificidade e de Acurácia entre os codecs estudados. Já o melhor desempenho para a Sensibilidade foi alcançado pelo codec Speex32, que neste estudo se mostrou propício à identificação da presença de desvios vocais em sinais desviados.

Ainda para a análise perceptivo-auditiva, identificou-se que o codec GSM *Full Rate* mostrou-se pouco adequado, obtendo os piores índices para avaliação da Sensibilidade e da Acurácia entre os codecs estudados. Já o codec Speex32 não se mostrou uma boa escolha para avaliar pacientes que não possuam desvios vocais, pois apresentou o pior resultado para a Especificidade.

Foi observada, ainda, uma maior preservação do desvio rugosidade para a maioria dos codecs. Já o desvio soproidade foi mantido na maior parte dos sinais, apenas pelos codecs com taxas de amostragem acima de 8 kamostras/s.

Para a análise acústica foram utilizadas as métricas de *jitter*, *shimmer*, GNE, frequência fundamental e desvio padrão da frequência fundamental, onde destacam-se positivamente o codec Opus24 em relação ao *jitter* e GNE e o codec G.711 Lei A em relação ao *shimmer*. No que se refere ao desempenho dos codecs em relação à frequência fundamental e ao desvio padrão da frequência fundamental, foram obtidos resultados muito semelhantes.

Os piores resultados da análise acústica ficaram a cargo dos codecs GSM *Full Rate* e SILK16, sendo o primeiro destacando-se negativamente em relação ao *jitter* e GNE e o segundo em relação ao *shimmer*. Destaca-se que a medida GNE refletiu os resultados obtidos por Lopes, Cavalcante e Costa (2014), que indicou uma correlação negativa desta medida em relação ao desvio soproidade, fato observado neste trabalho, onde verificou-se que o GSM *Full Rate* obteve o pior resultado para o GNE e também foi o codec que teve

maior dificuldade em preservar esse tipo de desvio após as transmissões para ambas as bases de dados estudadas.

Dentre os codecs utilizados neste estudo, o Opus24 mostrou ser o mais adequado para ser utilizado em avaliações da qualidade vocal à distância, pois alcançou índices acima de 68% em todas as taxas, tanto na análise perceptivo-auditiva quanto na acústica, já o GSM *Full Rate* se mostrou precário, uma vez que obteve índices abaixo de 52% em todas as taxas estudadas.

### **7.1. Contribuições da pesquisa**

As principais contribuições desta pesquisa são:

- Aplicação de tecnologias simples e de baixo custo para a avaliação da qualidade vocal à distância;
- Investigação do desempenho de codecs em manter a integridade de determinadas métricas acústicas, utilizados em diversas aplicações;
- Estudo do comportamento das principais métricas acústicas utilizadas na Fonoaudiologia para avaliação da qualidade vocal, quando submetidas à compressão dos codecs.

### **7.2. Sugestão para trabalhos futuros**

Como sugestões para trabalhos futuros, destaca-se:

- Repetir o experimento e analisar os resultados para outros codecs;
- Repetir o experimento, porém variando o cenário de transmissão objetivando analisar o desempenho dos codecs para cada um dos novos cenários de estudo;
- Utilizar um simulador para introduzir os efeitos do canal de transmissão a fim de quantificar as variações das medidas acústicas produzidas por ele;
- Desenvolver um software de apoio ao diagnóstico de desvios vocais a distância, identificando e aplicando automaticamente o codec mais adequado para o tipo de métrica ou característica perceptiva que se deseja avaliar;
- Desenvolver um classificador que possa identificar a distância o tipo de desvio vocal sem o auxílio de um profissional da voz.

## REFERÊNCIAS

- BARAVIEIRA, P. B., BRASOLOTTO, A. G., MONTAGNOLI, A. N., SILVERIO, K. C. A., YAMASAKI, R., BEHLAU, M. **Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica.** CoDAS, 2016.
- BEHLAU, M. MADAZIO, G. LUCERO, J., SCHOENTGEN, J. **Um novo paradigma no ensino da avaliação auditiva de vozes - uso de amostras sintetizadas.** XXI Congresso Brasileiro de Fonoaudiologia. Porto de Galinhas, 2013.
- BRYANT, R., MADSEN, L., MEGGELEN, J. V. **Asterisk: The Definitive Guide.** 4. ed. OREILLY. 2013.
- CARRASCO, E. R., OLIVEIRA, G., BEHLAU, M. **Análise Perceptivo-Auditiva e Acústica da Voz de Indivíduos Gagos.** **Revista CEFAC,** São Paulo, 2010. Disponível em <http://www.scielo.br/pdf/rcefac/2010nahead/179-09.pdf>.
- CHERNEY, L. VUUREN, S. V. **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.
- CIELO, C. A., RIBEIRO, V. V., BASTILHA, G. R., SCHILLING, N. O. **Qualidade de vida em voz, avaliação perceptivoauditiva e análise acústica da voz de professoras com queixas vocais.** Audiology - Communication Research, volume 20. São Paulo, 2015.
- COLTON, R. H., CASPER, J. K., LEONARD, R. **Understanding voice problems: a physiological perspective for diagnosis and treatment.** Lippincott Williams and Wilkins. 3. ed. 2011.
- COSTA, K. L. D.; SANTOS, N. F. **Telemedicina: uma nova forma de assistência médica ao paciente,** Revista Principia, Ano 4, nº 8, p.18-23, João Pessoa: Órgão de Divulgação Científica e Tecnológica do CEFET-PB, 2001.
- COSTA, S. L. do N. C. **Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas.** UFCG. Tese de doutorado, 2008.
- DAS, D. **Development of e-health application for rural community of Bangladesh.** Journal of Theoretical & Applied Information Technology, Volume 21. 2010.
- DEMIRIS, G. **Integration of telemedicine in graduate medical informatics education.** J Am Med Inform Assoc. 2003.
- DIAS, A. E. LIMONGI, J. C. P., BARBOSA, E. R., HSING, W. T. **Telerreabilitação vocal na doença de Parkinson.** Departamento de Neurologia, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo. São Paulo. 2016.
- DIGIUM, Inc. **What is Asterisk?.** 2017. Disponível em: <http://www.asterisk.org>. Acesso em 10/01/2017.
- ENGLERT, M., MADAZIO, G., GIELOW, I., LUCERNO, J., BEHLAU, M. **Erro Perceptivo-Auditivo de Vozes Humanas e Sintetizadas.** Congresso Brasileiro de Fonoaudiologia, 2016.
- ETSI, European Telecommunications Standards Institute. **Mobile technologies GSM.** 2017. Acesso em: 23/01/2017. Disponível em: <http://www.etsi.org/technologies-clusters/technologies/mobile/gsm>.
- FERRARI, R. C. C., JESUS, R. M., SHINODA, A. A., SCHWEITZER, C. M. **Análise de desempenho dos codecs suportados pelo Asterisk com e sem criptografia.** Visão universitária, volume 1. 2014.

- FONSECA, P. N. L. **Push-to-Talk no Celular II: Protocolos**. 2007. Acesso em: 19/09/2017. Disponível em: <http://www.teleco.com.br/tutoriais/tutorialpushtotalk2>.
- FRAZÃO, M. do C. de L.; COSTA, S. L. do N. C.; CORREIA, S. E. N. **Análise Acústica de Sinais de Vozes Patológicas por meio de Características Temporais**. Revista Príncipia, jun. 2013, João Pessoa, PB, nº. 22, p. 28-32. Disponível em: [periodicos.ifpb.edu.br/index.php/principia/article/download/123/98](http://periodicos.ifpb.edu.br/index.php/principia/article/download/123/98).
- GOUDARZI, Mohammad, SUN, Lingfen, IFEACHOR, Emmanuel. **Modelling Speech Quality for NB and WB SILK Codec for VoIP Applications**. Fifth International Conference on Next Generation Mobile Applications and Services. School of Computing and Mathematics, University of Plymouth, UK. 2011.
- GONÇALVES, F. E. de A. **Asterisk PBX: Guia de configuração**. 1.4. ed. Voffice, 2005.
- MOECKE, M. **PCM - Modulação por Código de Pulso: Detalhes da aplicação para telefonia digital**. São Jose - SC. 2006.
- HOWARD, D. ANGUS, J. **Acoustics and Psychoacoustics Applied - Part 2: Audiometry and psychoacoustic testing**. 2010. Disponível em: <http://www.embedded.com/print/4015908>. Acesso em 11/12/2016.
- JAMIESON, David. **Speech Coders and Associated PSQM Values. Test set 1**. Vocal Technologies. 2017. Acesso em: 23/04/2017. Disponível em: <https://www.vocal.com/speech-coders/associated-psqm-values>
- JAMIELSON, D. **SILK Codec**. Vocal Technologies. 2016. Acesso em: 23/10/2017. Disponível em: <https://www.vocal.com/speech-coders/silk/>.
- JOKISCH, Oliver. MARUSCHKE, Michael. **Audio and Speech Coding & Transcoding in Web Real-Time Communication**. Leipzig University of Telecommunications (HfTL). Leipzig, Germany. 2016.
- KARAPANTAZIS, Stylianos. PAVLIDOU, Fotini-Niovi. **VoIP: A comprehensive survey on a promising technology**. Computer Networks. Elsevier, 2009.
- KELLER, Alexandre. **Asterisk na prática**. 2. ed. São Paulo: Novatec, 2011.
- KIM, Ys. **Telemedicine in the USA with focus on clinical applications and issues**. Yonsei Med J. 2004.
- KUROSE, J. F., ROSS, K. W. **Redes de computadores e a Internet: uma abordagem top-down**. 5. ed. São Paulo: Addison Wesley, 2010.
- 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.
- LATHI, B. P. DING, Z. **Sistemas de comunicações analógicos e digitais modernos**. 4. ed. LTC, 2012.
- LAU, L., BRETALAR, M. **Epidemiology of Parkinson's disease**. The Lancet Neurology. 2006. Disponível em: [http://dx.doi.org/10.1016/S1474-4422\(06\)70471-9](http://dx.doi.org/10.1016/S1474-4422(06)70471-9). Acesso em: 12/02/2017.
- LOPES, L. W., CAVALCANTE, D. P., COSTA, P. O. **Intensidade do desvio vocal: integração de dados perceptivo-auditivos e acústicos em pacientes disfônicos**. Revista CODAS, vol 5, pp.382-400, 2014.
- 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.
- MACEDO, D. D. J. **Um Estudo de Estratégias de Sistemas Distribuídos aplicadas a Sistemas de Telemedicina**. UFSC. 2008.
- MALDONADO, J. M. S. de V., MARQUES, A. B., CRUZ, A. **Telemedicina: desafios à sua difusão no Brasil**. Caderno Saúde Pública, volume 32, Rio de Janeiro. 2016.

- MARTÍNEZ, A., VILLARROEL, V., SEOANE, J., POZO, F. **Analysis of information and communication needs in rural primary health care in developing countries**. IEEE Transactions on Information Technology in Biomedicine 9(1). 2015.
- MARTINS, A. E. A., LOLE, D. S., CLARK, F. S. P., MEDEIROS, V. C. **Técnicas de codificação: Pulse Code Modulation** por lei  $\mu$  e lei A. Rio de Janeiro, 2010. Disponível em: [http://www.gta.ufrj.br/grad/10\\_1/codec/Tecnicas\\_de\\_Codificacao.html](http://www.gta.ufrj.br/grad/10_1/codec/Tecnicas_de_Codificacao.html). Acesso em: 04/05/2016.
- MASHIMA, P. A., BIRKMIRE-PETERS, D. P., HOLTEL, M. R., SYMS, M. J. **Telehealth applications in speech-language pathology**. Journal of Healthcare Information Management. 1999.
- MESTON, R. **Sorting Through GSM Codecs: A Tutorial**. EE Times, 2003.
- MONTGOMERY, C. **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>
- PINHEIRO, F. S. R. **PCM - Pulse code modulation**: Departamento de engenharia elétrica, Universidade Federal do Rio Grande do Norte. 2008. Disponível em: <http://www.dee.ufrn.br/pcm.pdf>. Acesso em 13/04/2016.
- RCHANDRA, A. ITU G.711. 2013. Disponível em: <http://www.voip-info.org/wiki/view/ITU+G.711>. Acesso em 21/012017.
- RNP, **Rede Nacional de Ensino e Pesquisas**. Acesso em 18/02/2017. Disponível em: <http://rnp.br>.
- ROGER, L. L. B. **Digitalização de um sinal analógico**. 2012. Disponível em: <http://www.ceset.unicamp.br/~leobravo>. Acesso em 13/04/2016.
- RUTE, **Rede universitária de Telemedicina**. Acesso em: 18/02/2017. Disponível em: <http://rute.rnp.br>
- SAÚDE, Ministério da. Sobre o Telessaúde. 2017. Acesso em 16/10/2017. Disponível em <http://telessaude.saude.ms.gov.br/portal/sobre-o-telessaude-2>.
- SCHOLL, J., LAMBRINOS, L., LINDGREN, A. **Rural Telemedicine Networks Using Storeand-Forward Voice-over-IP**. European Federation for Medical Informatics. IOS Press, 2009.
- SCHULZRINNE, Henning. **Summary of codec characteristics**. Department of Computer Science. Columbia University. 2008. Acesso em 12/03/2017. Disponível em: <http://www.cs.columbia.edu/~hgs/audio/codecs.html>
- SOUZA, M. A., SCHICORSKI, A. D., FERREIRA, K. Z. A., SENGER, L. J., CORTES, O. A. C. **Avaliação de técnicas de codificação de voz para VoIP**. UTFPR, 2007.
- TANENBAUM, A. S. WETHERAL, D. **Redes de Computadores**. 5. ed. São Paulo: Pearson, 2011.
- TEIXEIRA, J. P.; FERREIRA, D. B.; CARNEIRO, S. M. **Análise Acústica Vocal – Determinação do jitter e do shimmer para Diagnóstico de Patologias da Fala**, 2011. Escola Superior de Tecnologia e Gestão Instituto Politécnico de Bragança - Bragança, POR. Disponível em [https://bibliotecadigital.ipb.pt/bitstream/10198/7282/1/artigo\\_publicado.pdf](https://bibliotecadigital.ipb.pt/bitstream/10198/7282/1/artigo_publicado.pdf). Acessado em 10/01/18.
- UNION, International Telecommunication. ITU-T recommendation G.711. Geneva, 1988.
- UNION, International Telecommunication. Recommendation G.726. Geneva, 1990.
- UNION, International Telecommunication. Recommendation P.862. Geneva, 2001. Acesso em 12/03/2017. Disponível em: <http://www.itu.int/rec/T-REC-P.862>
- VIEIRA, V. J. D. **Avaliação de Distúrbios da Voz por meio de Análise de Quantificação de Recorrência**. IFPB. Dissertação de mestrado, 2014.

- VOS, K. **Opus Interactive Audio Codec**. Acesso em 23/10/2017. Disponível em <http://opus-codec.org>.
- WEN, C. L. **Telemedicina e Telessaúde – Um panorama no Brasil**. Revista iP - Informática Pública. Vol. 2, p. 07-15, 2008.
- WOOTTON, R. **Recent advances: Telemedicine**. BMJ. 2001.
- YAMASAKI, R. **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.
- XIPH. **Speex: A Free Codec For Free Speech**. The Xiph.Org Foundation. Acesso em: 12/03/2017. Disponível em: <https://speex.org>.
- ZHANG, Y., MCGILLIGAN, C., ZHOU, L., VIG, M., JIANG, J. **Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps**. The Journal of the Acoustical Society of America, vol. 115, p. 2270, 2004.
- ZHU, Y, WITT, R. E., MACCALLUM, J. K., JIANG, J. J. **Effects of the Voice over Internet Protocol on Perturbation Analysis of Normal and Pathological Phonation**. Folia Phoniatica et Logopaedica. 2010.

## APÊNDICE A

### MEDIDAS ACÚSTICAS DOS SINAIS SINTETIZADOS

Este apêndice refere-se aos dados das medidas acústicas extraídas dos sinais sintetizados com o auxílio do software VoxMetria, versão 4, da CTS Informática, antes e após as transmissões.

Tabela A.1. – Valores de *jitter* em porcentagem (%) obtidos antes e após as transmissões dos sinais sintetizados

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	7,67	7,08	3,84	5,88	2,19	7,00	4,84	Rugoso
2	0,07	0,07	0,06	0,16	4,10	0,84	1,37	Normal
3	1,29	1,53	1,03	0,83	3,62	1,81	0,68	Rugoso
4	3,08	2,23	3,23	3,41	3,16	4,33	3,65	Rugoso
5	10,91	11,52	11,21	8,19	11,05	10,01	14,54	Rugoso
6	0,99	1,21	0,49	1,05	3,54	1,17	1,84	Rugoso
7	0,22	0,23	0,26	0,28	0,64	0,23	1,20	Normal
8	8,95	11,29	10,65	10,53	9,90	10,18	12,39	Rugoso
9	1,90	2,13	2,47	1,59	3,13	3,50	7,78	Rugoso
10	0,57	0,72	0,13	0,25	0,98	1,15	1,43	Soproso
11	0,04	0,04	0,07	0,15	0,05	0,05	2,88	Normal
12	13,45	13,46	13,03	12,29	12,82	12,01	14,53	Rugoso
13	4,26	5,08	0,11	0,79	2,92	2,48	2,79	Soproso
14	0,08	0,08	0,40	0,16	0,14	0,09	0,74	Normal
15	0,04	0,04	0,07	0,14	0,30	0,43	2,55	Normal
16	6,33	8,57	3,49	1,87	4,94	7,45	2,14	Rugoso
17	0,11	0,08	0,07	0,13	1,00	0,30	3,46	Normal
18	0,06	0,06	0,10	0,15	0,24	0,15	1,66	Normal
19	0,07	0,07	0,12	0,17	7,08	0,21	3,33	Normal
20	0,57	0,72	0,15	0,25	0,60	1,05	1,79	Rugoso
21	6,85	7,37	3,96	7,55	6,84	7,24	4,32	Rugoso
22	0,09	0,58	0,49	0,32	1,54	0,08	1,22	Normal
23	4,96	5,37	0,07	1,21	4,84	4,92	2,87	Rugoso
24	0,26	0,20	0,24	0,50	1,29	0,23	0,23	Normal
25	3,79	4,26	0,13	2,58	0,29	2,47	1,72	Rugoso
26	1,76	1,63	2,76	1,15	3,32	1,44	5,72	Rugoso
27	0,99	1,14	0,49	0,75	1,54	1,93	5,14	Rugoso
28	6,86	6,86	0,15	0,22	4,15	3,97	0,75	Normal
29	0,15	0,74	0,15	0,29	0,15	0,16	1,37	Soproso
30	7,12	7,03	0,18	0,41	3,12	4,89	0,39	Soproso
31	6,33	8,57	0,29	2,39	8,95	7,25	4,10	Rugoso
32	0,02	0,02	0,03	0,13	3,57	0,04	2,42	Normal

33	0,00	0,01	0,04	1,03	0,10	0,56	0,60	Normal
34	0,04	0,05	0,11	0,15	0,24	0,05	2,75	Normal
35	3,08	3,59	2,44	3,46	2,76	3,53	1,56	Rugoso
36	7,12	7,71	1,64	0,41	5,46	6,48	2,76	Rugoso

Tabela A.2. – Valores de *shimmer* em porcentagem (%) obtidos antes e após as transmissões dos sinais sintetizados

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	16,60	15,69	20,74	16,71	14,24	17,22	34,60	Rugoso
2	2,67	2,48	2,61	7,91	12,27	9,30	17,06	Normal
3	22,63	21,30	20,82	24,66	20,50	26,31	22,70	Rugoso
4	15,83	17,60	20,77	21,39	16,84	16,32	28,92	Rugoso
5	33,63	32,71	35,00	28,11	26,44	33,41	44,99	Rugoso
6	21,58	20,52	15,77	21,49	24,42	20,99	18,52	Rugoso
7	6,43	6,59	14,21	9,03	9,33	7,43	10,53	Normal
8	26,67	30,30	29,39	30,81	27,41	27,94	40,91	Rugoso
9	39,63	40,38	30,96	38,11	36,52	37,86	42,06	Rugoso
10	8,69	9,44	8,53	11,34	11,34	11,00	21,44	Soproso
11	1,33	1,32	11,52	7,11	1,82	8,72	22,70	Normal
12	50,12	49,09	57,83	52,25	48,73	53,96	58,88	Rugoso
13	8,30	8,91	11,56	9,02	9,35	11,12	18,27	Soproso
14	2,35	2,42	12,93	6,32	3,41	6,60	11,84	Normal
15	1,92	1,85	11,53	7,33	5,36	7,44	22,16	Normal
16	16,38	18,89	16,49	15,39	13,71	19,72	25,05	Rugoso
17	5,39	5,75	8,91	10,52	6,02	10,59	16,08	Normal
18	3,47	3,47	13,16	8,00	6,54	6,65	15,64	Normal
19	2,67	2,41	13,21	7,97	12,88	5,76	23,30	Normal
20	8,69	9,57	10,34	11,68	9,85	10,04	22,94	Rugoso
21	21,14	23,66	18,41	25,52	19,61	21,33	24,84	Rugoso
22	7,03	6,97	13,57	11,72	8,22	8,64	13,03	Normal
23	7,30	7,69	4,65	8,71	12,17	7,70	18,68	Rugoso
24	12,66	11,63	12,65	17,46	11,32	12,30	18,30	Normal
25	7,54	7,82	5,82	8,68	8,43	7,55	17,39	Rugoso
26	33,20	35,56	24,74	32,75	31,92	35,15	45,13	Rugoso
27	21,58	21,37	15,77	22,05	21,04	24,37	27,86	Rugoso
28	11,26	11,26	4,86	7,57	13,18	10,23	11,70	Normal
29	8,50	8,55	13,64	11,78	8,02	8,86	20,06	Soproso
30	11,24	11,52	12,26	11,77	12,30	9,99	12,43	Soproso
31	16,38	18,88	20,18	16,90	18,84	17,41	28,76	Rugoso
32	0,79	0,99	2,20	4,81	6,38	6,91	24,97	Normal
33	0,19	0,58	8,87	7,27	3,57	4,55	11,39	Normal
34	2,07	2,32	11,44	6,29	6,84	4,47	22,90	Normal
35	15,83	18,57	14,28	20,36	17,41	17,96	25,36	Rugoso

36	11,24	12,82	7,62	11,94	10,35	10,52	21,10	Rugoso
----	-------	-------	------	-------	-------	-------	-------	--------

Tabela A.3. – Valores de GNE em Decibel (dB) obtidos antes e após as transmissões dos sinais sintetizados

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	0,66	0,49	0,47	0,23	0,60	0,60	0,78	Rugoso
2	0,94	0,85	0,63	0,62	0,58	0,91	0,72	Normal
3	0,97	0,89	0,75	0,55	0,93	0,95	0,75	Rugoso
4	0,95	0,80	0,65	0,30	0,88	0,93	0,66	Rugoso
5	0,95	0,77	0,49	0,21	0,94	0,92	0,61	Rugoso
6	0,93	0,89	0,81	0,51	0,92	0,92	0,75	Rugoso
7	0,93	0,81	0,67	0,48	0,70	0,90	0,72	Normal
8	0,95	0,77	0,54	0,37	0,92	0,92	0,62	Rugoso
9	0,97	0,89	0,71	0,59	0,93	0,95	0,70	Rugoso
10	0,86	0,74	0,62	0,33	0,74	0,83	0,74	Soproso
11	0,94	0,85	0,67	0,68	0,54	0,93	0,72	Normal
12	0,97	0,90	0,53	0,50	0,92	0,96	0,71	Rugoso
13	0,86	0,77	0,61	0,40	0,69	0,85	0,75	Soproso
14	0,94	0,85	0,71	0,54	0,55	0,92	0,78	Normal
15	0,94	0,84	0,65	0,45	0,54	0,87	0,76	Normal
16	0,69	0,54	0,50	0,23	0,56	0,67	0,73	Rugoso
17	0,96	0,92	0,80	0,41	0,95	0,94	0,82	Normal
18	0,93	0,86	0,71	0,49	0,60	0,90	0,75	Normal
19	0,94	0,84	0,70	0,61	0,58	0,91	0,72	Normal
20	0,86	0,72	0,60	0,32	0,72	0,83	0,72	Rugoso
21	0,93	0,76	0,60	0,22	0,82	0,89	0,56	Rugoso
22	0,98	0,89	0,80	0,57	0,96	0,95	0,78	Normal
23	0,91	0,82	0,63	0,53	0,58	0,90	0,75	Rugoso
24	0,97	0,92	0,82	0,53	0,95	0,96	0,80	Normal
25	0,91	0,82	0,61	0,43	0,63	0,88	0,77	Rugoso
26	0,97	0,91	0,73	0,53	0,90	0,94	0,78	Rugoso
27	0,93	0,89	0,81	0,52	0,92	0,91	0,68	Rugoso
28	0,94	0,86	0,65	0,55	0,63	0,93	0,79	Normal
29	0,82	0,68	0,66	0,24	0,70	0,80	0,72	Soproso
30	0,86	0,69	0,66	0,30	0,73	0,81	0,78	Soproso
31	0,69	0,54	0,47	0,23	0,56	0,67	0,73	Rugoso
32	0,94	0,82	0,57	0,62	0,52	0,92	0,73	Normal
33	0,94	0,81	0,68	0,47	0,54	0,91	0,72	Normal
34	0,94	0,84	0,66	0,55	0,57	0,93	0,71	Normal
35	0,95	0,80	0,68	0,30	0,87	0,92	0,69	Rugoso
36	0,86	0,69	0,65	0,29	0,71	0,82	0,75	Rugoso

Tabela A.4. – Valores de frequência fundamental em Hertz (Hz) obtidos antes e após as transmissões dos sinais sintetizados

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	76,77	76,75	78,49	76,75	76,81	76,74	126,72	Rugoso
2	153,55	153,43	153,19	153,44	153,50	153,19	153,41	Normal
3	90,16	90,13	90,16	90,14	90,06	90,08	90,23	Rugoso
4	76,79	76,79	76,81	76,78	76,82	76,74	139,84	Rugoso
5	83,76	80,71	78,22	79,46	86,29	78,59	104,34	Rugoso
6	89,93	89,77	90,02	89,79	89,88	89,72	89,96	Rugoso
7	153,62	76,85	132,20	153,66	121,10	153,69	153,67	Normal
8	76,85	76,85	86,53	76,76	76,89	79,87	107,55	Rugoso
9	89,58	89,42	89,37	89,37	89,73	89,59	89,46	Rugoso
10	76,59	76,64	152,90	132,45	76,81	152,92	153,26	Soproso
11	153,30	153,32	153,41	153,34	153,29	153,42	153,27	Normal
12	151,35	168,62	202,55	179,79	143,84	193,68	203,09	Rugoso
13	153,32	153,40	153,52	153,44	80,25	153,47	153,35	Soproso
14	153,63	153,70	153,60	153,72	153,58	153,81	153,65	Normal
15	153,34	153,41	153,53	153,26	153,50	153,47	153,47	Normal
16	84,59	76,74	127,53	76,76	82,19	76,74	153,11	Rugoso
17	89,93	89,72	89,72	89,73	89,74	89,85	89,94	Normal
18	153,34	153,43	153,39	153,45	153,46	153,48	153,45	Normal
19	153,35	153,43	153,22	153,43	153,33	153,46	153,36	Normal
20	76,59	76,64	148,38	131,50	82,03	152,87	153,17	Rugoso
21	76,77	76,80	76,72	76,80	76,79	79,96	85,54	Rugoso
22	89,91	89,78	89,96	89,79	89,91	90,02	98,39	Normal
23	153,29	153,20	153,11	153,21	153,48	153,45	153,46	Rugoso
24	89,94	89,82	89,91	89,73	89,83	90,04	90,02	Normal
25	153,30	76,71	153,41	153,42	153,39	153,36	153,35	Rugoso
26	90,16	89,86	89,90	89,88	90,15	90,13	89,95	Rugoso
27	89,93	89,78	90,02	89,80	89,80	89,93	89,91	Rugoso
28	153,33	153,43	153,35	153,45	153,49	153,50	153,37	Normal
29	76,75	76,66	110,84	153,30	132,11	77,72	153,76	Soproso
30	76,68	76,71	153,10	153,40	76,87	153,29	153,15	Soproso
31	84,59	76,74	118,08	76,77	76,82	76,79	152,51	Rugoso
32	153,41	153,42	153,44	153,40	153,22	153,54	153,45	Normal
33	153,33	153,40	153,45	153,45	153,47	153,46	153,36	Normal
34	153,08	153,12	153,25	153,12	153,26	153,15	153,17	Normal
35	76,79	76,78	76,81	76,78	76,88	76,81	148,71	Rugoso
36	76,68	153,36	153,14	153,40	76,86	153,30	153,11	Rugoso

Tabela A.5. – Valores do desvio padrão da frequência fundamental em Hertz (Hz) obtidos antes e após as transmissões dos sinais sintetizados

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	0,80	0,76	9,96	0,78	0,85	0,79	36,30	Rugoso
2	0,75	0,44	1,73	0,50	0,55	1,74	0,56	Normal
3	1,33	0,93	1,49	0,98	1,28	1,44	1,58	Rugoso
4	0,91	0,89	1,04	0,89	0,93	0,95	29,11	Rugoso
5	21,58	16,55	9,53	13,93	24,80	10,31	53,98	Rugoso
6	1,45	1,09	1,28	1,12	1,36	1,18	1,53	Rugoso
7	1,28	1,00	34,58	1,25	38,26	1,26	1,27	Normal
8	1,22	1,16	43,85	1,07	1,26	24,60	37,27	Rugoso
9	1,70	1,34	1,45	1,35	1,61	1,66	1,44	Rugoso
10	0,70	0,73	2,20	34,02	0,99	2,24	0,88	Soproso
11	0,93	0,83	0,93	0,82	0,90	0,90	0,85	Normal
12	71,44	72,34	59,15	69,73	69,83	65,46	59,35	Rugoso
13	0,73	0,43	0,70	0,49	15,93	0,56	0,44	Soproso
14	1,11	0,99	1,11	1,02	1,37	1,16	1,01	Normal
15	0,76	0,42	0,69	1,26	0,58	0,58	0,50	Normal
16	25,51	0,75	37,48	0,79	19,08	0,79	4,76	Rugoso
17	1,38	0,95	0,92	0,95	1,03	1,43	1,64	Normal
18	0,73	0,44	0,66	0,49	0,49	0,57	0,48	Normal
19	0,75	0,44	0,69	0,49	1,12	0,54	0,53	Normal
20	0,70	0,73	18,74	34,51	19,00	2,17	0,85	Rugoso
21	0,91	0,91	0,93	0,91	0,91	24,81	25,39	Rugoso
22	1,55	1,18	1,73	1,19	1,39	1,39	33,02	Normal
23	0,86	1,58	2,11	1,82	0,56	0,48	0,53	Rugoso
24	1,45	1,29	1,39	1,02	1,23	1,29	1,27	Normal
25	0,76	0,72	0,62	0,49	0,57	0,58	0,47	Rugoso
26	1,43	1,35	1,41	1,37	1,30	1,34	1,41	Rugoso
27	1,45	1,09	1,28	1,13	1,29	1,55	1,47	Rugoso
28	0,82	0,47	0,73	0,53	0,68	0,63	0,49	Normal
29	0,87	1,04	38,15	1,64	34,28	7,80	3,97	Soproso
30	0,76	0,72	1,78	0,49	1,01	0,53	1,70	Soproso
31	25,51	0,75	38,74	0,79	0,91	0,88	3,61	Rugoso
32	1,05	0,88	0,95	0,88	1,80	1,14	0,97	Normal
33	0,71	0,41	0,50	0,48	0,55	0,54	0,51	Normal
34	1,07	0,99	1,11	0,99	1,26	1,09	1,20	Normal
35	0,91	0,89	1,03	0,89	0,98	0,93	18,33	Rugoso
36	0,76	0,41	1,72	0,49	1,03	0,59	1,67	Rugoso

## APÊNDICE B

### MEDIDAS ACÚSTICAS DOS SINAIS REAIS

Este apêndice refere-se aos dados das medidas acústicas extraídas dos sinais reais com o auxílio do software VoxMetria, versão 4, da CTS Informática, antes e após as transmissões.

Tabela B.1. – Valores de *jitter* em porcentagem (%) obtidos antes e após as transmissões dos sinais reais

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	0,42	0,76	0,54	1,49	0,83	0,69	1,97	Rugoso
2	0,67	0,51	0,41	2,23	0,87	0,98	0,52	Soproso
3	0,06	0,70	0,06	3,36	0,06	1,72	2,64	Soproso
4	0,45	4,44	8,65	3,00	0,99	2,00	6,05	Rugoso
5	0,10	1,16	0,14	1,78	1,11	0,52	0,17	Normal
6	0,07	0,08	0,10	1,11	0,33	0,18	3,16	Soproso
7	0,10	0,66	0,11	0,38	0,23	0,24	0,37	Soproso
8	0,07	0,90	0,10	1,41	0,87	0,36	0,82	Soproso
9	0,11	0,13	0,09	0,74	0,18	0,20	1,48	Soproso
10	0,34	0,76	0,39	0,75	0,62	0,38	1,25	Soproso
11	2,74	2,20	0,97	2,83	2,19	1,87	1,68	Rugoso
12	0,17	0,21	0,09	1,68	0,38	0,20	2,25	Soproso
13	6,21	7,50	7,95	6,62	6,51	6,27	10,40	Rugoso
14	0,14	0,20	2,68	8,46	0,47	3,38	4,44	Soproso
15	6,36	6,17	5,93	8,14	6,28	5,86	6,97	Rugoso
16	0,36	0,54	0,55	1,02	0,63	0,55	2,66	Rugoso
17	0,66	0,65	0,46	2,97	0,70	0,60	1,16	Normal
18	0,07	0,09	0,12	0,87	0,48	0,11	0,41	Normal
19	0,48	0,41	0,33	0,52	0,43	0,47	3,12	Soproso
20	0,16	0,22	0,63	1,02	0,58	0,29	0,57	Soproso
21	0,11	0,52	0,18	0,85	0,26	0,15	1,27	Soproso
22	0,13	1,73	0,24	1,61	0,19	0,32	2,19	Rugoso
23	0,14	0,36	0,28	1,96	1,49	0,31	0,84	Soproso
24	0,11	0,71	0,18	0,51	0,21	0,26	0,50	Normal
25	0,20	0,22	0,13	0,54	0,34	0,26	3,68	Normal
26	0,20	0,25	0,19	5,47	0,26	0,28	1,51	Soproso
27	0,25	0,28	0,24	0,62	0,64	0,30	2,57	Rugoso
28	0,25	0,49	0,86	1,14	0,72	0,61	2,72	Soproso
29	0,21	1,33	0,27	2,97	1,26	0,32	3,66	Soproso
30	0,12	0,20	0,12	0,99	0,43	0,16	2,50	Soproso
31	2,74	2,11	1,12	3,16	2,47	1,82	1,49	Rugoso
32	0,36	0,38	0,68	0,94	0,40	0,54	1,98	Rugoso
33	0,06	1,91	2,08	0,23	2,14	2,12	2,04	Normal

34	0,13	1,80	0,12	0,28	0,16	0,25	3,06	Rugoso
35	0,14	0,51	0,15	1,97	1,35	0,67	0,68	Normal
36	0,25	0,42	1,11	0,62	0,53	0,59	1,06	Soproso

Tabela B.2. – Valores de *shimmer* em porcentagem (%) obtidos antes e após as transmissões dos sinais reais

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	6,48	8,07	13,28	10,76	8,65	7,90	23,30	Rugoso
2	5,82	7,50	14,11	15,23	9,40	6,44	16,79	Soproso
3	3,66	6,90	5,98	15,71	3,68	11,12	25,49	Soproso
4	6,61	18,37	35,83	15,74	7,90	12,60	34,40	Rugoso
5	4,68	7,81	10,98	18,78	7,95	11,04	13,85	Normal
6	1,90	2,16	5,96	14,21	4,94	5,65	19,45	Soproso
7	2,11	4,25	6,01	6,42	2,76	3,24	12,29	Soproso
8	2,89	5,41	8,09	14,21	5,44	4,67	15,59	Soproso
9	1,92	2,10	5,36	12,55	2,76	5,19	15,25	Soproso
10	2,80	4,31	5,93	6,28	4,47	3,63	16,18	Soproso
11	6,17	6,10	7,45	12,48	5,91	6,49	19,00	Rugoso
12	2,90	3,47	4,16	19,53	5,11	3,68	22,05	Soproso
13	32,76	37,65	34,96	33,89	36,33	36,78	47,30	Rugoso
14	7,78	8,11	20,10	47,19	9,84	20,25	27,09	Soproso
15	28,98	30,13	31,68	34,40	29,06	29,79	36,12	Rugoso
16	16,34	13,27	18,66	18,24	18,70	18,54	31,97	Rugoso
17	11,54	13,24	12,11	20,37	13,20	12,73	15,18	Normal
18	1,14	1,41	11,09	12,77	4,36	3,74	10,69	Normal
19	4,84	6,52	6,92	7,17	4,80	5,54	31,21	Soproso
20	2,27	2,51	10,63	11,40	4,54	5,10	14,02	Soproso
21	2,55	4,01	6,63	10,24	3,04	3,25	17,16	Soproso
22	4,23	5,51	14,75	19,64	5,22	8,13	24,86	Rugoso
23	2,20	2,86	16,10	14,29	6,37	4,27	11,56	Soproso
24	4,11	7,33	8,22	10,81	4,47	8,54	11,74	Normal
25	1,48	2,19	5,34	8,68	2,51	3,22	23,25	Normal
26	1,23	1,56	4,11	16,13	2,51	4,13	17,50	Soproso
27	5,27	5,43	7,18	12,46	6,90	7,25	21,61	Rugoso
28	2,80	5,24	18,65	12,88	8,45	8,26	24,66	Soproso
29	2,23	3,93	13,91	18,93	7,81	6,68	24,81	Soproso
30	2,22	2,83	5,23	13,78	5,99	5,38	23,76	Soproso
31	6,17	5,78	7,95	12,76	7,27	6,19	17,63	Rugoso
32	16,34	13,56	20,70	17,31	17,33	19,34	31,39	Rugoso
33	3,66	9,60	30,71	6,92	10,88	13,00	21,68	Normal
34	4,23	5,50	5,80	8,72	4,61	7,81	24,52	Rugoso
35	2,20	3,45	9,12	14,16	6,21	5,14	12,31	Normal
36	2,80	4,87	20,69	7,76	6,72	8,06	16,71	Soproso

Tabela B.3. – Valores de GNE em Decibel (dB) obtidos antes e após as transmissões dos sinais reais

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	0,86	0,73	0,80	0,48	0,86	0,85	0,82	Rugoso
2	0,62	0,43	0,55	0,38	0,62	0,60	0,75	Soproso
3	0,90	0,79	0,85	0,44	0,90	0,89	0,76	Soproso
4	0,68	0,37	0,39	0,33	0,68	0,65	0,73	Rugoso
5	0,91	0,90	0,84	0,54	0,91	0,91	0,88	Normal
6	0,76	0,60	0,72	0,37	0,77	0,75	0,65	Soproso
7	0,91	0,73	0,89	0,65	0,91	0,90	0,89	Soproso
8	0,67	0,52	0,56	0,44	0,67	0,65	0,73	Soproso
9	0,96	0,89	0,94	0,63	0,96	0,96	0,82	Soproso
10	0,90	0,83	0,88	0,67	0,90	0,89	0,87	Soproso
11	0,97	0,88	0,94	0,70	0,97	0,96	0,85	Rugoso
12	0,78	0,67	0,74	0,36	0,78	0,77	0,75	Soproso
13	0,88	0,82	0,66	0,57	0,89	0,86	0,73	Rugoso
14	0,77	0,65	0,69	0,48	0,78	0,74	0,70	Soproso
15	0,88	0,87	0,76	0,60	0,88	0,87	0,81	Rugoso
16	0,73	0,63	0,58	0,51	0,73	0,71	0,71	Rugoso
17	0,83	0,66	0,74	0,41	0,83	0,82	0,70	Normal
18	0,94	0,86	0,88	0,53	0,94	0,93	0,83	Normal
19	0,37	0,31	0,31	0,33	0,37	0,36	0,71	Soproso
20	0,93	0,78	0,85	0,59	0,93	0,92	0,85	Soproso
21	0,85	0,80	0,80	0,64	0,85	0,85	0,74	Soproso
22	0,87	0,78	0,80	0,52	0,87	0,86	0,81	Rugoso
23	0,95	0,87	0,85	0,52	0,95	0,94	0,90	Soproso
24	0,93	0,86	0,85	0,59	0,94	0,92	0,86	Normal
25	0,98	0,94	0,95	0,75	0,99	0,98	0,89	Normal
26	0,92	0,76	0,87	0,49	0,92	0,90	0,78	Soproso
27	0,92	0,87	0,86	0,62	0,92	0,91	0,81	Rugoso
28	0,65	0,52	0,54	0,35	0,65	0,61	0,73	Soproso
29	0,92	0,86	0,78	0,59	0,92	0,91	0,84	Soproso
30	0,68	0,52	0,51	0,44	0,69	0,67	0,76	Soproso
31	0,97	0,88	0,94	0,71	0,97	0,96	0,85	Rugoso
32	0,73	0,63	0,57	0,52	0,73	0,71	0,72	Rugoso
33	0,90	0,79	0,72	0,43	0,90	0,89	0,78	Normal
34	0,87	0,78	0,85	0,66	0,87	0,86	0,82	Rugoso
35	0,95	0,87	0,91	0,59	0,95	0,94	0,90	Normal
36	0,65	0,53	0,52	0,36	0,66	0,62	0,72	Soproso

Tabela B.4. – Valores de frequência fundamental em Hertz (Hz) obtidos antes e após as transmissões dos sinais reais

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	191,70	191,51	191,51	191,68	191,68	191,68	177,43	Rugoso
2	203,91	204,09	204,01	203,88	204,14	203,97	203,90	Soproso
3	75,19	75,24	75,21	75,22	75,19	75,46	147,08	Soproso
4	72,18	71,56	72,97	72,68	72,60	72,14	72,43	Rugoso
5	114,72	114,20	114,71	114,62	114,73	114,67	114,71	Normal
6	192,77	192,80	192,77	192,77	192,75	192,77	192,82	Soproso
7	177,70	177,79	177,71	177,71	177,70	177,69	177,70	Soproso
8	137,44	137,41	137,43	137,44	137,43	137,43	137,46	Soproso
9	194,34	194,36	194,36	194,43	194,36	194,37	187,90	Soproso
10	261,93	262,01	263,10	262,17	261,98	262,16	222,88	Soproso
11	253,27	254,56	254,78	254,58	254,67	253,23	237,97	Rugoso
12	208,89	208,91	208,91	200,36	208,92	208,89	208,96	Soproso
13	141,64	141,36	149,32	152,05	142,38	143,26	161,00	Rugoso
14	224,17	224,12	223,86	224,98	237,09	209,36	209,61	Soproso
15	115,60	115,67	113,23	119,97	114,28	115,76	115,45	Rugoso
16	103,48	103,55	103,66	103,34	103,75	103,85	103,89	Rugoso
17	205,05	205,16	205,10	205,08	204,48	204,44	203,42	Normal
18	190,33	189,88	189,79	190,24	189,86	189,88	189,86	Normal
19	204,45	204,64	204,49	204,53	204,45	204,50	204,66	Soproso
20	222,09	221,78	222,14	222,00	222,15	221,81	221,78	Soproso
21	178,64	178,71	178,69	178,59	178,62	178,62	177,30	Soproso
22	154,03	154,04	154,01	153,98	154,03	154,08	153,06	Rugoso
23	177,57	177,61	177,64	177,76	177,60	177,56	177,55	Soproso
24	220,06	219,99	220,15	220,04	220,06	220,20	220,10	Normal
25	211,47	211,49	211,46	211,48	211,45	211,50	203,32	Normal
26	257,10	257,10	257,06	229,02	257,20	257,37	257,10	Soproso
27	102,65	102,64	102,72	102,64	102,69	102,65	102,62	Rugoso
28	212,00	212,36	212,17	206,70	212,43	212,48	212,22	Soproso
29	164,48	164,42	164,43	161,08	164,47	164,47	164,43	Soproso
30	194,06	194,11	194,21	194,02	194,20	194,19	194,13	Soproso
31	253,27	254,56	249,74	251,11	253,01	252,94	240,92	Rugoso
32	103,48	103,64	103,59	103,40	103,31	103,55	103,96	Rugoso
33	75,19	75,37	79,43	75,19	75,18	120,28	75,21	Normal
34	154,03	154,05	154,03	154,00	154,01	154,07	152,74	Rugoso
35	177,57	177,59	177,52	177,75	177,58	177,60	177,60	Normal
36	212,00	212,29	212,16	212,00	212,58	212,34	212,03	Soproso

Tabela B.5. – Valores do desvio padrão da frequência fundamental em Hertz (Hz) obtidos antes e após as transmissões dos sinais reais

Número do sinal	Valor original	G.711 Lei A	Speex32	GSM Full Rate	LPCM16	Opus24	SILK16	Classificação inicial
1	1,85	1,90	4,06	1,94	1,84	1,89	33,94	Rugoso
2	2,73	2,65	2,74	2,81	2,71	2,87	2,77	Soproso
3	0,37	0,76	0,38	0,43	0,37	1,92	75,40	Soproso
4	0,97	1,31	1,55	1,57	1,37	1,23	1,59	Rugoso
5	1,20	4,05	1,20	2,26	1,20	1,17	1,20	Normal
6	1,56	1,60	1,59	1,64	1,57	1,65	1,62	Soproso
7	1,04	1,14	1,06	1,06	1,05	1,05	1,08	Soproso
8	1,39	1,37	1,44	1,47	1,41	1,41	1,41	Soproso
9	2,38	2,40	2,37	2,46	2,38	2,39	24,11	Soproso
10	31,17	31,05	27,37	30,69	31,18	30,67	64,12	Soproso
11	15,73	8,86	7,90	8,88	8,84	15,79	43,91	Rugoso
12	3,84	3,88	3,71	28,39	3,94	3,86	3,98	Soproso
13	61,04	61,88	70,62	69,62	60,01	61,91	76,84	Rugoso
14	48,37	48,17	48,58	47,31	23,43	62,44	59,86	Soproso
15	8,00	8,02	12,97	27,72	11,16	8,21	8,09	Rugoso
16	4,90	4,81	4,42	5,11	4,33	4,02	4,26	Rugoso
17	3,54	3,72	3,68	3,58	9,39	9,27	12,98	Normal
18	1,91	2,27	2,33	1,91	2,35	2,33	2,32	Normal
19	2,52	2,72	2,74	2,63	2,51	2,53	2,78	Soproso
20	4,72	2,13	5,18	4,59	4,81	2,12	2,19	Soproso
21	1,50	1,50	2,37	1,46	1,46	1,49	10,83	Soproso
22	1,20	1,19	1,35	2,16	1,20	1,21	8,55	Rugoso
23	1,33	1,32	1,45	1,63	1,27	1,33	1,35	Soproso
24	2,09	2,30	2,32	2,09	2,09	2,24	2,11	Normal
25	1,44	1,46	1,48	1,46	1,45	1,47	27,97	Normal
26	1,27	1,26	1,28	60,29	1,32	1,70	1,32	Soproso
27	4,06	4,07	4,01	4,07	4,01	3,98	4,09	Rugoso
28	1,42	1,89	2,52	26,87	2,11	2,28	1,70	Soproso
29	0,87	0,91	1,05	16,08	0,87	0,87	0,90	Soproso
30	3,12	3,51	3,60	3,47	3,86	3,73	3,40	Soproso
31	15,73	8,86	26,45	23,74	16,41	16,83	40,52	Rugoso
32	4,90	4,69	4,66	5,19	5,29	4,92	4,92	Rugoso
33	0,37	1,34	23,56	0,38	0,37	68,99	0,48	Normal
34	1,20	1,19	1,19	1,55	1,19	1,21	<b>9,87</b>	Rugoso
35	1,33	1,41	1,38	1,47	1,27	1,29	1,37	Normal
36	1,42	1,97	2,25	1,74	2,08	1,93	1,57	Soproso