

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SANTA CATARINA – CÂMPUS FLORIANÓPOLIS
DEPARTAMENTO ACADÊMICO DE ELETRÔNICA
CURSO SUPERIOR DE ENGENHARIA ELETRÔNICA**

VÍTOR FACCIO

**DESENVOLVIMENTO DE *SOFTWARE* PARA CORREÇÃO DE PICOTE
EM ARQUIVOS DE ÁUDIO: processamento digital de sinais em
Python para reconstrução de sinal picotado com ênfase em auxílio
a ferramentas de análise de voz**

Florianópolis, 2023.

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
SANTA CATARINA – CÂMPUS FLORIANÓPOLIS
DEPARTAMENTO ACADÊMICO DE ELETRÔNICA
CURSO SUPERIOR DE ENGENHARIA ELETRÔNICA**

VÍTOR FACCIO

**DESENVOLVIMENTO DE *SOFTWARE* PARA CORREÇÃO DE PICOTE
EM ARQUIVOS DE ÁUDIO: processamento digital de sinais em
Python para reconstrução de sinal picotado com ênfase em auxílio
a ferramentas de análise de voz**

Trabalho de Conclusão de Curso submetido
ao Instituto Federal de Educação, Ciência e
Tecnologia de Santa Catarina como parte
dos requisitos para obtenção do título de
Engenheiro Eletrônico.

Orientador:
Prof. Pedro Giassi Junior, Dr.

Florianópolis, 2023.

Ficha de identificação da obra elaborada pelo autor.

Faccio, Vítor

Desenvolvimento de software para correção de picote em arquivos de áudio: processamento digital de sinais em Python para reconstrução de sinal picotado com ênfase em auxílio a ferramentas de análise de voz / Vítor Faccio; orientação de Pedro Giassi Jr. - Florianópolis, SC, 2023.

Trabalho de Conclusão de Curso (TCC) - Instituto Federal de Santa Catarina, Câmpus Florianópolis. Bacharelado em Engenharia Eletrônica. Departamento Acadêmico de Eletrônica.

Inclui Referências.

1. Processamento digital de sinais. 2. Picote de áudio. 3. Ocultação de perda de pacotes. 4. Biometria de voz. 5. Transcrição de áudio. I. Giassi Jr, Pedro. II. Instituto Federal de Santa Catarina. III. Desenvolvimento de software para correção de picote em arquivos de áudio.

DESENVOLVIMENTO DE SOFTWARE PARA CORREÇÃO DE PICOTE EM ARQUIVOS DE ÁUDIO: processamento digital de sinais em *Python* para reconstrução de sinal picotado com ênfase em auxílio a ferramentas de análise de voz

VÍTOR FACCIO

Este trabalho foi julgado adequado para obtenção do Título de Engenheiro Eletrônico em fevereiro de 2023 e aprovado na sua forma final pela banca examinadora do Curso de Engenharia Eletrônica do Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina.

Florianópolis, 15 de fevereiro de 2023.

Banca examinadora:



Documento assinado digitalmente
PEDRO GIASSI JUNIOR
Data: 22/02/2023 14:56:01-0300
CPF: 219.227.438-60
Verifique as assinaturas em <https://v.ifsc.edu.br>

Pedro Giassi Junior, Dr.



Documento assinado digitalmente
ALEXANDRE LEIZOR SZCZUPAK
Data: 24/02/2023 17:23:57-0300
Verifique em <https://verificador.iti.br>

Alexandre Leizor Szczupak, Dr.



Documento assinado digitalmente
CARLOS GONTARSKI SPERANZA
Data: 24/02/2023 12:52:41-0300
Verifique em <https://verificador.iti.br>

Carlos Gontarski Speranza, Dr.

AGRADECIMENTOS

Agradeço a todas as pessoas que contribuíram com a elaboração deste trabalho e de minha formação como engenheiro eletrônico.

Em especial, agradeço ao professor Pedro Giassi Jr. pela orientação dedicada e cuidadosa com meu trabalho.

Ao Instituto Federal de Santa Catarina, particularmente ao Câmpus Florianópolis, instituição que me acolhe desde o Curso Técnico em Eletrônica.

Aos professores da Graduação em Engenharia Eletrônica, pela participação determinante em minha formação. À professora Cláudia Regina Silveira, pela leitura atenciosa deste TCC.

A todos os servidores do Câmpus Florianópolis, por contribuírem com minha formação e trajetória acadêmica, especialmente aos servidores do Departamento Acadêmico de Eletrônica.

Pela disponibilidade de comporem a banca examinadora deste TCC, aos professores Alexandre Leizor e Carlos Speranza. Obrigado por suas contribuições que qualificam meu trabalho.

Aos colegas de curso, com quem vivi vivenciei momentos importantes para minha evolução pessoal e acadêmica.

À empresa Khomp, que me concedeu campo de estágio, a partir do qual realizei a pesquisa.

Aos profissionais de saúde que me acompanharam durante toda a graduação, particularmente à dra. Concetta e as fisioterapeutas Catarina e Luciana.

À minha namorada Luana, pelo apoio extenso em todos os momentos, além do carinho e paciência durante os últimos meses.

Aos meus pais Roberto e Juçara, por me apoiarem e se esforçarem para que eu pudesse completar essa etapa da minha vida, além de servirem como meus grandes exemplos.

A todos estes e aos que não foram citados, mas que contribuíram de alguma forma com minha pesquisa e com este TCC, meu muito obrigado!

*"Você se apoiou nos ombros de gênios
para realizar algo o mais rápido possível e,
antes mesmo de saber o que tinha, você
patenteou, embalou e agora está vendendo.*

Bem...

*Seus cientistas estavam tão preocupados se poderiam
que não pararam para pensar se deveriam."*

- Michael Crichton

RESUMO

A perda de dados nas comunicações por Voz sobre Protocolo de Internet (*Voice over Internet Protocol* – VoIP) é um problema recorrente nos meios de telefonia modernos que causa deformidades no áudio recebido por um aparelho. A VoIP é uma forma de transmissão de voz dividida em pacotes, portanto o não recebimento de uma parcela de sinal pode acarretar pequenos intervalos de silêncio, chamados de picotes. Nesse contexto surgem técnicas de ocultação de perda de pacotes (*packet loss concealment* – PLC) que buscam reconstruir segmentos de áudio com base em partes adjacentes recebidas corretamente. Neste trabalho aborda-se a implementação de diferentes técnicas de reconstrução de áudio para correção de arquivos com características de perda de pacotes, como os provenientes de comunicação VoIP, com foco em auxiliar ferramentas de análise de voz a obterem melhores resultados. As ferramentas de análise visadas são módulos de reconhecimento de identidade por biometria de voz e de transcrição de áudio. Explora-se métodos de reconstrução de sinal em obras e documentações já concebidas, como o algoritmo de PLC apresentado no Anexo A do *codec* G.711. Implementa-se em *Python* funções de replicação de período de tom (*pitch waveform replication* – PWR), soma sobreposta (*overlap adding* – OLA) e codificação por predição linear (*Linear Predictive Coding* – LPC) em sentidos frontal (*forward*), inverso (*backward*) e bilateral. Obtém-se resultados promissores com a aplicação de todos os algoritmos nos testes com biometria de voz, especialmente a estratégia baseada em PWR. Para os testes com ferramenta de transcrição de áudio, de outro modo, obtém-se resultados negativos com as técnicas baseadas em LPC, sendo o algoritmo baseado em PWR o único com resultados que favorecem a ferramenta. Os códigos implementados podem ser úteis no tratamento de sinal em sistemas que façam gravação de chamadas por VoIP e analisem seu conteúdo para fins diversos, havendo domínio das ferramentas utilizadas e noção de como os algoritmos se demonstram eficazes.

Palavras-chave: Processamento digital de sinais. Picote de áudio. Ocultação de perda de pacotes. Biometria de voz. Transcrição de áudio.

ABSTRACT

Data loss in Voice over Internet Protocol (VoIP) is a recurring problem in modern telecommunication, which causes distortions in the audio received by equipment. VoIP is a transmission form based in divided packets, thus not receiving a segment of signal may result in small silence intervals, called chops. In this context arise packet loss concealment (PLC) techniques, which seek to reconstruct audio segments based on correctly received adjacent parts. This work discusses the implementation of different audio reconstruction techniques to correct files with packet loss characteristics, such as those originating from VoIP, with a focus on helping voice analysis tools to obtain better results. The targeted analysis tools are identity recognition by voice biometrics and audio transcription modules. Signal reconstruction methods are explored in previous works and documentation, such as the PLC algorithm presented in Annex A of codec G.711. Code is implemented in Python with functions of pitch waveform replication (PWR), overlap adding (OLA) and linear prediction coding (LPC) in forward, backward and bilateral directions. All algorithms have positive results in voice biometrics tests, especially the PWR-based strategy. For tests with the audio transcription tool, on the other hand, negative results are obtained with techniques based on LPC, with the PWR-based algorithm being the only one with results favoring the tool. The implemented codes may be useful in signal processing in systems that record VoIP calls and analyze their content for different purposes, being aware of the tools used and how the algorithms prove to be effective.

Keywords: Digital signal processing. Audio chopping. Packet loss concealment. Voice biometrics. Speech recognition.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Exposição de sinal de voz em função de tempo contínuo e tempo discreto	23
Figura 2 –	Representação em domínio da frequência de sinal original $x_c(t)$ com banda limitada em Ω_N	24
Figura 3 –	Representação em domínio da frequência de sinal discreto formado a partir de amostragem de $x_c(t)$	25
Figura 4 –	Representação em domínio da frequência de sinal discreto formado a partir de amostragem de $x_c(t)$ com frequência de amostragem inferior a limite de banda do sinal	25
Figura 5 –	Exemplos de sinal das vogais /a/, /i/ e /u/ no domínio do tempo e da frequência	26
Figura 6 –	Exemplo de pacote UDP	28
Figura 7 –	Exemplo de pacote RTP	29
Figura 8 –	Visualização em programa Audacity 3.1.3 de ondas provenientes de mesma gravação de voz, com sinal original no gráfico superior e acometido por picotes no inferior	31
Figura 9 –	Exemplo de mesmo sinal sem e com picotes, detalhado	32
Figura 10 –	Representação de sistema de reconhecimento de voz	34
Figura 11 –	Processo inicial de obtenção dos resultados para áudios de referência e com picote não corrigidos	39
Figura 12 –	Processo de obtenção de resultados para ferramenta de biometria de voz	40
Figura 13 –	Processo de obtenção de resultados para ferramenta de transcrição de áudio	40
Figura 14 –	Geração de <i>datasets</i> de áudios corrigidos	41
Figura 15 –	Sinal original utilizado na demonstração dos métodos de PLC	43
Figura 16 –	Magnitude da DFT do sinal original utilizado na demonstração dos métodos de PLC	43
Figura 17 –	Sinal picotado utilizado na demonstração dos métodos de PLC	44
Figura 18 –	Correção de picote por repetição de frame anterior completo	45
Figura 19 –	Exemplo de replicação de tom do sinal	47
Figura 20 –	Detalhe da região de detecção de período de pitch anterior a um picote	50

Figura 21 –	Ordem de replicação das amostras do período de pitch anterior a um picote no espaço de seu tempo	51
Figura 22 –	Sinais 1 e 2 utilizados na demonstração de OLA	54
Figura 23 –	Aplicação de OLA simples sobre sinais distintos	55
Figura 24 –	Janelas comumente usadas em processamento digital de sinais	56
Figura 25 –	Janelas crescentes triangular e Hanning prontas para aplicação nos sinais operados por OLA	57
Figura 26 –	Aplicação de janelas triangular e Hanning sobre segmentos interseccionantes dos sinais 1 e 2	58
Figura 27 –	Operação de OLA sobre sinais com janelamento triangular e Hanning	59
Figura 28 –	Transições entre primeiro frame do sinal 1 e segundo frame do sinal 2 com segmentos calculados por OLA com janelamento triangular e Hanning	59
Figura 29 –	Sinais com PLC por método de PWR puro e por PWR com processo de OLA nas fronteiras anterior e posterior ao picote	61
Figura 30 –	Reconstrução de frame perdido de sinal vocalizado com método de LPC discutido	65
Figura 31 –	Reconstrução de frame perdido de sinal não-vocalizado com método de LPC discutido	66
Figura 32 –	Reconstrução de frame perdido de sinal vocalizado com método de LPC backward	68
Figura 33 –	Comparação entre frames sintetizados por meio de LPC <i>forward</i> , <i>backward</i> e bilateral	69
Figura 34 –	Reconstrução de frame perdido de sinal vocalizado com método de LPC bilateral	70
Figura 35 –	Trechos de áudios buscados em versão original e com presença de picotes	73
Figura 36 –	Conteúdo de cada arquivo de áudio do dataset	76
Figura 37 –	Exemplo de forma de onda completa em arquivo de áudio utilizado	75
Figura 38 –	Trecho de áudios com picotes, de referência e corrigido por algoritmo I	83

Figura 39 – Detalhe de replicação de período de pitch de sinal de voz não-vocalizado com PWR e transição entre frames com OLA janelada	84
Figura 40 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC II	88
Figura 41 – Formas de onda completa de arquivos de áudio picotado, de referência e corrigido por algoritmo de PLC II com destaque para artefatos sonoros	89
Figura 42 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC II	89
Figura 43 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC III	92
Figura 44 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC III	93
Figura 45 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC IV	96
Figura 46 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC IV	97
Figura 47 – Representação gráfica de resultados obtidos em testes de algoritmos com ferramenta de biometria de voz	101
Figura 48 – Representação gráfica de resultados obtidos em testes de algoritmos com ferramenta de transcrição de áudio	102

LISTA DE ABREVIATURAS E SIGLAS

ACK	<i>Acknowledgement</i> (Resposta de confirmação)
AMDF	<i>Average Magnitude Difference Function</i> (Função de Diferença de Magnitude Média)
API	<i>Application Programming Interface</i> (Interface de Programação de Aplicação)
CDT	<i>Cumulative Distribution Transform</i> (Transformada de Distribuição Cumulativa)
CER	<i>Character Error Rate</i> (Taxa de Erro de Caracteres)
DFT	<i>Discrete Fourier Transform</i> (Transformada Discreta de Fourier)
DSP	<i>Digital Signal Processing</i> (Processamento Digital de Sinais)
FEC	<i>Forward Error Correction</i> (Correção Adiantada de Erro)
FFT	<i>Fast Fourier Transform</i> (Transformada Rápida de Fourier)
IFSC	Instituto Federal de Santa Catarina
IP	<i>Internet Protocol</i> (Protocolo de Internet)
LPC	<i>Linear Predictive Coding</i> (Codificação por Predição Linear)
MFCC	<i>Mel-Frequency Cepstral Coefficients</i> (Coeficientes Mel-Cepstrais)
MOS	<i>Mean Opinion Score</i> (Pontuação Média de Opinião)
OLA	<i>Overlap Adding</i> (Soma Sobreposta)
PCM	<i>Pulse Code Modulation</i> (Modulação por Código de Pulsos)
PLC	<i>Packet Loss Concealment</i> (Ocultação de Perda de Pacotes)
PSTN	<i>Public Switched Telephone Network</i> (Rede Pública de Telefonia Comutada)
PWR	<i>Pitch Waveform Replication</i> (Replicação de Tom do Sinal)
QoS	<i>Quality of Service</i> (Qualidade de Serviço)
RTP	<i>Real-time Transport Protocol</i> (Protocolo de Transporte em Tempo Real)
UDP	<i>User Datagram Protocol</i> (protocolo de datagrama do usuário)
VoIP	<i>Voice over Internet Protocol</i> (Voz sobre Protocolo de Internet)
WER	<i>Word Error Rate</i> (Taxa de Erro de Palavras)

LISTA DE QUADROS

Quadro 1 –	Locutores envolvidos com a gravação do dataset utilizado	75
Quadro 2 –	Faixa de valores utilizados na geração de picotes do dataset	76
Quadro 3 –	Resultados com ferramenta de biometria de voz para dataset de referência	78
Quadro 4 –	Resultados com ferramenta de biometria de voz para dataset com picotes	80
Quadro 5 –	Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets de referência e acometido por picotes	81
Quadro 6 –	Valores de CER para comparação entre transcrições de áudio dos datasets de referência e acometido por picotes	82
Quadro 7 –	Resultados com ferramenta de biometria de voz para dataset corrigido pelo algoritmo I	85
Quadro 8 –	Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets de referência, acometido por picotes e corrigido por algoritmo de PLC I	86
Quadro 9 –	Valores de CER para comparação entre transcrições de áudio dos datasets de referência e com picotes corrigidos por algoritmo de PLC I	86
Quadro 10 –	Comparação entre resultados entre médias de CER para datasets com picote e corrigido por algoritmo de PLC I	87
Quadro 11 –	Valores de variáveis em implementação de PLC II com LPC em <i>Python</i>	87
Quadro 12 –	Resultados com ferramenta de biometria de voz para dataset corrigido pelo algoritmo II	90
Quadro 13 –	Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets de referência, acometido por picotes e corrigido por algoritmo de PLC II	91
Quadro 14 –	Valores de CER para comparação entre transcrições de áudio dos datasets de referência e com picotes corrigidos por algoritmo de PLC II	91
Quadro 15 –	Comparação entre resultados entre médias de CER para datasets com picote e corrigido por algoritmo de PLC II	92

Quadro 16 – Resultados com ferramenta de biometria de voz para dataset corrigido pelo algoritmo III	94
Quadro 17 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets de referência, acometido por picotes e corrigido por algoritmo de PLC III	95
Quadro 18 – Valores de CER para comparação entre transcrições de áudio dos datasets de referência e com picotes corrigidos por algoritmo de PLC III	95
Quadro 19 – Comparação entre resultados entre médias de CER para datasets com picote e corrigido por algoritmo de PLC III	96
Quadro 20 – Resultados com ferramenta de biometria de voz para dataset corrigido pelo algoritmo III	98
Quadro 21 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets de referência, acometido por picotes e corrigido por algoritmo de PLC IV	99
Quadro 22 – Valores de CER para comparação entre transcrições de áudio dos datasets de referência e com picotes corrigidos por algoritmo de PLC IV	99
Quadro 23 – Comparação entre resultados entre médias de CER para datasets com picote e corrigido por algoritmo de PLC IV	100
Quadro 24 – Comparação entre variações de médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre datasets corrigidos por algoritmos de PLC I, II, III e IV e dataset acometido por picotes	100
Quadro 25 – Comparação entre variações de médias de CER para <i>datasets</i> corrigidos por algoritmos de PLC I, II, III e IV e <i>dataset</i> picotado . . .	101

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Justificativa	18
1.2	Definição do problema	19
1.3	Objetivo geral	19
1.4	Objetivos específicos	20
1.5	Estrutura do trabalho	20
2	REFERENCIAL TEÓRICO	22
2.1	Sinais digitais	22
2.2	Frequências da voz humana	26
2.3	Perda de pacotes em comunicação VoIP e picote em sinais de áudio	27
2.4	Ferramentas de análise de voz	32
2.4.1	Biometria e reconhecimento de voz	32
2.4.2	Transcrição de áudio	34
2.5	Ocultação de perda de pacotes	35
3	METODOLOGIA	37
3.1	Métodos aplicados	38
4	APRESENTAÇÃO DOS RESULTADOS	42
4.1	Técnicas de reconstrução de áudio	42
4.1.1	<i>Pitch Waveform Replication – PWR</i>	44
4.1.1.1	<i>Detecção de período de pitch</i>	47
4.1.1.2	<i>Replicação do sinal</i>	50
4.1.2	<i>Overlap Add – OLA</i>	52
4.1.3	<i>Linear Prediction Coding – LPC</i>	61
4.1.4	LPC com sentido inverso	66
4.1.5	LPC bilateral	68
4.2	Ferramentas de análise de áudio	70
4.3	Conjunto de arquivos de áudio	72
4.3.1	Descrição dos arquivos componentes do <i>dataset</i>	74
4.3.2	Adaptação de frequência de amostragem	76
4.3.3	Insuficiência dos arquivos acometidos por picote e geração de <i>dataset</i> adequado	77
4.4	Testes com arquivos de referência	78
4.5	Testes com arquivos acometidos por picotes	80
4.6	Testes com arquivos corrigidos	82
4.6.1	Algoritmo I – <i>Pitch Wave Replication</i> baseado no Apêndice I do codec ITU-T G.711 (1999)	83
4.6.2	Algoritmo II – <i>Linear Prediction Coding</i> baseado em ElSabrouty, Bouchard e Aboulnasr (2004)	87

4.6.3	Algoritmo III – <i>Linear Prediction Coding</i> com sentido inverso	92
4.6.4	Algoritmo IV – <i>Linear Prediction Coding</i> bilateral	96
4.7	Resultados e discussões	100
	REFERÊNCIAS	103
	APÊNDICES	107
	APÊNDICE A – Frases transcritas a partir dos áudios de referência	107
	APÊNDICE B – Frases transcritas a partir dos áudios acometidos por picotes	111
	APÊNDICE C – Frases transcritas a partir dos áudios corrigidos pelo algoritmo I	115
	APÊNDICE D – Frases transcritas a partir dos áudios corrigidos pelo algoritmo II	119
	APÊNDICE E – Frases transcritas a partir dos áudios corrigidos pelo algoritmo III	122
	APÊNDICE F – Frases transcritas a partir dos áudios corrigidos pelo algoritmo IV	126

1 INTRODUÇÃO

A indústria de comunicação de voz por meio da *internet* está em constante expansão, tanto em questão de faturamento quanto em número de usuários. Esse setor é composto por empresas de telecomunicações e telefonia, como a companhia *Cisco Systems Inc.* (CISCO SYSTEMS INC., 2017) e, exemplificando nacionalmente, a Khomp Indústria e Comércio Ltda. (KHOMP INDÚSTRIA E COMÉRCIO LTDA., 2022), que oferecem equipamentos e soluções dedicados. Também são populares alguns serviços informáticos desse meio, como Microsoft Skype, Google Hangouts e Apple FaceTime (HARTE; GILLEN; HINES, 2015). De acordo com o Global Market Insights (2021 *apud* HATFIELD, 2022), o mercado mundial desse setor valia mais de US\$ 30 bilhões em 2020 e existe a estimativa de haver uma taxa de crescimento anual de 15% entre 2021 e 2027, chegando a cerca de US\$ 95 bilhões ao final do período. O *site* VoIP do Brasil (2017) afirma que o número de assinantes de serviços de comunicação de voz por *internet* cresce cerca de 14% ao ano desde 2009.

A forma de comunicação em questão é chamada de "voz sobre protocolo de *internet*" (*Voice over Internet Protocol* – VoIP). Trata-se de um grupo de tecnologias capaz de transmitir voz e mídias por meio de uma rede de *internet*, o que se vivencia como a possibilidade de executar chamadas de telefone a partir de uma conexão de *internet* banda larga (SHAW; SHARMA, 2016, p. 16). Por utilizar a mesma infraestrutura de rede em que os usuários navegam na *internet*, a VoIP se opõe à necessidade de haver uma linha exclusiva de operadora de telefonia comum, o que simplifica sistemas de grandes empresas e oferece redução de custos (KARAPANTAZIS; PAVLIDOU, 2009). Harte, Gillen e Hines (2015) salientam a vantagem de tal estrutura em oferecer melhores algoritmos e mais banda para chamadas com melhor qualidade em relação ao uso da Rede Pública de Telefonia Comutada (*Public Switched Telephone Network* – PSTN).

Apesar de ser uma forma de comunicação vantajosa, a tecnologia VoIP está sujeita a todas as complicações envolvidas com a dependência da *internet*. De acordo com Karapantazis e Pavlidou (2009), por se tratar de um sistema cuja informação trafega em pequenos pacotes, problemas na rede podem causar perda de dados – nesse contexto, os dados se referem a pedaços da voz transmitida. A perda do pacote pode ocorrer por diversos motivos, como atraso, devido ao fato de a

rede estar sobrecarregada, ou de esse mesmo pacote ser descartado quando atinge seu destino. Segundo os autores, a perda de quantidades significativas de pacote durante uma chamada de voz é sentida pelo ouvinte na forma de picotes, ou seja, cortes com duração muito pequena.

O picote é um dos três principais causadores de perda de qualidade de serviço (*Quality of Service – QoS*) em um sistema VoIP, junto ao atraso ponta-a-ponta e *jitter* (KARAPANTASIS; PAVLIDOU, 2009). Essas duas outras grandezas se caracterizam, respectivamente, pelo intervalo entre um usuário pronunciar um som e o outro ouvi-lo e pela variação deste primeiro ao longo de pacotes consecutivos. Todos eles afetam diretamente a experiência do usuário que se comunica (HARTPENCE, 2013), mas o picote é o problema em foco neste trabalho.

O reconhecimento de padrões em sinais de fala é a base para diversas ferramentas como a transcrição de áudio ou a identificação de uma pessoa que está falando. O primeiro exemplo é descrito por Bakri *et al.* (2015) como o processo que converte um sinal sonoro de fala em um conjunto de palavras ou frases. O segundo se trata de uma forma de biometria, o que consiste em métodos para se reconhecer uma pessoa com base em suas características fisiológicas ou comportamentais únicas (SALES; SALES; SILVA, 2021). A voz pode se distinguir em função de características como sotaque, pronúncia, articulação, nasalidade, tom, volume e velocidade da fala. Além disso, tais parâmetros podem ser percebidos de formas diferentes se houver alguma distorção do meio em que trafegam, como, por exemplo, uma chamada telefônica.

Conforme Silva *et al.* (2019), a biometria de voz tem ganhado espaço em aplicações modernas com a finalidade de autenticação de pessoas e controle de acesso, ou seja, uma ferramenta de segurança que age como uma senha. Seu uso é simples e não requer ferramentas de captura especializadas, bastando haver um microfone ligado ao aparelho do usuário. Em um contexto em que os dispositivos digitais se apresentam cada vez mais móveis e dependentes de comunicação externa, é benéfico utilizar-se de um fator de autenticação que acompanhe o consumidor do produto.

A biometria de voz apresenta uma solução ao reconhecimento de pessoas, mas, como citado anteriormente, o meio de transporte é capaz de interferir no áudio recebido pelo aparelho que realiza a checagem (SALES; SALES; SILVA, 2021). Se essa voz transita por meio de VoIP, é possível que ocorram picotes no

sinal e que essas imperfeições afetem o funcionamento da biometria. Referindo-se à qualidade de serviço notada pelo ouvinte, Karapantazis e Pavlidou (2009) afirmam que a qualidade de voz em uma chamada VoIP é aceitável se as perdas de pacotes não ultrapassarem 2%.

A escolha do picote em arquivos de áudio como foco deste trabalho foi tomada com base em relatos de funcionários da equipe de Soluções/Cloud APIs da empresa Khomp Indústria e Comércio Ltda., na qual o autor participou de programa de estágio durante os anos de 2021 e 2022, tendo contato com diversos temas relacionados a redes e telefonia VoIP. Os outros seis funcionários foram questionados verbalmente durante uma reunião a respeito dos fatores mais problemáticos no uso de ferramenta de biometria de voz. O picote de áudio foi referenciado com unanimidade. Embora o incômodo que o picote causa durante uma ligação não seja tão perceptível ao ouvido humano quanto um atraso ponta-a-ponta, que atrapalha o andamento de uma conversa (KARAPANTAZIS; PAVLIDOU, 2009), aborda-se neste texto o efeito do picote em ferramentas digitais de análise de voz, como o reconhecimento de pessoa falante e de conteúdo falado.

Este trabalho investiga os efeitos que a inserção de picotes em gravações de voz causa em ferramentas como a biometria de voz, com base em uma série de áudios confeccionados a partir de gravações de diferentes locutores para simular um equipamento sobrecarregado em uma rede. Foram utilizados para a avaliação desses arquivos dois *softwares* de análise de voz: o *software* de biometria de voz da empresa Khomp Indústria e Comércio Ltda., cedido gentilmente para o desenvolvimento deste trabalho, e a biblioteca *Python SpeechRecognition*, uma ferramenta gratuita de transcrição de áudio para texto com acesso a diversos métodos de conversão. Além de buscar entender a natureza desses efeitos, implementa-se em *Python* algoritmos para correção dos picotes a fim de se obter melhores resultados nas ferramentas utilizadas, no sentido de aproximar-se do que ocorre com áudios sem a simulação forçada de perda de pacotes.

1.1 Justificativa

O contexto de ampliação da comunicação por VoIP em equipamentos modernos dá luz à fragilidade de sistemas variados em relação a distorções no sinal de áudio recebido. Ferramentas de análise de voz em aparelhos ligados à internet,

como a biometria ou a transcrição de áudio, são vulneráveis a fatores externos ao computador, que afetam diretamente a eficiência deles (SALES; SALES; SILVA, 2021).

De acordo com Lecomte *et al.* (2015), deve fazer parte do projeto de codificadores e decodificadores (chamados de *codecs*) a robustez contra erros de rede. Esses equipamentos são responsáveis, respectivamente, por transformar um sinal analógico de áudio em uma série de amostras com valores binários e trazer de volta a sequência de amostras a uma forma de onda analógica (HARTPENCE, 2013). A decodificação de um sinal acometido por picotes forma um áudio errôneo se comparado ao gerado na fonte.

Este trabalho se justifica pela necessidade de haver em um sistema de biometria de voz ou de transcrição de áudio uma certa robustez, mesmo na presença de perda de pacotes. No caso de um projeto já implementado que utiliza um *codec* comercial, é viável a inclusão de uma etapa de processamento de sinal independente deste equipamento para buscar um nível de qualidade de áudio aceitável.

1.2 Definição do problema

Como enunciado anteriormente, um projeto que lida com o processamento de voz como uma ferramenta de biometria de voz ou transcrição de áudio é suscetível a problemas de rede se o tráfego do áudio ocorrer por meio de VoIP. Sendo essa forma de comunicação um mercado em expansão, é razoável que se busque uma forma de compensar erros como a perda de pacotes que seja inserível em um sistema digital.

No sentido de fortalecer aplicações envolvidas com a análise de gravações de voz, tornando seu funcionamento mais resistente à perda de pacotes na comunicação VoIP, este trabalho questiona: qual o desempenho dos principais algoritmos para correção de picotes aplicados a ferramentas de análise de voz?

1.3 Objetivo geral

Este trabalho tem como meta analisar algoritmos de correção de picotes em arquivos de áudio como aqueles provenientes de sinais de VoIP, no contexto do

reconhecimento de voz e transcrição de áudio, examinando o impacto em *softwares* de análise de voz.

1.4 Objetivos específicos

Tendo em vista o panorama almejado, definem-se os objetivos específicos como os seguintes:

- a) revisar em bibliografia os principais métodos de correção de picote disponíveis;
- b) definir uma ferramenta de biometria de voz e uma de transcrição de áudio para uso sobre os arquivos selecionados;
- c) construir uma base de áudios apropriada contendo tanto arquivos de referência quanto acometidos por picote, verificando sua validade no escopo deste projeto e obtendo resultados do uso das ferramentas para cada conjunto;
- d) implementar algoritmos computacionais de correção de picote baseados nos métodos pesquisados, de maneira a reconstruir os arquivos de áudio danificados, utilizando linguagem Python e bibliotecas de acesso público;
- e) aplicar os arquivos corrigidos em ferramentas de biometria de voz e transcrição de áudio;
- f) comparar os três grupos de resultados obtidos e observar o efeito da correção de picote sobre a aproximação aos valores de referência, analisando o melhor modelo.

1.5 Estrutura do trabalho

Este trabalho é centrado na análise do desempenho de métodos de correção de picotes em sinais de áudio quando aplicados a testes executados a partir de ferramentas de biometria de voz e transcrição de áudio.

O primeiro capítulo apresenta as informações gerais a respeito do tema tratado, justificativa e definição de problema, além da explicitação do objetivo geral e todos os específicos.

Em seguida é feita uma exploração mais profunda acerca do tema do trabalho e sua origem, fornecendo um estudo sobre o referencial teórico. Discute-se o gênero de sinais em tempo discreto, abordando a questão de frequência de

amostragem e teorema de amostragem de Nyquist-Shannon. Trata-se também da natureza da voz humana e sua composição frequencial, além da tecnologia VoIP e o problema de picote de áudio. Em seguida são vistos os fundamentos das ferramentas de análise de voz utilizadas durante os testes, a biometria de voz e a transcrição de áudio. Ao final do capítulo discorre-se sobre métodos de ocultação de perda de pacotes.

No capítulo 3 é tratada a metodologia do trabalho, expondo inicialmente as classificações do tipo de pesquisa executado, informando os caracteres de objetivos e abordagem. Após isso, descreve-se os passos tomados para a realização do trabalho.

Por fim, o capítulo 4 percorre todas as etapas práticas executadas, além de apresentar os algoritmos de correção de picote pesquisados, suas bases matemáticas e resultados em formas de onda. São tratados os esforços para estabelecer um conjunto de áudios adequado, explicita-se as características das ferramentas de análise de voz utilizadas e seus métodos de teste e apresenta-se todos os resultados numéricos decorrentes dos diversos testes propostos. Por fim, discute-se os resultados obtidos.

2 REFERENCIAL TEÓRICO

Realiza-se, neste capítulo, um exame sobre os temas pertinentes ao projeto. É proveitoso contextualizar a fonte do problema combatido e entender o sistema em que essa ocorre, para que a solução encontrada seja mais facilmente implementável. Percorre-se, também, a fundamentação teórica dos processos realizados, desde os conceitos de processamento digital de sinais (*digital signal processing* – DSP) até os algoritmos implementados e testados para a realização do projeto.

Este capítulo apresenta, inicialmente, uma análise sobre sinais digitais, embasando sinais de áudio no domínio do tempo, frequência de amostragem, domínio da frequência e Transformada Discreta de Fourier (*Discrete Fourier Transform* – DFT), frequências da voz humana e a mudança de frequência de amostragem (*resampling*). Em seguida, aborda-se o conceito do picote de áudio, contextualizando a comunicação VoIP, a decodificação de sinal e a perda de pacotes, além de salientar formas básicas de decodificadores lidarem com esse problema. Posteriormente, são descritas as ferramentas de análise de voz, a biometria de voz e a transcrição de áudio, observando seus princípios de funcionamento. Por fim, são expostas técnicas de reconstrução de sinal para correção de picote, além de aplicações em algoritmos conhecidos e suas limitações.

2.1 Sinais digitais

De acordo com Oppenheim e Schafer (2010), “sinal” é o termo utilizado para nomear algo que carrega informação, geralmente sobre o estado ou comportamento de um sistema físico. São representados matematicamente como funções de uma ou mais variáveis independentes. Sinais de voz, por exemplo, são representados como funções do tempo. Conforme os autores, a variável independente pode ser contínua ou discreta.

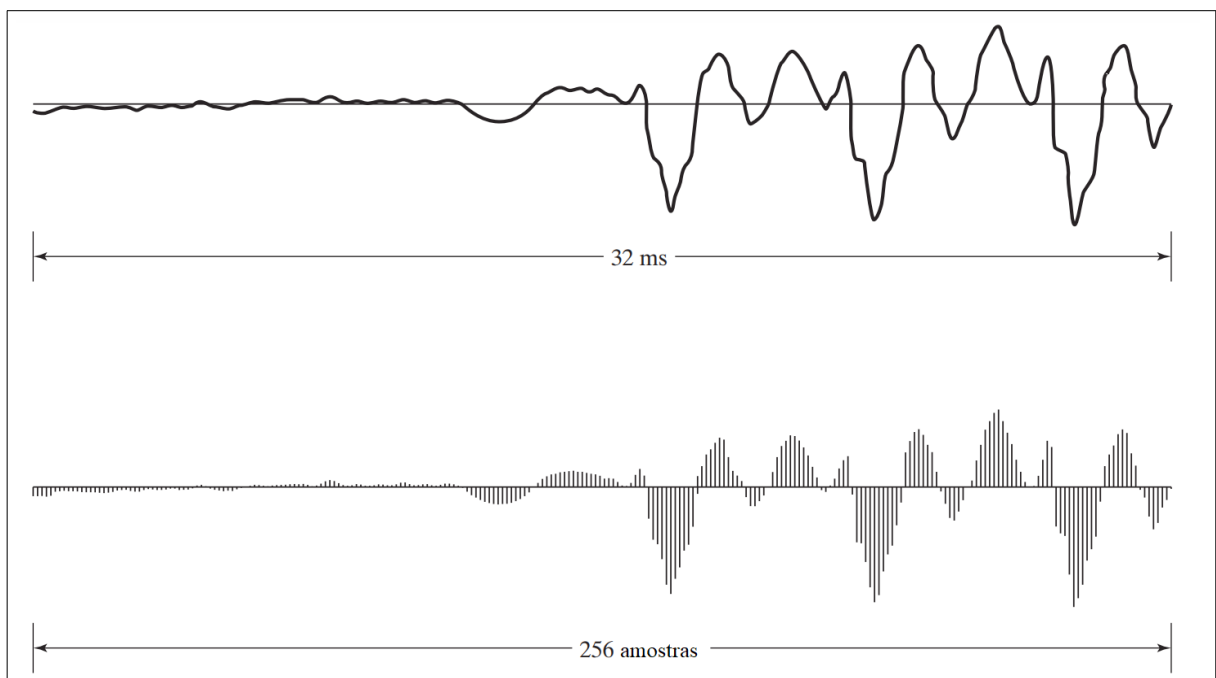
Abordando o caso pertinente a este trabalho, “sinais em tempo discreto são representados matematicamente como uma sequência de números.” (OPPENHEIM; SCHAFER, 2010, p. 10, tradução nossa). A voz pode ser entendida originalmente como a variação da pressão acústica em função do tempo, fenômeno que ocorre de maneira contínua. Em contextos práticos, um sinal contínuo no tempo

pode basear um sinal discreto por meio de amostragens periódicas. Oppenheim e Schafer (2010) descrevem um sinal discreto x derivado de um sinal analógico x_A conforme a Equação 1, em que a variável n é um número inteiro e T representa o período de amostragem.

$$x[n] = x_A(n \cdot T), \quad -\infty < n < \infty \quad (1)$$

A Figura 1 ilustra um sinal acústico de voz em sua forma em tempo contínuo, no gráfico superior, e o mesmo sinal após um processo de amostragem no gráfico inferior. Conforme as escalas horizontais, o sinal tem duração de 32ms e, após amostrado, contém 256 amostras. O período de amostragem T informado por Oppenheim e Schafer (2010) é de 125μs.

FIGURA 1 – Exposição de sinal de voz em função de tempo contínuo e tempo discreto



Fonte: Adaptado de Oppenheim e Schafer (2010, p. 10).

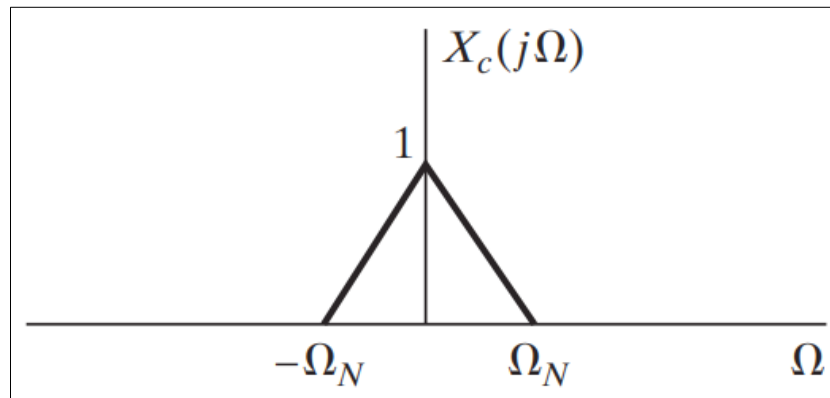
O valor inverso do período de amostragem é chamado de frequência de amostragem (OPPENHEIM; SCHAFER, 2010), ou *sample rate*, representado por f_s ; esse valor indica em *hertz* quantas amostras são obtidas do sinal em um segundo. A mesma grandeza pode ser expressa em radianos por segundo pela frequência

angular de amostragem, ou Ω_s , conforme a Equação 2 (OPPENHEIM; SCHAFER, 2010, p. 154).

$$\Omega_s = \frac{2\pi}{T} \quad (2)$$

O estudo da amostragem no domínio da frequência se beneficia do uso de frequência angular de amostragem. Oppenheim e Schafer (2010) descrevem um exemplo de transformada de Fourier de um sinal amostrado. Considerando um sinal em tempo contínuo $x_c(t)$, sua representação no domínio da frequência pode ser tomada como sua transformada de Fourier $X_c(j\Omega)$. Os autores ilustram o sinal com limite de banda em Ω_N , da forma vista na Figura 2.

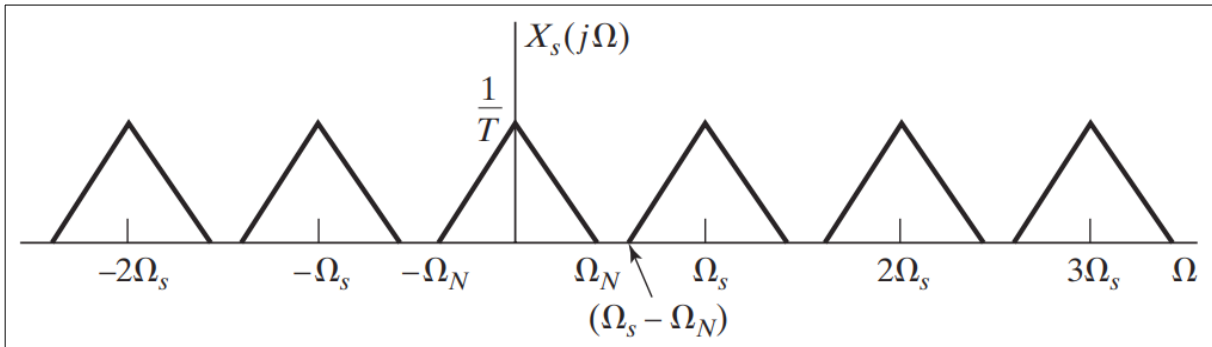
FIGURA 2 – Representação em domínio da frequência de sinal original $x_c(t)$ com banda limitada em Ω_N



Fonte: Oppenheim e Schafer (2010, p. 157).

Considera-se então $X_s(j\Omega)$ a transformada de Fourier de um sinal em tempo discreto, formado a partir da amostragem de $x_c(t)$ e com frequência angular de amostragem igual a Ω_s . Segundo Oppenheim e Schafer (2010), essa transformada consiste em cópias de X_c repetidas periodicamente no eixo Ω , distanciadas de múltiplos de Ω_s . Este fato é ilustrado na Figura 3. Destaca-se na parte inferior do gráfico o segmento $(\Omega_s - \Omega_N)$, que representa a distância entre os pontos do eixo em que a banda positiva da cópia central e a banda negativa da primeira cópia à direita atingem zero.

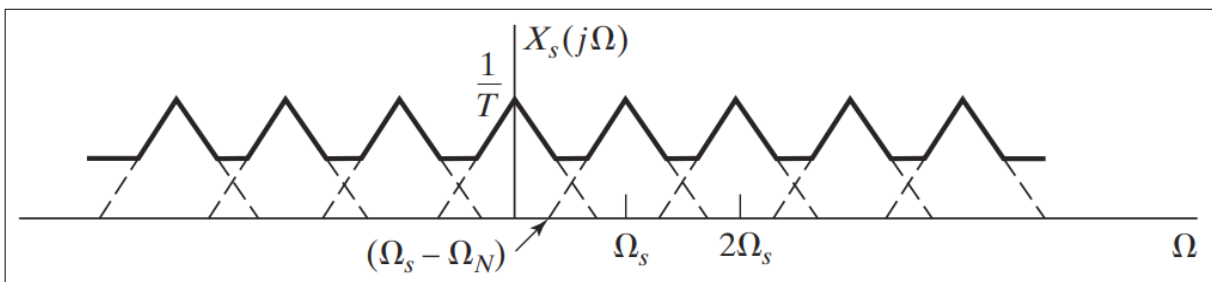
FIGURA 3 – Representação em domínio da frequência de sinal discreto formado a partir de amostragem de $x_c(t)$



Fonte: Oppenheim e Schafer (2010, p. 157);

Os mesmos autores (2010) constatam que, caso os valores de Ω_s e Ω_N forem tais que aquela subtração tenha resultado negativo, X_s contará com a sobreposição de réplicas de X_c ; para um dado sinal original, isso pode ocorrer caso a frequência angular de amostragem seja menor que duas vezes o limite de banda do sinal. Tal fenômeno, denominado “recobrimento” (*aliasing*), é visto na Figura 4.

FIGURA 4 – Representação em domínio da frequência de sinal discreto formado a partir de amostragem de $x_c(t)$ com frequência de amostragem inferior a limite de banda do sinal



Fonte: Oppenheim e Schafer (2010, p. 157);

Ainda conforme Oppenheim e Schafer (2010), X_s pode ser recuperado e transformado de volta em um sinal contínuo, fiel ao original, aplicando-se um filtro passa-baixa caso as réplicas de X_c estiverem devidamente separadas, como é o caso da Figura 3. A transformada de Fourier que se observa na Figura 4, entretanto, não é passível de recuperação e não representa fielmente o sinal original. A situação ilustra o Teorema de Amostragem de Nyquist-Shannon.

O Teorema de Amostragem de Nyquist-Shannon determina que, para um sinal $x_c(t)$ limitado em banda com $X_c(j\Omega) = 0$ para $|\Omega| \geq \Omega_N$, $x_c(t)$ é unicamente determinado pelas suas amostras $x[n] = x_c(nT)$, $n = 0, \pm 1, \pm 2, \dots$ se obedecer à

Equação 3. Ela pode ser entendida também em função de frequências, em *hertz*, f_s e f_N .

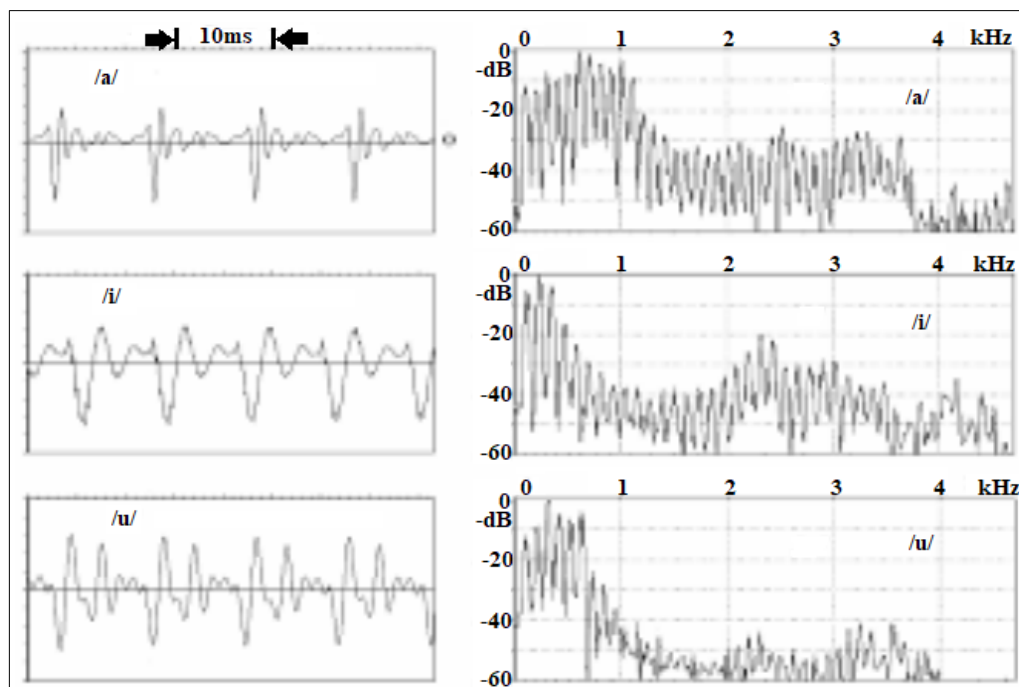
$$\Omega_s = \frac{2 \cdot \pi}{T} \geq 2 \cdot \Omega_N \quad (3)$$

2.2 Frequências da voz humana

A fala humana é um fenômeno extremamente complexo e que envolve a emissão de numerosas informações, oportunizada pela ação não-linear de órgãos humanos como o trato vocal e articuladores dependentes das características pessoais do locutor (SALES; SALES; SILVA, 2021). A vocalização é influenciada por características como sotaque, pronúncia, articulação, nasalidade, tom, volume e velocidade de fala.

Sinais de voz contêm uma riqueza de componentes harmônicos que diferenciam os fonemas pronunciados pelo locutor (RASHID *et al.*, 2008). A Figura 5 exibe exemplos do sinal da pronúncia das vogais /a/, /i/ e /u/ respectivamente. Os gráficos representam os sinais no domínio do tempo à esquerda e da frequência à direita.

FIGURA 5 – Exemplos de sinal das vogais /a/, /i/ e /u/ no domínio do tempo e da frequência



Fonte: Adaptado de Roden (1996) *apud* Rashid *et al.* (2008, p. 2).

Todos os sinais vistos na Figura 5 possuem frequência fundamental f_{fun} em cerca de 100 Hz, mas suas frequências formantes variam. As formantes f_1 , f_2 e f_3 da vogal /a/ se localizam, neste caso, em torno de 600 Hz, 1000 Hz e 3000 Hz respectivamente. Para a vogal /i/ estas estão em 200 Hz, 2300 Hz e 3000 Hz. Por fim, para a vogal /u/, os valores respectivos são 300 Hz, 600 Hz e 2300 Hz.

O ser humano possui limitações em termos de audição e vocalização. Segundo Hartpence (2013), o ser humano é capaz de ouvir sons entre 20 Hz e 20 kHz, havendo um estreitamento dessa faixa para cada indivíduo em função de suas características fisiológicas. A vocalização possível se encontra na faixa de pouco mais de zero a 5 kHz. Esse segundo valor se trata de um registro máximo, ao passo que conversas regulares não tendem a atingir essa extensão.

A maior parte do som vocalizado por pessoas é de baixa frequência, portanto padrões de telecomunicação como o ITU-T Recommendation G.107 consideram a faixa de transmissão de 300 Hz a 3,4 kHz adequada (HARTPENCE, 2013). De acordo com Lathi e Green (2014), testes subjetivos mostram que o entendimento da voz por humanos não é afetado se componentes acima de 3,4 kHz e abaixo de 300 Hz são suprimidos. Ainda, segundo os autores, o foco sobre a transmissão de voz em telecomunicações é a inteligibilidade, não a alta fidelidade do som. Isso permite que se aplique sobre os sinais de voz uma filtragem passa-baixa em 3,4 kHz e que eles sejam fielmente reconstruídos na telefonia com uma frequência de amostragem de 8 kHz, que está acima da frequência de amostragem de Nyquist do sinal de 6,8 kHz.

2.3 Perda de pacotes em comunicação VoIP e picote em sinais de áudio

A tecnologia VoIP é descrita por Hartpence (2013) como o envio de voz e/ou vídeo por meio de uma rede baseada em protocolo de *internet* (*internet protocol* – IP), ou por Elsabrouty, Bouchard e Aboulnasr (2004, p. 1, tradução nossa) como “[...] a transmissão de voz pacotizada sobre redes IP [...]”. Redes de *internet* são comutadas por pacote de dados e cada um é enviado semi-autonomamente, tem seu próprio cabeçalho IP e é encaminhado para seu destino por roteadores separadamente (HARTPENCE, 2013).

Uma característica importante da comunicação VoIP é a natureza da conversa entre pessoas, em que o atraso da mensagem ou perda de sinal causa

uma diminuição de qualidade a ponto de tornar o serviço inutilizável (HARTPENCE, 2013). Isso se demonstra, por exemplo, durante chamadas telefônicas em que os usuários precisam esperar segundos para ouvirem uns aos outros ou têm suas palavras cortadas. Esse panorama manifesta a necessidade da tecnologia VoIP de se trabalhar com dados em tempo real, ou seja, num regime sensível ao tempo (HARTPENCE, 2013).

Situações não envolvidas com tempo real, como a transferência de arquivos ou o acesso a um *site*, são mais comuns na rede de *internet* e não sofrem de maneira impeditiva com o atraso de uma parcela da informação ou sua perda, já que essa pode ser reenviada (HARTPENCE, 2013). De acordo com Lee, Zitouni e Zhou (2004), na ocasião de uma rede congestionada, pacotes de dados podem ser descartados ou transferidos por outras rotas, o que faz com que a chegada no seu destino ocorra com prazos diferentes. Na transmissão de dados em tempo real o descarte de um pacote não pode contar com a sua recuperação, visto que o receptor precisa lidar com um fluxo constante de pacotes de voz para a reprodução do som. O mesmo acontece com pacotes atrasados, visto que sua emissão não pode ser em uma ordem diferente do que o locutor está proferindo.

Por se tratar de uma comunicação em tempo real que requer atrasos mínimos e não aceita o reenvio de pacotes, a tecnologia VoIP geralmente utiliza protocolo de datagrama do usuário (*User Datagram Protocol* – UDP) para seus pacotes de dados (HARTPENCE, 2013). Esse formato não registra a sequência de envio nem faz uso de respostas de confirmação (*acknowledgement* – ACK), fatores indesejáveis neste tipo de comunicação. A Figura 6 ilustra um exemplo de pacote UDP. Nele é possível ver as informações de porta fonte e destino (*Source port* e *Destination port*); no entanto, nenhuma outra informação é relevante à conexão.

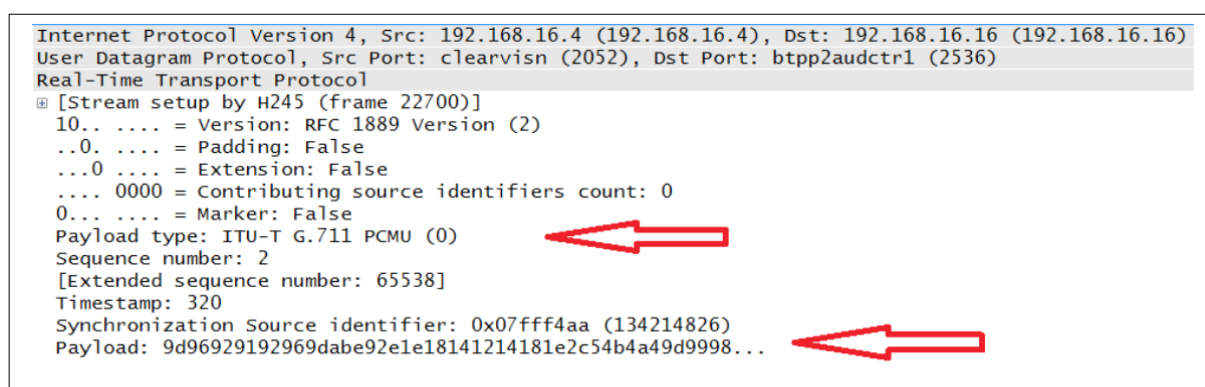
FIGURA 6 – Exemplo de pacote UDP

```
Ethernet II, Src: Avaya_ef:48:f8 (00:04:0d:ef:48:f8), Dst: GiantEle_05:cb:11 (00:09:6e:05:cb:11)
Internet Protocol Version 4, Src: 192.168.16.4 (192.168.16.4), Dst: 192.168.16.23 (192.168.16.23)
User Datagram Protocol, Src Port: clearvisn (2052), Dst Port: tsb2 (2742)
  Source port: clearvisn (2052)
  Destination port: tsb2 (2742)
  Length: 180
  Checksum: 0xd11b [validation disabled]
Real-Time Transport Protocol
```

Fonte: Hartpence (2013, p. 7).

Ao fim do pacote UDP, visto na Figura 6, nota-se o encapsulamento de uma mensagem em protocolo de transporte em tempo real (*Real-time Transport Protocol – RTP*) (HARTPENCE, 2013). Chamadas VoIP usualmente enviam os dados dos sinais de voz por meio desse aparato, como mostra a Figura 7. A flecha vermelha superior indica o tipo de *codec* usado para criar o dado enviado, enquanto a flecha inferior aponta a carga de sinal de voz (*payload*). Esse *payload* caracteriza uma janela de áudio, chamada de *frame*, que representa a sequência de amostras de um período do áudio a ser reproduzido. Cada pacote enviado é responsável, portanto, por carregar um bloco de amostras com comprimento de determinados milissegundos do áudio.

FIGURA 7 – Exemplo de pacote RTP



```

Internet Protocol Version 4, Src: 192.168.16.4 (192.168.16.4), Dst: 192.168.16.16 (192.168.16.16)
User Datagram Protocol, Src Port: clearvisn (2052), Dst Port: btpp2audctrl (2536)
Real-Time Transport Protocol
  [Stream setup by H245 (frame 22700)]
  10.. .... = Version: RFC 1889 Version (2)
  ..0. .... = Padding: False
  ...0 .... = Extension: False
  .... 0000 = Contributing source identifiers count: 0
  0... .... = Marker: False
  Payload type: ITU-T G.711 PCMU (0)
  Sequence number: 2
  [Extended sequence number: 65538]
  Timestamp: 320
  Synchronization Source identifier: 0x07fff4aa (134214826)
  Payload: 9d96929192969dabe92e1e18141214181e2c54b4a49d9998...
  
```

Fonte: Hartpence (2013, p. 96).

Segundo Lee, Zitouni e Zhou (2004), há a possibilidade de haver a perda de pacotes em uma comunicação VoIP dentro de qualquer rede IP existente, independente de qualidade da gestão ou superdimensionamento de equipamentos. Nesse contexto, Gündüzhan e Momtahan (2001) afirmam que tal problema se traduz na perda do conjunto de amostras do sinal no seu respectivo intervalo de tempo.

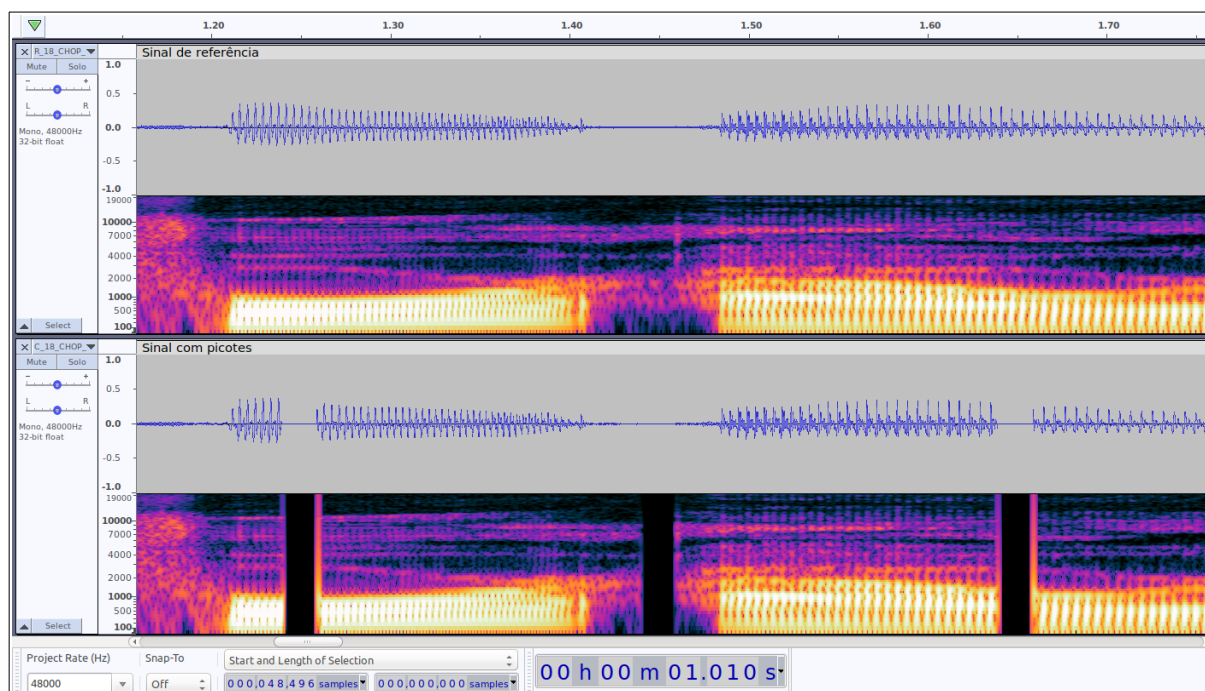
É necessário que o aparelho telefônico recebendo esse sinal lide com a falta daquelas amostras e mantenha o funcionamento dos seus auto-falantes, tarefa que é conduzida pelo *codec* do dispositivo. O *codec* é descrito por Karapantazis e Pavlidou (2009) como o algoritmo que permite que se transporte sinais analógicos em linhas digitais. De acordo com Hartpence (2013), o *codec* é encarregado de transformar um sinal analógico em uma série de amostras e prover valores a elas – ou transformar valores de amostras de volta em sinal analógico, dependendo de qual a ponta da comunicação a discussão se refere. Ainda, conforme esse autor, tal

processo é denominado Modulação por Código de Pulsos (*Pulse Code Modulation* – PCM).

Diversos autores detalham o comportamento básico de um *codec* ao identificar a perda de um pacote. Elsabrouty, Bouchard e Aboulhasr (2004) afirmam que a forma mais simples de se resolver esse problema é reproduzir um pacote mudo, ou seja, silêncio, no período da perda, mas que essa alternativa causa cortes na voz e deteriora a qualidade da chamada mesmo em taxas de perda de pacote muito baixas. Gündüzhan e Momtahan (2001), por sua vez, afirmam que o método mais simples é reproduzir uma réplica do *frame* anterior e que essa forma tem performance melhor que a inserção de silêncio, embora ainda não seja suficiente para aplicações que requerem alta qualidade de áudio. Lee, Zitouni e Zhou (2004) citam essas duas maneiras e adicionam a possibilidade de o *codec* inserir um período de ruído branco ou semi-aleatório, o que é corroborado por Kondo e Nakagawa (2006). De acordo com Hartpence (2013), tais abordagens não podem ser consideradas algoritmos de correção de perda de pacote.

Neste trabalho, considera-se a inserção de silêncio como padrão de comportamento no caso da perda de um pacote, o que é chamado de “picote” (o termo é traduzido ao inglês como “*chop*”, ou “*choppy*” referindo-se a “picotado”). Ilustra-se, na Figura 8, um exemplo de picotes. As duas formas de onda pertencem a arquivos de áudio do conjunto de dados utilizado nesta pesquisa, descrita no tópico 4.1, e foram visualizadas por meio do programa Audacity 3.1.3. Ambas as ondas são provenientes da mesma gravação de voz humana, porém a de cima se trata do sinal original, de referência, enquanto a de baixo simula a existência de picotes. Estão presentes os sinais no domínio do tempo na parte de cima de cada um e em espectrograma na parte de baixo. Para a segunda representação utilizou-se janela Hann de 256 amostras.

FIGURA 8 – Visualização em programa Audacity 3.1.3 de ondas provenientes de mesma gravação de voz, com sinal original no gráfico superior e acometido por picotes no inferior

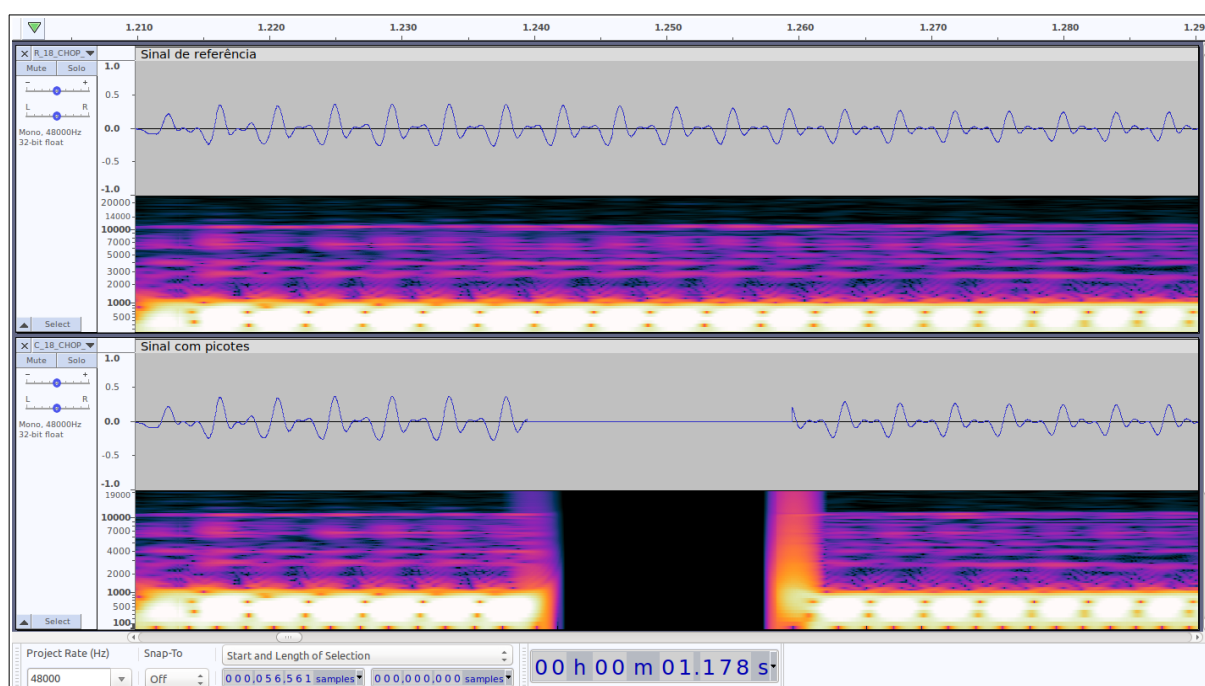


Fonte: Autoria própria (2023).

Percebe-se, no espectrograma de baixo, três regiões de escuro completo. Comparando as formas de onda em azul, é possível observar nas regiões dos dois vazios exteriores as amostras sendo levadas a zero. Essas regiões zeradas são os *frames* perdidos. Cada *frame*, neste áudio, tem duração de 20ms. Na faixa escura do meio ocorre a mesma situação; mas, nesse período, a gravação contou com silêncio e o sinal que se perdeu foi o ruído de fundo, invisível na figura.

Verificando mais detalhadamente o primeiro dos três picotes da Figura 8, obtém-se a imagem ilustrada na Figura 9. No âmbito de diferenças entre os sinais, além do espaço com energia zero, destacam-se as bordas dos *frames* anterior e posterior com o picote. O espectrograma mostra a distorção causada devido ao fato de as janelas de análise nos locais captarem um sinal descontínuo, com queda abrupta. Além de ter sua repercussão na audição humana, o que ocorre como um estalo, esse fenômeno também pode atrapalhar a análise de sinais por ferramentas variadas. O tópico 2.4 aborda o uso de métodos de análise no domínio da frequência para obtenção de dados em ferramentas de análise de voz.

FIGURA 9 – Exemplo de mesmo sinal sem e com picotes, detalhado



Fonte: Autoria própria (2023).

2.4 Ferramentas de análise de voz

A voz humana carrega diversas informações extremamente complexas que o cérebro do ouvinte identifica subconscientemente e com facilidade, como a mensagem passada e identidade, localização no espaço e emoção do locutor (SALES; SALES; SILVA, 2021). Ferramentas de análise de voz têm a tarefa de adquirir essas informações em sinais de voz, o que para Tebelskis (*apud* SALES; SALES; SILVA, 2021) é o reconhecimento de padrões multinível, cujo sinal é examinado e estruturado em uma hierarquia de unidades de fonemas, palavras, frases e sentenças.

2.4.1 Biometria e reconhecimento de voz

De acordo com Rashid *et al.* (2008), a identificação biométrica é uma alternativa a senhas alfanuméricas para autenticação e acesso. Esses métodos são propensos a hackeamento ou roubo de informações, enquanto características biométricas são únicas ao indivíduo e o corpo se torna sua “senha”. Segundo Rabiner e Schafer (1978), técnicas de verificação têm como meta decidir se uma pessoa falante corresponde a quem alega ser. Técnicas de identificação, no entanto,

devem reconhecer a qual registro uma pessoa falante corresponde dentre uma série de vozes documentadas.

Rashid *et al.* (2008, p. 2, tradução nossa) afirmam que “sistemas de reconhecimento de voz incorporam diversas variáveis ou parâmetros no reconhecimento da voz ou padrão de fala de uma pessoa, incluindo tom (*pitch*), dinâmicas e forma de onda.”. A tarefa realizada para reconhecimento de voz é precedida por uma etapa de pré-processamento de sinal, que envolve a medida de níveis de energia, a eliminação de ruídos com filtros digitais e a extração de características do sinal.

Extratores de características são descritos por Silva *et al.* (2019) como métodos computacionais que transformam dados complexos em outros compactados por meio de funções matemáticas, preservando as principais e mais necessárias características do sinal e reduzindo o volume de informações. Os autores citam como exemplos de métodos para a extração de características a codificação por predição linear (*Linear Predictive Coding* – LPC), coeficientes mel-cepstrais (*Mel-Frequency Cepstral Coefficients* – MFCC), transformada de distribuição cumulativa (*Cumulative Distribution Transform* – CDT) e aritmética intervalar. O LPC é apresentado da seguinte forma:

Linear Predictive Coding (LPC) é o principal método para extrair coeficientes de sinais de voz humana. O princípio deste método se baseia na análise de amostras anteriores contidas no sinal; a predição linear aponta que uma amostra pode ser predita (valor aproximado da amostra) por uma combinação linear dos valores das amostras anteriores, considerando a correlação entre elas (COSTA, 2013). A estimação de cada amostra se dá por meio de uma combinação linear de p amostras anteriores, na qual é possível afirmar que um valor de p maior fornece um modelo mais preciso. Como o método LPC se utiliza dos parâmetros presentes no sinal de voz (fala) que representam o trato vocal, qualquer mudança na anatomia do trato vocal compromete os coeficientes LPC obtidos. (SILVA *et al.*, 2019, p. 81).

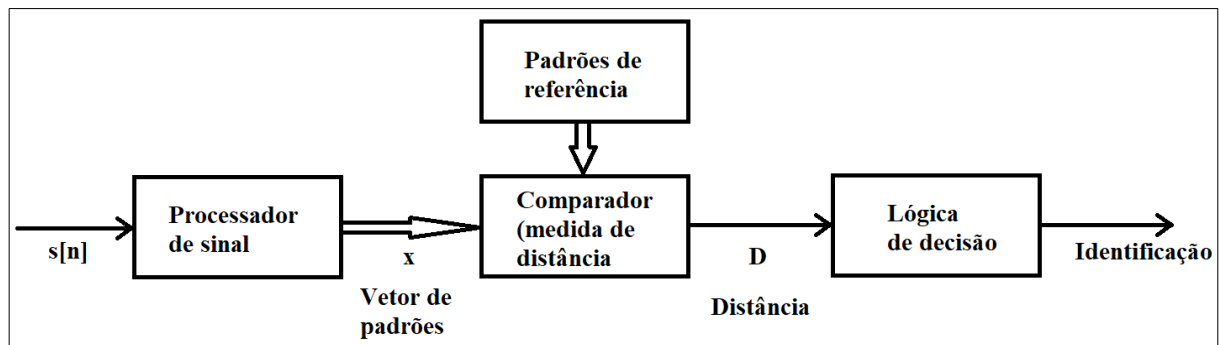
O MFCC, por sua vez, é um método que se baseia na escala Mel de frequências, que possui características mais perceptíveis ao ouvido humano, capaz de discernir muito bem pequenas mudanças entre baixas e altas frequências (CUADROS *et al.*, 2007 *apud* SILVA *et al.*, 2019, p.4). De acordo com Silva *et al.* (2019), a escala Mel contribui nos processos computacionais aproximando os resultados adquiridos aos que seriam percebidos naturalmente por um ser humano. A relação entre escala Mel e Hertz é logarítmica, como mostra a Equação 4 (SILVA

et al., 2019, p. 82). As variáveis f_{mel} e f_{hertz} representam as frequências em escala Mel e em hertz respectivamente.

$$f_{mel} = 1125 \cdot \ln \left(1 + \frac{f_{hertz}}{700} \right) \quad (4)$$

Rabiner e Schafer (1978) afirmam que o reconhecimento de voz se trata basicamente de um problema de reconhecimento de padrões, cuja primeira etapa é constituída de técnicas de processamento digital de sinais. Segundo a descrição dos autores, um vetor de padrões pode ser obtido de um sinal $s[n]$ a partir de um processamento digital de voz que preserve as características relativas à identidade do locutor. Esse conjunto resultante é comparado a uma série de padrões de referência previamente obtidos e, após uma lógica de decisão, tem-se a escolha de melhor candidato dentre as alternativas disponíveis. A Figura 10 apresenta um diagrama desse sistema.

FIGURA 10 – Representação de sistema de reconhecimento de voz



Fonte: Adaptado de Rabiner e Schafer (1978, p. 476).

2.4.2 Transcrição de áudio

Rabiner e Schafer (1978, p. 7, tradução nossa) descrevem brevemente a transcrição de áudio como “[...] a conversão de uma forma de onda acústica ao equivalente escrito da sua mensagem.”. Os autores dão como exemplo de aplicação da transcrição de áudio uma máquina de escrever operada por voz. Trazendo o conceito a tempos atuais, é possível citar o uso da transcrição em aparelhos celulares e a escrita de mensagens de texto por locução de voz.

A estrutura de análise da transcrição de áudio é muito similar à de identificação de usuário vista anteriormente (RABINER; SCHAFER, 1978). Com a mesma abordagem de processamento de sinal para obtenção de padrões de características e a comparação com padrões de referência, tem-se, nesse caso, como objetivo a identificação de quais fonemas, palavras e frases foram proferidos.

2.5 Ocultação de perda de pacotes

No cenário de projetos de equipamentos para VoIP é necessário atentar-se à possibilidade inerente da perda de pacotes (LEE; ZITOUNI; ZHOU, 2004, p. 1). Os problemas causados pela falta do sinal são significativos e trazem à tona a necessidade de se incluir nos *codecs* alternativas mais inteligentes do que as vistas anteriormente (HARTPENCE, 2013), como inserção de período de silêncio ou replicação do *frame* anterior. O dano pode ser mitigado utilizando-se técnicas de ocultação de perda de pacotes (*Packet Loss Concealment* – PLC) ou correção adiantada de erro (*Forward Error Correction* – FEC). Esses processos podem ser classificados como baseados no aparelho emissor (*sender-based*) ou receptor (*receiver-based*) do pacote (SANNECK *et al.*, 1996).

Métodos FEC são fundamentados no envio de informação adicional em um pacote de voz para contornar a possível perda de outro. Karapantazis e Pavlidou (2009, p. 2053) descrevem FEC como “[...] a adição de redundância à informação transmitida utilizando um algoritmo pré-definido, desta forma permitindo que o receptor detecte e corrija erros dentro de um limite”. A informação redundante, usada para determinar o sinal perdido em outros pacotes, é derivada do pacote original, utilizando técnicas de teoria de código (CARLE; BIRSACK, 1999).

Métodos FEC são classificados como *sender-based*. A introdução de dados de recuperação nos pacotes de VoIP dificulta o uso de FEC em ambientes de transmissão heterogênea, em que aparelhos transmissores utilizam ferramentas de áudio diferentes dos receptores (SANNECK *et al.*, 1996). O uso de métodos de PLC permite que os receptores mitiguem a perda de pacote conforme suas necessidades específicas de qualidade, pois geralmente são abordagens *receiver-based*.

Ao contrário de fazer uso da rede para enviar mais dados como a FEC, os métodos de PLC inferem o sinal perdido por meio do processamento digital dos *frames* de voz recebidos anterior ou posteriormente (HARTPENCE, 2013). De

acordo com o apêndice I da Recomendação G.711 (ITU-T, 1999), o objetivo da PLC é gerar um sinal de voz sintético para cobrir a perda de pacote em um fluxo de dados, idealmente havendo o mesmo timbre e características espectrais do sinal perdido sem criar deformações não-naturais. Devido ao sinal da voz humana ser “[...] geralmente localmente estacionário, é possível utilizar o passado dele para gerar uma aproximação razoável do segmento faltante [...]” (ITU-T, 1999, p. 1, tradução nossa), que “[...] provavelmente terá características similares à do pacote recebido anteriormente.” (LEE; ZITOUNI; ZHOU, 2004, p. 1, tradução nossa). Esse fato é fortalecido por Perkins, Hodson e Hardman (1998, p. 44, tradução nossa), que afirmam:

Esquemas de ocultação de erro se baseiam em uma substituição para o pacote perdido que seja igual ao original. Isso é possível pois sinais de áudio, em particular de voz, possuem grande autossimilaridade a curto prazo. Assim sendo, estas técnicas [de PLC] funcionam para taxas de perda de pacote relativamente pequenas (menores que 15%) e para pacotes pequenos (até 40ms). Quando a perda se aproxima do comprimento de um fonema (até 100ms), estas técnicas quebram, pois fonemas inteiros podem ser perdidos pelo ouvinte. (PERKINS; HODSON; HARDMAN, 1998, p. 44, tradução nossa)

O *codec* de um aparelho receptor funciona recebendo e decodificando pacotes VoIP ao longo do tempo, de forma a compor um fluxo reprodutível de amostras de áudio. Conforme Gündüzhan e Momtahan (2001), ao identificar a falta de um pacote, o algoritmo de PLC reconstrói um novo sinal para haver a substituição; em caso contrário, é feita a decodificação normalmente. Este estudo trata da reconstrução de sinal em arquivos de áudio, ou seja, não é compreendido o projeto de um *codec* ou o recebimento de pacotes, pois todo o sinal está disponível previamente. Apesar disso, é possível recorrer a métodos de reconstrução de sinal similares aos utilizados em PLC para fazer o mesmo processo nos *frames* zerados de um arquivo.

3 METODOLOGIA

Este trabalho busca implementar e comparar algoritmos para correção de picotes em arquivos de áudio, com a finalidade de providenciar melhores resultados quando aplicados a ferramentas de análise de voz por inteligência artificial como biometria de voz e transcrição de áudio. Para atingir esse objetivo, é necessário haver uma pesquisa extensa a respeito de diversos fatores. A pesquisa é descrita por Gil (2008, p. 26) como “o processo formal e sistemático de desenvolvimento do método científico.”, tendo como objetivo fundamental “descobrir respostas para problemas mediante o emprego de procedimentos científicos.”.

O mesmo autor (2002) afirma que é possível classificar as pesquisas entre três grupos em função dos seus objetivos gerais: pesquisas exploratórias, descritivas e explicativas. Este trabalho conta com uma pesquisa de cunho descritivo, visto que tal grupo tem como objetivo principal “a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis.” (GIL, 2002, p. 42). Conforme Prodanov e Freitas (2013, p. 53), “uma das características mais significativas das pesquisas descritivas é a utilização de técnicas padronizadas de coleta de dados, como o questionário e a observação sistemática.”. Esses são os intuitos de se realizar uma pesquisa acerca do comportamento de ferramentas de análise de áudio, arquivos de áudio acometidos por picotes e algoritmos de correção destes.

Ainda no sentido de classificar a pesquisa executada, Menezes *et al.* (2019) discorrem sobre a distinção à vista de abordagem, dividindo-a em qualitativa e quantitativa. Essa característica é capaz de informar “[...] quais os direcionamentos que o estudo tomará em sua feitura.” (MENEZES *et al.*, 2019, p. 29). Enquanto a abordagem quantitativa é marcada pelo uso de técnicas e recursos estatísticos, a fim de se classificar e analisar informações, a qualitativa é marcada pela interpretação de fenômenos e atribuição de significados (UNIVERSIDADE FEDERAL DE SANTA CATARINA, 2005). Este trabalho conta com dados numéricos, mas não os trata estatisticamente, a pesquisa é focada em aclarar fenômenos e métodos, portanto, deve ser classificada como qualitativa.

3.1 Métodos aplicados

O cumprimento dos diversos objetivos específicos deste trabalho requer uma série de procedimentos técnicos. A seguir, são descritas as etapas para a realização desta pesquisa, que envolve desde a construção de um referencial teórico quanto à aplicação de técnicas de processamento de áudio em arquivos problemáticos e seu uso em ferramentas profissionais de análise de voz.

Este trabalho foi iniciado com uma pesquisa bibliográfica extensa, a fim de se firmar um referencial teórico voltado a processamento digital de sinais, picote em arquivos de áudio, ferramentas de análise de voz com aprendizado de máquinas e algoritmos de correção de picote, além de uma contextualização sobre tecnologia VoIP, *codecs* e os cenários que causam o surgimento de tal problema.

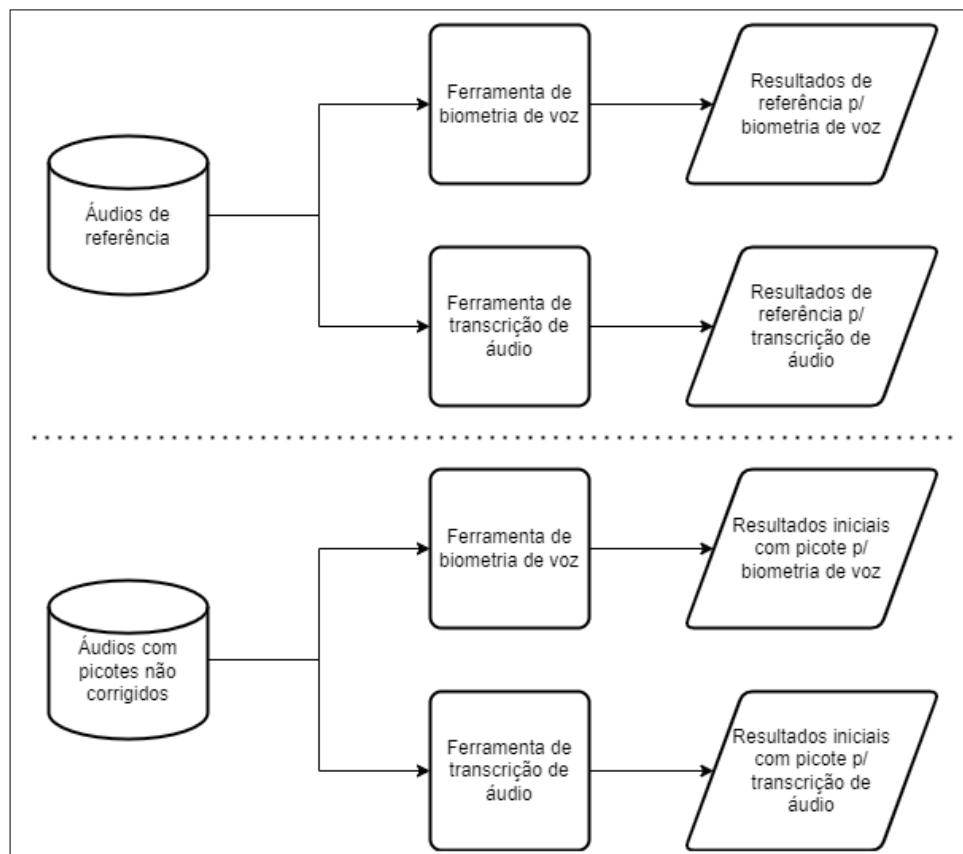
A primeira tarefa executada envolvendo procedimentos técnicos foi a definição das ferramentas de análise de voz. O tema deste trabalho foi elaborado tendo em vista a interface de programação de aplicação (*Application Programming Interface* – API) de biometria de voz da empresa Khomp Indústria e Comércio Ltda., por ter se mostrado um ponto com espaço para melhorias, suporte por parte da empresa e uma boa estrutura para testes. Verificaram-se as características necessárias de arquivos de áudio para uso de tal dispositivo, como número de canais, limitações de duração e frequência de amostragem. Fez-se o mesmo processo para uma ferramenta de transcrição de áudio, conferindo sua disponibilidade como biblioteca na linguagem de programação ou acesso aberto para envio dos áudios por meio da *internet*.

Em seguida, foi necessário obter um conjunto de dados (*dataset*) para examinar as ferramentas. Buscaram-se áudios com vozes humanas de diversos locutores, em duplas com e sem presença de picote, havendo consistência sobre similaridade entre arquivos dentro de cada par, metodologia de produção e a qualidade desses conjuntos. Verificou-se a conformidade dos áudios angariados com as ferramentas de análise citadas anteriormente e fizeram-se as adaptações necessárias, monitorando a usabilidade deles em relação à frequência de amostragem e aspectos dos sinais no domínio do tempo e da frequência.

Em posse de todas as ferramentas e *dataset* necessários, foi feita a análise primária dos áudios com a biometria de voz e a transcrição de áudio. Cada ferramenta foi executada uma vez com o conjunto de áudios de referência e outra

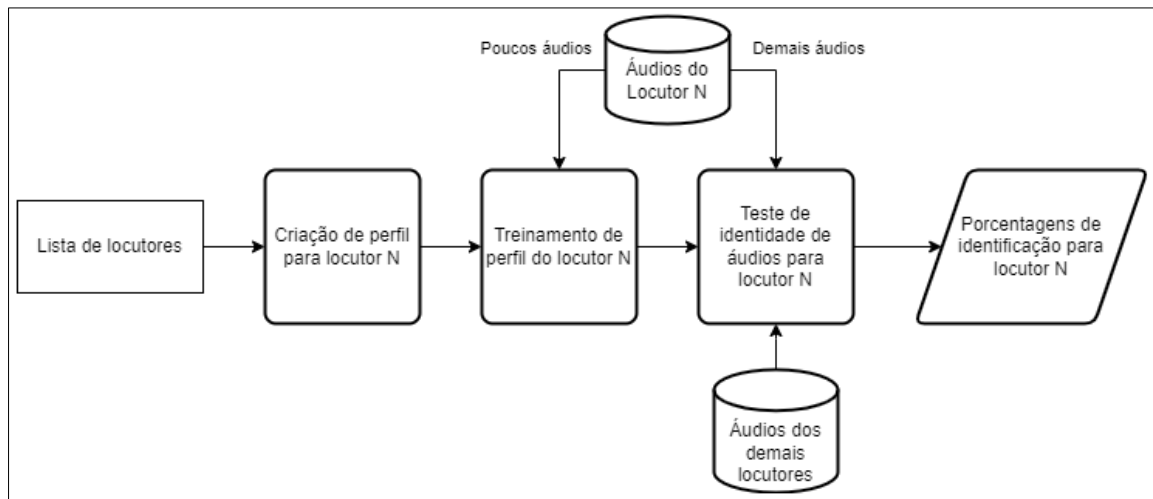
com o de áudios afetados por picote não corrigidos. A Figura 11 ilustra basicamente o processo, exibindo a obtenção dos resultados para cada conjunto de dados em cada ferramenta.

FIGURA 11 – Processo inicial de obtenção dos resultados para áudios de referência e com picote não corrigidos



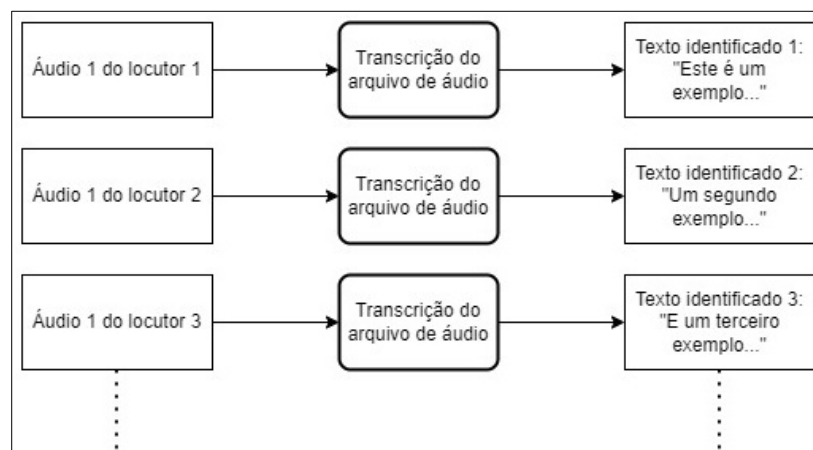
Fonte: Autoria própria (2023).

Dentro do processo visto na Figura 11, para a obtenção de resultados da biometria de voz, foi treinado na ferramenta um perfil para cada locutor com seus três primeiros áudios, para armazenagem de suas características vocais. Após isso, foi testada a porcentagem de confiança da identidade para cada outro áudio do conjunto, tanto do mesmo locutor quanto dos demais. A medida de confiança se refere à similaridade entre as características vocais contidas em uma gravação de voz e as pertencentes a um perfil treinado. A Figura 12 traz um esquema ilustrativo desse procedimento, que é melhor descrito no tópico 4.2. Utilizou-se de exemplo um locutor “N” do *dataset*, o que deve ser feito para cada um dos quatro locutores do conjunto.

FIGURA 12 – Processo de obtenção de resultados para ferramenta de biometria de voz

Fonte: Autoria própria (2023).

A ferramenta de transcrição de áudio contou com um processo mais simples: executou-se a análise de cada áudio dos conjuntos salvando seu resultado textual em um arquivo para, posteriormente, haver uma diferenciação de caracteres entre pares de arquivos de referência e picotados. Observa-se, na Figura 13, a aplicação dos diversos áudios de um *dataset* à ferramenta e a obtenção do respectivo texto identificado para cada arquivo.

FIGURA 13 – Processo de obtenção de resultados para ferramenta de transcrição de áudio

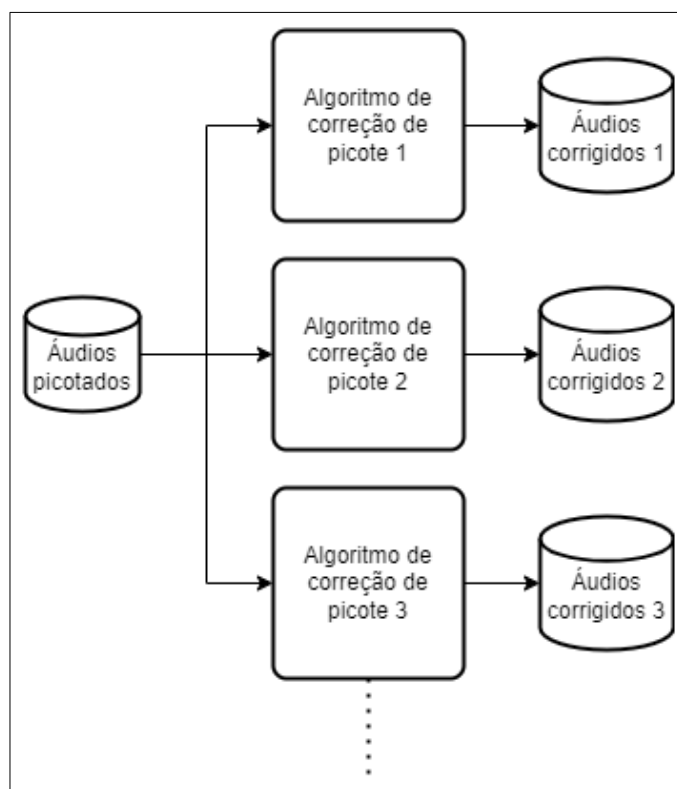
Fonte: Autoria própria (2023).

O passo seguinte foi a implementação de algoritmos de correção de picotes. Tal tarefa foi realizada apreciando métodos e tecnologias visitadas na fundamentação teórica e implementando-as em *Python*. As equações de processamento digital de sinais foram implementadas manualmente como descritas

nos textos, contando em alguns casos com adaptações para concordância entre autores.

Com as ferramentas de processamento dos arquivos picotados implementadas, fez-se, para cada uma, a aplicação no conjunto de áudios acometidos por picotes, gerando uma terceira classe de arquivos – os áudios corrigidos. A Figura 14 permite a visualização desse passo.

FIGURA 14 – Geração de *datasets* de áudios corrigidos



Fonte: Autoria própria (2023).

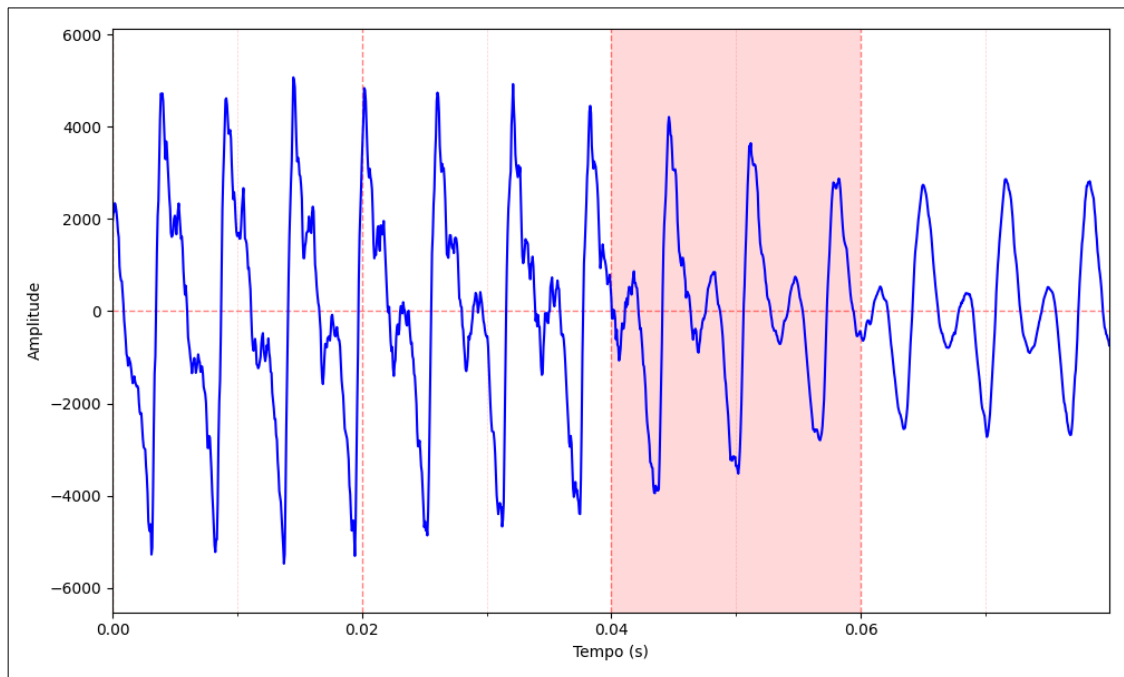
Por fim, repetiu-se a etapa de inspeção dos áudios com as ferramentas de análise de voz, porém com os conjuntos de áudios corrigidos. Os resultados foram obtidos e salvos da mesma forma, podendo, assim, ser comparados com os de antes da aplicação dos algoritmos implementados. Os resultados foram comparados numericamente para determinar a melhora fornecida pelas estratégias propostas.

4 APRESENTAÇÃO DOS RESULTADOS

Este capítulo apresenta os resultados obtidos na aplicação dos algoritmos de correção de picote. São abordadas as ferramentas de análise de voz utilizadas e suas formas de aplicação e medição, além do processo de obtenção de uma base de áudios apropriada e suficiente para os testes. Ademais, discorre-se sobre os algoritmos implementados, suas origens em abordagens teóricas, as formas de onda obtidas e os resultados numéricos para cada caso. Por fim, os resultados de todos os algoritmos são comparados e identifica-se os de melhor desempenho para uso com cada ferramenta de análise de voz.

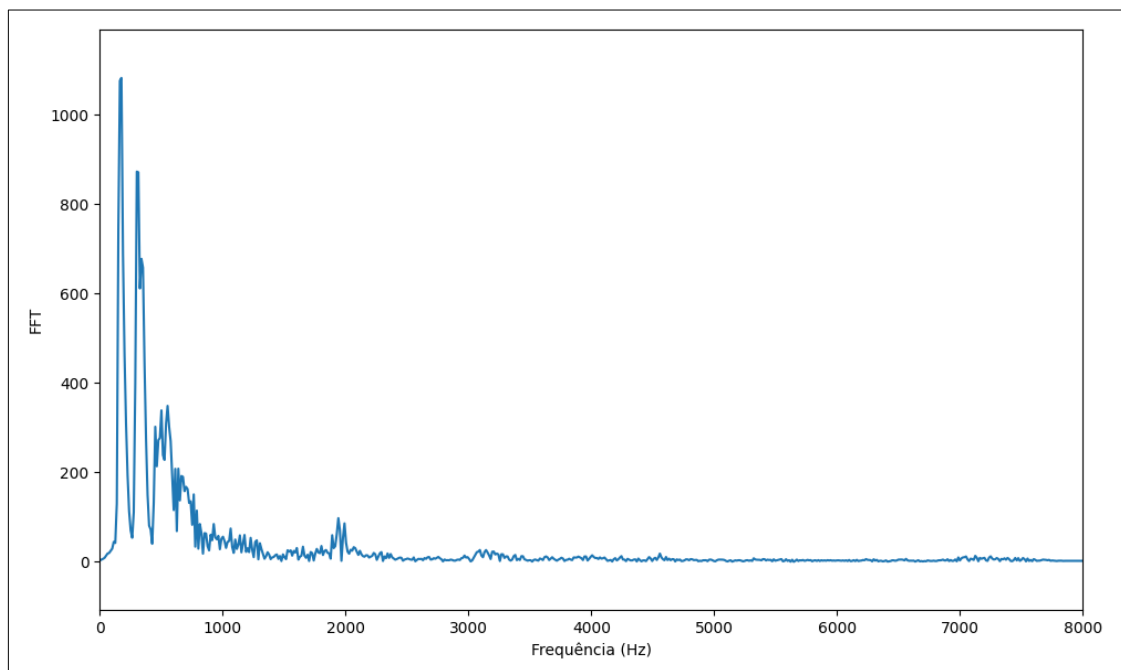
4.1 Técnicas de reconstrução de áudio

Os tópicos a seguir apresentam técnicas de reconstrução de sinal utilizadas em PLC. As demonstrações gráficas foram realizadas implementando os métodos de processamento em *Python 3* e aplicando sobre um sinal de voz real, extraído do *dataset* utilizado nos testes. Retirou-se um segmento de 80ms do arquivo de referência “*R_01_CHOP_FA.wav*” após sua conversão para 16 kHz, como tratado no tópico 4.1.2, com características que evidenciam as particularidades das técnicas de reconstrução. Foi escolhido o período de 6,06s a 6,14s do áudio. A Figura 15 apresenta o sinal citado, onde é possível notar que ele se comporta de maneira semelhante à estacionária nas regiões inicial e final, mas há uma transição entre elas no meio da forma de onda.

FIGURA 15 – Sinal original utilizado na demonstração dos métodos de PLC

Fonte: Autoria própria (2023).

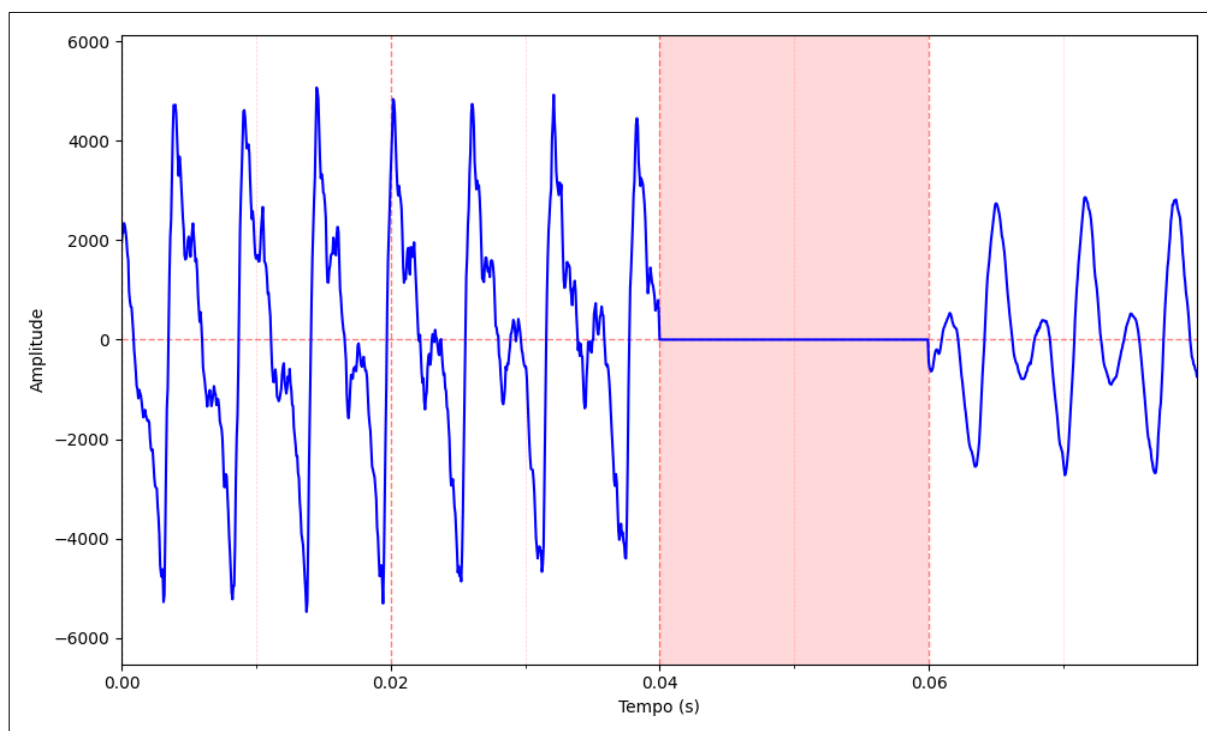
A Figura 16 apresenta a magnitude da DFT do sinal visto anteriormente, onde destacam-se as principais frequências formantes. Esses valores encontram-se em torno de 175 Hz, 305 Hz, 550 Hz e 1950 Hz.

FIGURA 16 – Magnitude da DFT do sinal original utilizado na demonstração dos métodos de PLC

Fonte: Autoria própria (2023).

A forma de onda tem frequência de amostragem de 16 kHz, duração de 80ms e considera-se dividida em quatro frames de 20ms cada. Para apresentar os métodos de PLC, o terceiro frame do sinal é transformado em um picote, ou seja, tem todas as suas amostras levadas a zero. Essa região é marcada em vermelho no gráfico da Figura 17, a qual apresenta o sinal de demonstração com essa modificação.

FIGURA 17 – Sinal picotado utilizado na demonstração dos métodos de PLC



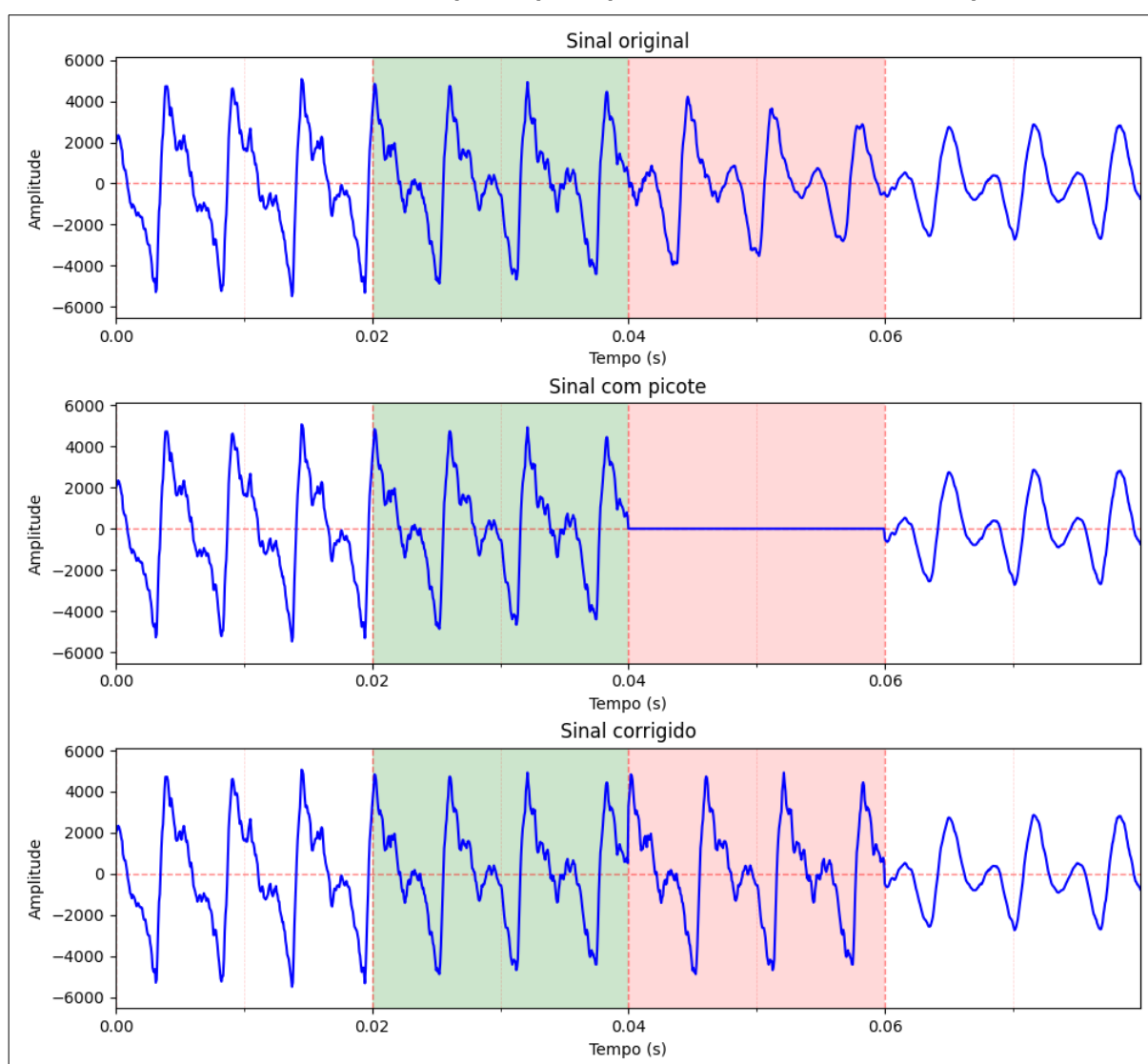
Fonte: Autoria própria (2023).

4.1.1 Pitch Waveform Replication – PWR

Como visto anteriormente, uma das formas de o *codec* lidar com a perda de um pacote de voz é a replicação completa do *frame* anterior. Embora seja uma forma com melhor performance que a inserção de silêncio (GÜNDÜZHAN; MONTAHAN, 2001), esse procedimento não pode ser classificado como PLC porque tais algoritmos “tentam determinar o que o sinal seria ou deveria ser caso o pacote não tivesse sido perdido.” (HARTPENCE, 2013, p. 134). Em outros termos, essas abordagens devem buscar conjecturar um *frame* ausente a partir das informações disponíveis – como os *frames* próximos.

Além da classificação do procedimento, a mera repetição do sinal pode gerar descontinuidades nas bordas do *frame* copiante, o que ainda torna o áudio desagradável ao ouvido humano (LEE; ZITOUNI; ZHOU, 2004). A Figura 18 ilustra essa ocorrência; nela pode-se perceber que o gráfico superior traz o sinal original de demonstração, o do meio conta com o picote e o inferior mostra o resultado de uma correção por replicação de *frame* anterior completo. As regiões em vermelho apontam a área de ocorrência do picote, o terceiro *frame*, a ser corrigido. As zonas em verde delimitam a faixa do sinal utilizada para replicação.

FIGURA 18 – Correção de picote por repetição de *frame* anterior completo



Fonte: Autoria própria (2023).

É perceptível, no sinal corrigido, a ocorrência de descontinuidade nas bordas do frame em vermelho – neste caso, destaca-se a inicial, aos 40ms. No

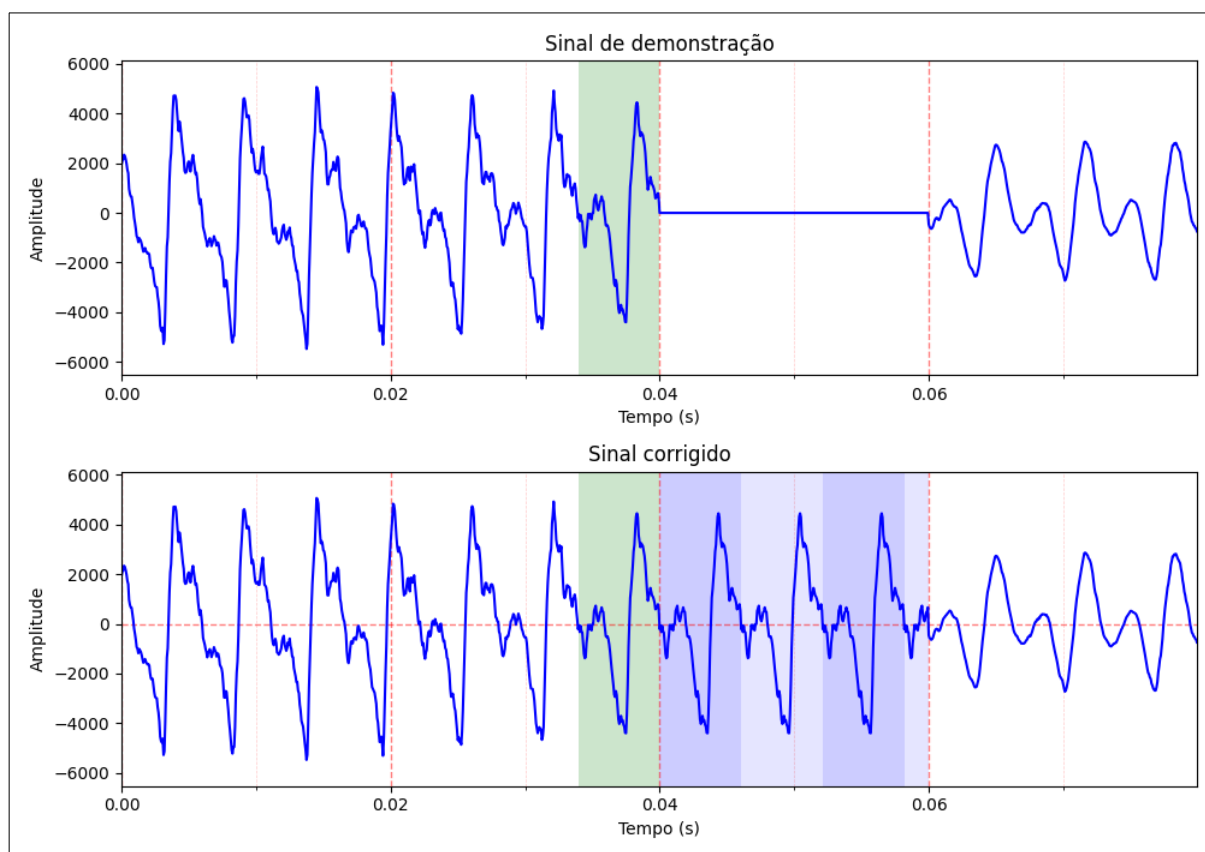
sentido de mitigar esse problema, surge a replicação de tom do sinal (*Pitch Waveform Replication – PWR*).

O tom (*pitch*) de um sinal corresponde à sua frequência fundamental (GOLD, 1977), ou seja, a primeira frequência com energia considerável da série harmônica deste. Equivalentemente, o período de *pitch* se refere ao período fundamental do sinal. Tendo em vista que sinais de voz são considerados quase-estacionários, ou seja, sofrem poucas mudanças num período curto de 10 a 30ms (ELSABROUTY; BOUCHARD; ABOULNASR, 2004); uma região de uma forma de onda de voz pode conter vários períodos de *pitch* com formatos semelhantes.

O intuito da PWR é preencher o espaço do *frame* perdido com cópias repetidas do período de *pitch* anterior ao picote, o que se trata de apenas uma parcela do *frame* anterior, ao invés do *frame* anterior completo (TOYOSHIMA; SHIMAMURA, 2014). Faz-se uso de ferramentas matemáticas como a autocorrelação normalizada, autocorrelação não-normalizada ou função de diferença de magnitude média (*average magnitude difference function – AMDF*) para calcular o período fundamental do sinal (ITU-T, 1999), a fim de se obter uma repetição que se encaixe sem descontinuidades. A transição entre o *frame* anterior e o picote tende a ser natural, mas, de acordo com Toyoshima e Shimamura (2014, p. 91), “o demérito da PWR é criar uma descontinuidade na borda entre o pacote perdido e o próximo recebido”.

O processo da PWR é dado em duas partes: a detecção do período de *pitch* e a sua replicação (ITU-T, 1999). A fim de se exemplificar inicialmente, a Figura 19 ilustra uma replicação de período de *pitch*. O sinal de demonstração utilizado, no gráfico superior, conta com a primeira metade do sinal de demonstração original e o resto em zero. O sinal inferior traz essa mesma seção e replica o período de *pitch* anterior à região nula até o final dos 80ms. As áreas marcadas em verde denotam o período de *pitch* calculado, enquanto cada zona azul se refere a uma replicação individual. Os tópicos adiante expõem como se dão os processos para essa correção.

FIGURA 19 – Exemplo de replicação de tom do sinal



Fonte: Autoria própria (2023).

Como se pode verificar na Figura 19, as transições entre os períodos replicados são suaves e sem desencontros, da mesma forma que se comporta o sinal até os 40ms. É vista ainda, entretanto, a desvantagem da transição entre um sinal replicado por PWR e um *frame* do sinal original recebido em seguida, aos 60ms, como citado anteriormente.

4.1.1.1 Detecção de período de pitch

Diversos autores recorrem à autocorrelação normalizada para a detecção de período de *pitch* em um sinal de voz. A correlação é uma ferramenta matemática utilizada para determinar a similaridade entre dois sinais $x[n]$ e $y[n]$, sendo que, se esses forem distintos, toma-se o termo “correlação cruzada”; enquanto que, se essa similaridade for aplicada a duas partes defasadas de um mesmo sinal, a ferramenta é referida como “autocorrelação” (LATHI; GREEN, 2014).

Gündüzhan e Momtahan (2001) fornecem um retrato da autocorrelação de um sinal digital em função da defasagem. Assim como a recomendação de algoritmo de PLC para o *codec* G.711 (ITU-T, 1999), esses autores afirmam que o período de *pitch* em amostras é estimado encontrando o pico da autocorrelação normalizada do sinal de voz em um segmento imediatamente anterior ao picote do áudio. A Equação 5 apresenta tal fórmula, em que P é o período de *pitch* encontrado, $e[n]$ é o sinal avaliado, M é o comprimento do sinal $e[n]$, C é o comprimento da janela de cálculo da autocorrelação, i é o atraso (*lag*) de cálculo da autocorrelação e $P_{\text{MÍNIMO}}$ e $P_{\text{MÁXIMO}}$ são os valores mínimo e máximo determinados para P respectivamente (GÜNDÜZHAN; MOMTAHAN, 2001).

$$P = \underset{i=P_{\text{MÍNIMO}} \dots P_{\text{MÁXIMO}}}{\arg \max} \left\{ \frac{\sum_{n=M-C}^M e[n-i] \cdot e[n]}{\sqrt{\sum_{n=M-C}^M e[n-i] \cdot e[n-i]}} \right\} \quad (5)$$

As variáveis $P_{\text{MÍNIMO}}$ e $P_{\text{MÁXIMO}}$ provêm da necessidade de se limitar a frequência de *pitch* detectada. Conforme Gündüzhan e Momtahan (2001), a detecção de *pitch* assume que as amostras do frame analisado são correlacionadas, o que não ocorre no caso da emissão de fonemas mudos (*unvoiced*). Impor um limite de período de *pitch* mínimo dificulta a ocorrência de artefatos tonais, ou seja, barulhos indesejados, no sinal recriado. Os fonemas mudos são aqueles produzidos sem a vibração das cordas vocais e tendem a ser abruptos, como as consoantes /p/, /k/ e /t/ (UNIVERSITY OF CALIFORNIA, 2015).

Visto que a detecção de *pitch* busca extrair uma característica local do sinal referente apenas à vizinhança do picote, somente um fragmento é necessário para realizar tal procedimento. Na Equação 5 é possível determinar, em termos das variáveis utilizadas, a faixa do sinal original a que $e[n]$ corresponde, ou seja, o trecho do sinal necessário para se executar o cálculo de determinação do *pitch*. O método de PLC do Anexo A do *codec* G.711 (ITU-T, 1999) faz a detecção encontrando o pico da autocorrelação dos 20ms de voz mais recentes ao picote com o sinal anterior em atrasos de 5ms a 15ms, correspondentes às frequências 200 Hz e 66 Hz. Os autores afirmam que tais limites de frequência foram escolhidos com base na banda utilizada pelo filtro pós-processamento do *codec* G.728. Gündüzhan e Momtahan

(2001), no entanto, propõem o limite de períodos de *pitch* em 2,5ms e 15ms, ou seja, frequências de 400 Hz e 66,7 Hz. Utiliza-se o segundo caso para as demonstrações. Para uma frequência de amostragem de 16 kHz, as frequências menor e maior geram os valores de $P_{\text{MÍNIMO}}$ e $P_{\text{MÁXIMO}}$ em 240 amostras e 40 amostras, respectivamente. Além disso, retira-se da fala de ITU-T que a janela de cálculo da autocorrelação é de 20ms, o que nesse cenário traz um C igual a 200 amostras.

O comprimento do sinal de análise $e[n]$ deve ser tal que os índices dos vetores da Equação 5 tenham valor mínimo zero, para não haver a busca em nenhum índice negativo. Observa-se que a única variável com relação negativa aos índices de e é i , que percorre de $P_{\text{MÍNIMO}}$ a $P_{\text{MÁXIMO}}$. Sendo o valor máximo de i igual a $P_{\text{MÍNIMO}}$, ou 240 amostras, os limites inferiores dos somatórios devem ter $n_{\text{mínimo}}$ como esse valor. Dessa forma tem-se a Equação 6, sendo M o comprimento do sinal $e[n]$ e C o tamanho da janela de cálculo da autocorrelação, ambos em amostras:

$$n_{\text{mínimo}} = M - C \quad (6)$$

Onde:

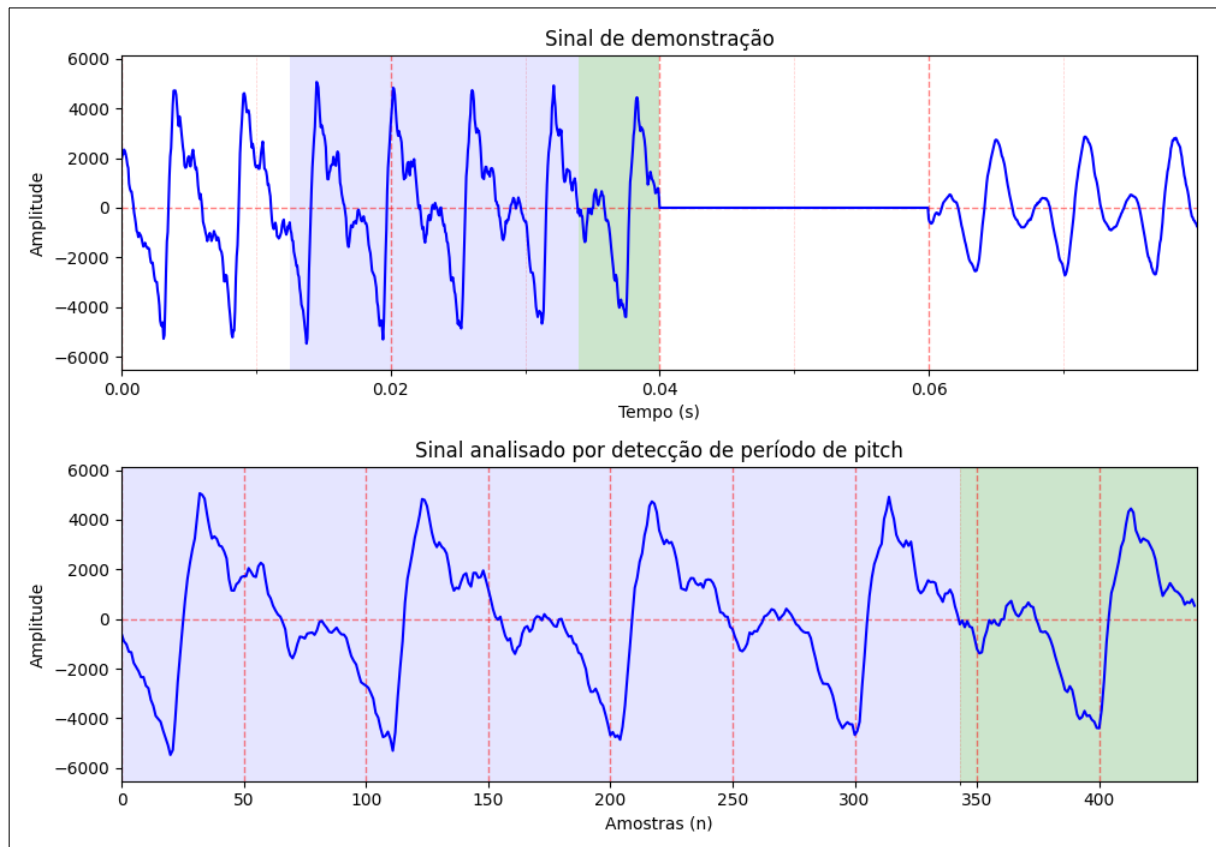
$$M = 240 \text{ amostras} + 200 \text{ amostras}$$

$$M = 440 \text{ amostras}$$

Com isso, conclui-se que o sinal da detecção de período de *pitch* deve compreender as 440 amostras anteriores ao picote, ou seja, as 320 amostras do *frame* anterior e as últimas 120 do anterior a esse.

Voltando ao sinal de demonstração, é possível calcular o período de *pitch* anterior ao picote. Implementando a Equação 5 em *Python* e aplicando os valores citados, encontra-se o valor de 97 amostras, o que com uma taxa de amostragem de 16 kHz indica que a frequência fundamental da região é de aproximadamente 164,9 Hz. A Figura 20 ilustra a forma de onda de demonstração no gráfico superior e o detalhe da região tomada para análise de detecção de *pitch* no inferior. A extensão incluindo ambas as áreas azul e verde caracteriza a região de detecção de *pitch*, ou seja, o sinal $e[n]$ da Equação 5. A zona em verde é a região anterior ao picote com comprimento igual ao valor calculado de 97 amostras.

FIGURA 20 – Detalhe da região de detecção de período de *pitch* anterior a um picote



Fonte: Autoria própria (2023).

Tendo posse do valor de *pitch* calculado, é possível estender o sinal ao longo do *frame* vazio.

4.1.1.2 Replicação do sinal

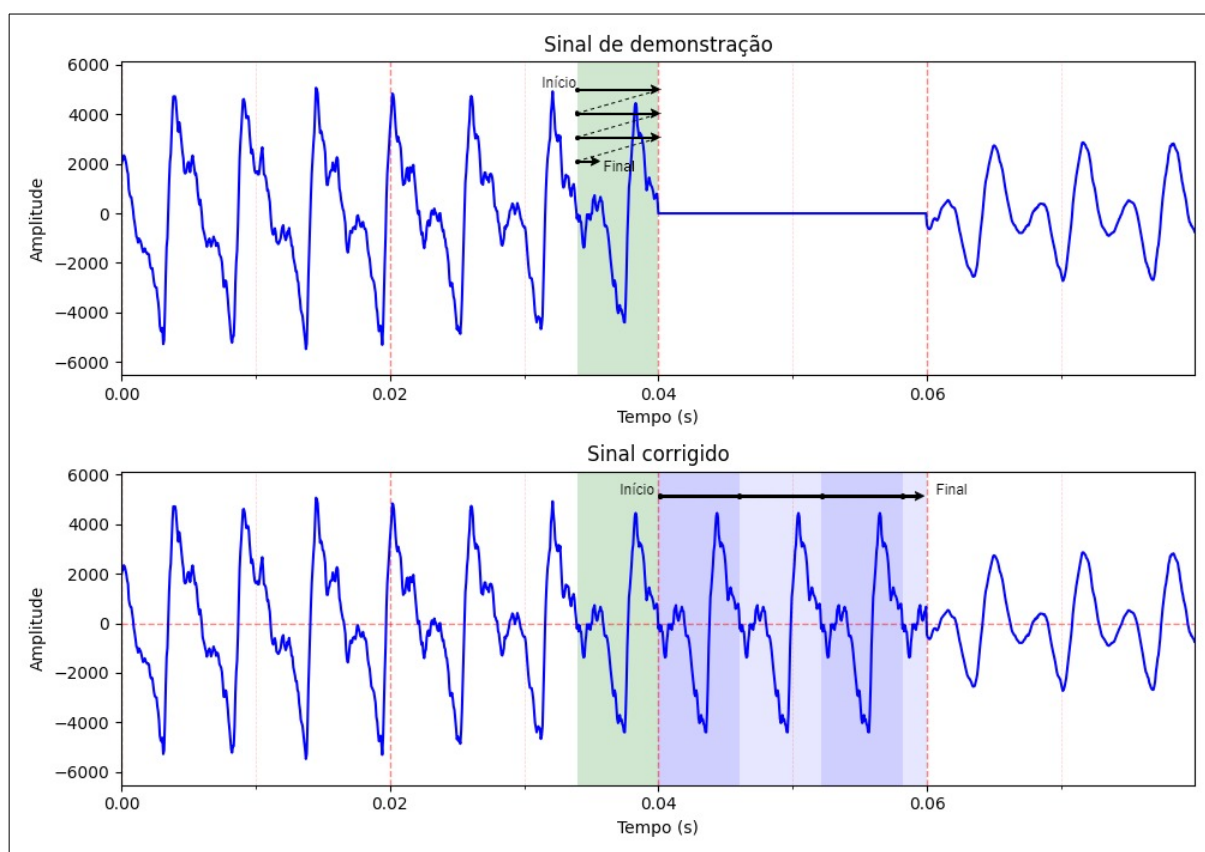
O processo de replicação de período de *pitch* do sinal é executado na PLC do Anexo A do *codec* G.711, pois a ferramenta apresentada pelos autores é baseada em PWR (ITU-T, 1999). Por ser um instrumento de funcionamento em tempo real, processando os pacotes de amostras recebidos, o algoritmo conta com *buffers* para salvar e manipular o sinal no tempo. O documento retrata a replicação do sinal sobre o período vazio e, também, expõe uma etapa de soma sobreposta (*overlap adding* – OLA), utilizada para suavizar as transições entre o *frame* sintetizado e os vizinhos, que é descrita adiante no tópico 2.4.2.

De acordo com ITU-T (1999), o *buffer* de *pitch* é um espaço da memória do equipamento reservado para salvar as últimas amostras do sinal recebido

corretamente quando ocorre uma perda de pacote. O sinal é replicado dentro do *frame* vazio estabelecendo-se um ponteiro no *buffer* de *pitch*, na posição da amostra localizada um período de *pitch* antes do final, percorrendo amostra por amostra em direção ao final do *buffer* e replicando na saída (*output*) cada valor. O *codec* G.711 trabalha com *frames* de 10ms. Segundo os autores, se o período de *pitch* é menor que tal valor, quando o ponteiro ultrapassa o final do *buffer* de *pitch*, sua posição é retomada em um período de *pitch* anterior ao fim do *buffer* antes de continuar a cópia de amostras. Se o período de *pitch* é curto, a região é replicada diversas vezes durante o período de perda de sinal.

A lógica de cópia das amostras no *codec* G.711 pode ser transportada para a análise de um arquivo de áudio, considerando que o processo é feito com o sinal inteiro à disposição, e não com *buffers* específicos. A Figura 21 ilustra a ordem da replicação das amostras.

FIGURA 21 – Ordem de replicação das amostras do período de *pitch* anterior a um picote no espaço de seu tempo



Fonte: Autoria própria (2023).

Como pode se observar a Figura 21, o gráfico superior apresenta o sinal de demonstração, com a região em verde representando o período de *pitch* anterior ao picote; já o gráfico inferior mostra a replicação desse sinal com PWR. Cada região demarcada em azul é a replicação de um período de *pitch*. As flechas do gráfico superior mostram a ordem das amostras levadas ao sinal sintetizado, percorrendo do início ao fim da região verde e retornando diversas vezes. Nota-se, no entanto, que o término da replicação ocorre de forma brusca: no gráfico inferior, visto que o *frame* seguinte não é vazio, a replicação do sinal é interrompida, independente de haver ou não descontinuidade entre os sinais. Para mitigar esta possibilidade, o Anexo A do G.711 (ITU-T, 1999) apresenta a janela de OLA.

4.1.2 Overlap Add – OLA

A operação matemática de OLA utilizada nesse cenário é homônima a uma técnica de processamento digital de DSP utilizada para segmentar sinais e aplicar filtros em trechos mais curtos, sendo uma alternativa à convolução direta de um sinal grande com um filtro (SMITH, 1999).

A partir do Anexo A do *codec* G.711 (1999), é possível entender a OLA utilizada em algoritmos de PLC. Trata-se de um processo de soma de dois segmentos de sinais distintos sobrepostos no tempo, ou seja, que se interseccionam, amostra por amostra, para formar um terceiro. O uso da OLA tem sua razão exposta da seguinte maneira:

É necessário haver uma suavização nas fronteiras do frame perdido, onde há a transição do sinal de voz original para o sintético e vice-versa. Sem a operação de suavização a saída de voz iria conter cliques audíveis, especialmente quando as amostras anteriores não possuem uma periodicidade clara ou quando há uma transição de fonema vocalizado para mudo durante o frame perdido. (GÜNDÜZHAN; MOMTAHAN, 2001, p. 781, tradução nossa).

A respeito da operação de suavização, Gündüzhan e Momtahan se referem ao fato de um sinal conter características diferentes no seu início e seu fim, havendo uma transição suave entre eles.

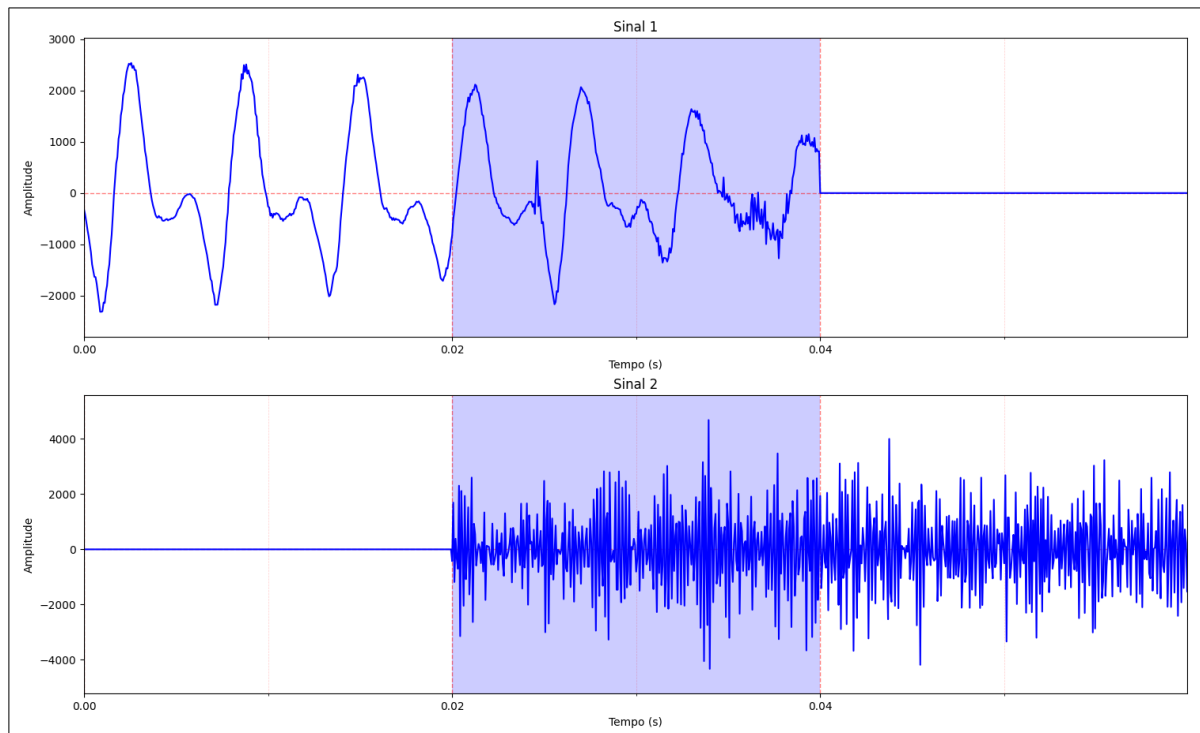
Considera-se uma soma sobreposta simples como a operação vista na Equação 7, em que $y[n]$ é o sinal resultante, $x_1[n]$ e $x_2[n]$ são os segmentos

sobrepostos a serem misturados e N é o comprimento destes em amostras (ITU-T, 1999).

$$y[n] = \sum_{n=0}^{N-1} x_1[n] + x_2[n] \quad (7)$$

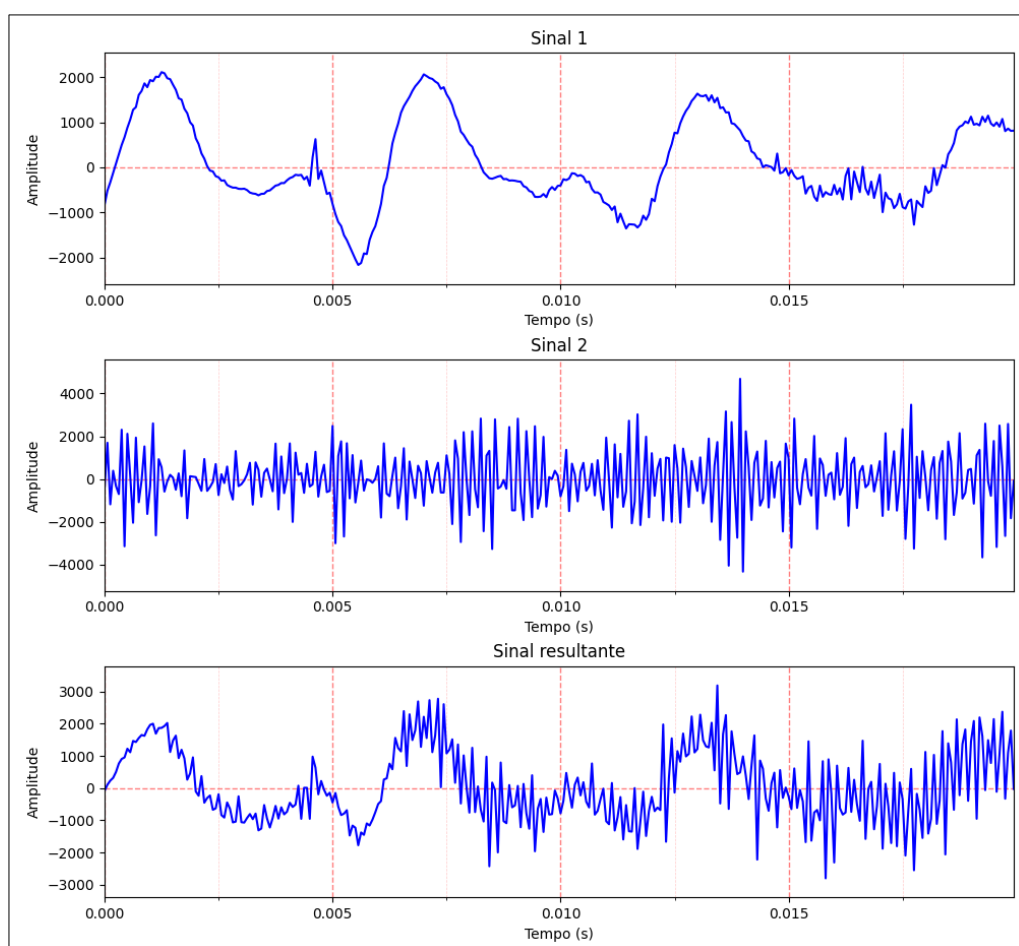
É possível deduzir, a partir da Equação 7, que os sinais $x_1[n]$ e $x_2[n]$ são representados em $y[n]$ da mesma forma para todos os índices n da operação, ou seja, as características individuais de ambos os sinais se manifestam ao longo de todo o período. Se $x_1[n]$ e $x_2[n]$ são sinais com estacionariedade de curto prazo como os de voz, a mera soma deles não acarreta a transição de características de $x_1[n]$ a $x_2[n]$ entre o começo e o final de $y[n]$. Além disso, a amplitude de $y[n]$ em uma OLA simples pode chegar a atingir a soma das amplitudes de $x_1[n]$ e $x_2[n]$. Sendo a OLA utilizada em processamento de áudio, isso pode resultar em uma porção do sinal com nível repentinamente elevado e, portanto, desagradável ao ouvinte.

Para ilustrar a operação de OLA nesse cenário, é proveitoso utilizar dois sinais com comportamentos muito distintos e observar se o resultado traz a transição entre início e final do intervalo proposta por Gündüzhan e Momtahan (2001). Foram escolhidas duas faixas diferentes do mesmo arquivo de áudio visto anteriormente, diferentes das usadas no sinal de demonstração. Retiraram-se duas seções de 40ms com instantes iniciais em 6,22s e 6,28s, denominados sinal 1 e 2, respectivamente, os quais foram arranjados de maneira que houvesse uma intersecção entre a segunda metade do sinal 1 e a primeira metade do sinal 2, com comprimento de um *frame*. A Figura 22 apresenta esse caso. As áreas em azul dos gráficos representam os segmentos sobrepostos em que a OLA é aplicada posteriormente.

FIGURA 22 – Sinais 1 e 2 utilizados na demonstração de OLA

Fonte: Autoria própria (2023).

Os sinais 1 e 2 contam com comportamentos diferentes, sendo o segundo formado por frequências mais altas que o primeiro. Ao aplicar os dois sinais em azul da Figura 22 à equação 7, obtém-se o resultado visto na Figura 23. Os gráficos superior e do meio são os segmentos interseccionados dos sinais 1 e 2, respectivamente; já o gráfico inferior é o sinal resultante. A partir da imagem pode se perceber, no sinal resultante, a presença das frequências baixas e altas dos dois sinais compositores em toda sua duração, o que denota um comportamento homogêneo, não um trânsito entre os diferentes aspectos.

FIGURA 23 – Aplicação de OLA simples sobre sinais distintos

Fonte: Autoria própria (2023).

A fim de se criar uma transição suave entre o começo e o final da região em que a OLA é aplicada, ITU-T (1999) apresenta a possibilidade de janelamento dos sinais. Trata-se da multiplicação das amostras de cada sinal por uma sequência decrescente ou crescente de fatores, de modo que um sinal tenha mais influência sobre o início de $y[n]$ e o outro, sobre o final.

ITU-T (1999) aponta o uso de dois tipos de janelas, triangular, ou de Barlett, e Hanning, ou Hann, indicando o uso da primeira em meios em que complexidade de código é uma questão limitante. Essa recomendação pode ser considerada em sistemas em que não há acesso direto à memória e os índices precisam ser calculados a cada interação. Considerando que isso não é um fator problemático em sistemas modernos, a complexidade da janela não é um fator determinante e ambas podem ser utilizadas a depender da escolha de estratégia. As janelas possuem valores de 0 a 1 e comprimento igual ao dos sinais em que se aplica a OLA.

Embora ITU-T (1999) se refira meramente a janelas triangular e Hanning, as descrições teóricas delas não são as mesmas apresentadas como soluções por esses autores. A janela triangular é descrita conforme a Equação 8 (OPPENHEIM; SCHAFER, 2010), com n indo de 0 ao índice M , de forma que esta possui $M + 1$ elementos. O comprimento da janela é referido a seguir, como M' .

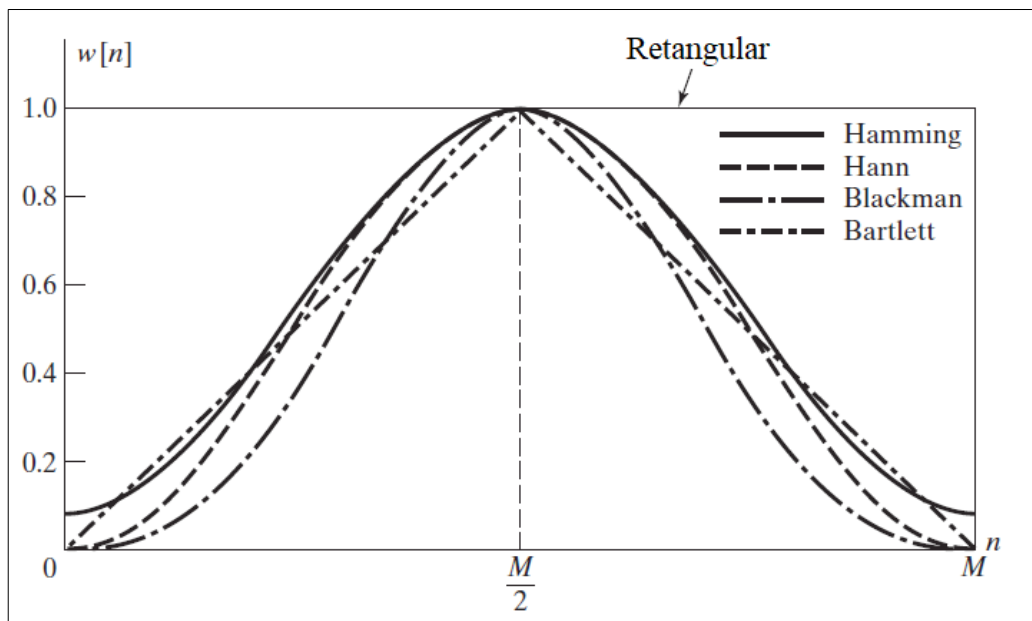
$$w_{\text{Bartlett}}[n] = \begin{cases} \frac{2 \cdot n}{M}, & 0 \leq n \leq M/2 \\ 2 - \frac{2 \cdot n}{M}, & M/2 < n \leq M \end{cases} \quad (8)$$

A janela Hanning é descrita por Lathi (2005) conforme a Equação 9, com as mesmas variáveis da equação anterior.

$$w_{\text{Hanning}}[n] = 0,5 + 0,5 \cdot \cos\left(\frac{2 \cdot \pi \cdot n}{M}\right), \quad 0 \leq n \leq M \quad (9)$$

A Figura 24 apresenta algumas janelas comumente usadas no processamento digital de sinais (OPPENHEIM; SCHAFER, 2010, p. 536). Entre elas estão as janelas Barlett e Hann. A imagem conta com as curvas das janelas traçadas como funções de uma variável contínua, para melhor legibilidade.

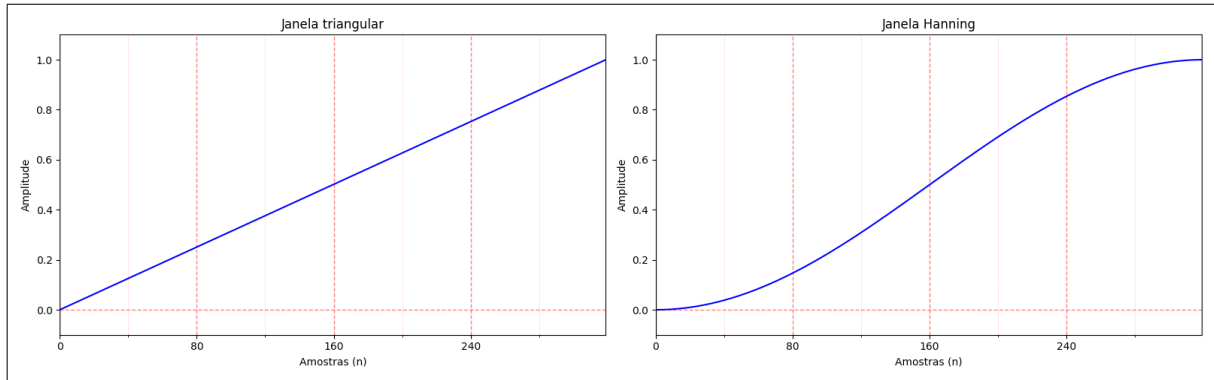
FIGURA 24 – Janelas comumente usadas em processamento digital de sinais



Fonte: Adaptado de Oppenheim e Schafer (2010, p. 536).

Conforme discutido, o janelamento de sinais para OLA pretende dar um formato decrescente e crescente aos sinais $x_1[n]$ e $x_2[n]$, respectivamente. As rampas crescentes podem ser retiradas das janelas triangular e Hanning como apresentadas anteriormente se considerada a forma até seu pico em 1. A partir das equações 8 e 9 é possível concluir que as janelas possuem pico em 1 quando M é par, ou seja, M' é ímpar. Busca-se uma série de valores com mesmo comprimento dos sinais em que a OLA é aplicada, ou seja, igual à variável N da Equação 7. Para garantir o correto tamanho da rampa crescente, com índices inicial e final iguais a 0 e 1, respectivamente, é possível formar uma janela como visto anteriormente com M' igual a $2 \cdot N + 1$ e se retirar os primeiros $M' / 2$ valores, arredondando a divisão para baixo. Para gerar a rampa decrescente, basta inverter a ordem dos índices da sequência obtida. A Figura 25 apresenta as janelas crescentes com formatos triangular à esquerda e Hanning à direita, ambas com comprimento de 320 amostras.

FIGURA 25 – Janelas crescentes triangular e Hanning prontas para aplicação nos sinais operados por OLA



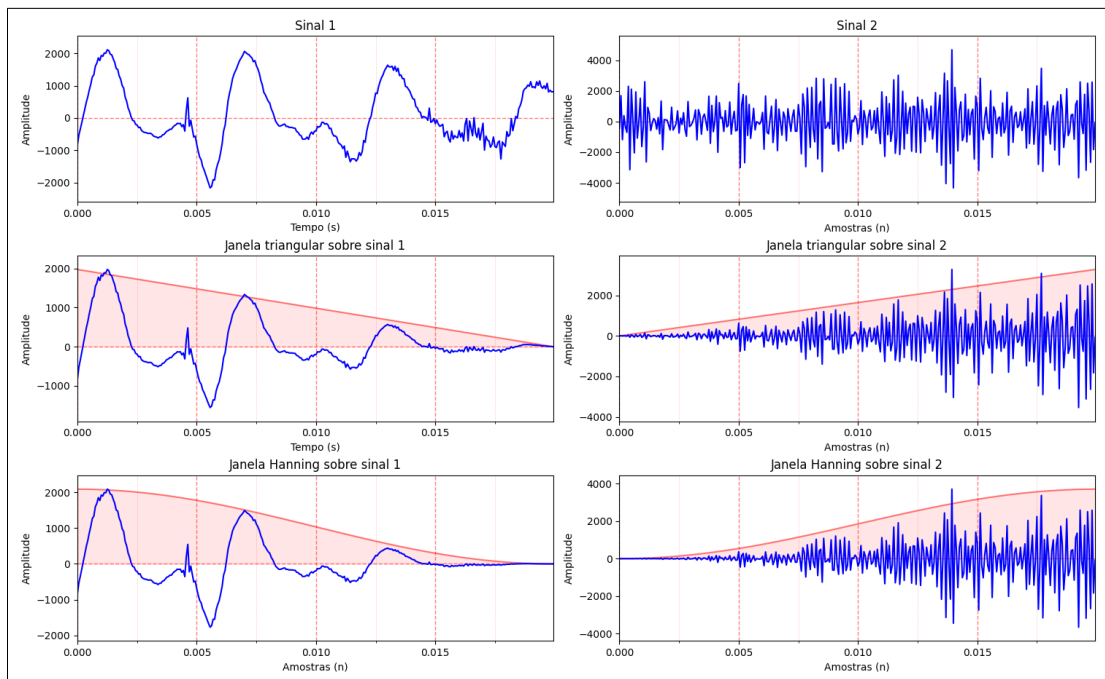
Fonte: Autoria própria (2023).

Tendo posse das janelas triangular ou Hanning expressas na Figura 25, denominando-as $w'_{triangular/Hanning}[n]$, é possível multiplicá-las pelos sinais e obter os sinais janelados, permitindo, assim, uma OLA janelada, como refere-se ITU-T (1999). A Equação 10 apresenta a fórmula da OLA janelada.

$$y[n] = \sum_{n=0}^{N-1} x_1[n] \cdot w'_{triangular/Hanning}[N-1-n] + x_2[n] \cdot w'_{triangular/Hanning}[n] \quad (10)$$

A Figura 26 mostra a aplicação do janelamento crescente e decrescente nos sinais descritos anteriormente. Os gráficos superiores mostram os segmentos interseccionantes dos sinais 1 e 2 à esquerda e à direita, respectivamente. Os gráficos do meio mostram os mesmos sinais quando multiplicados por uma janela triangular. Por fim, os gráficos inferiores repetem esta operação com janela Hanning.

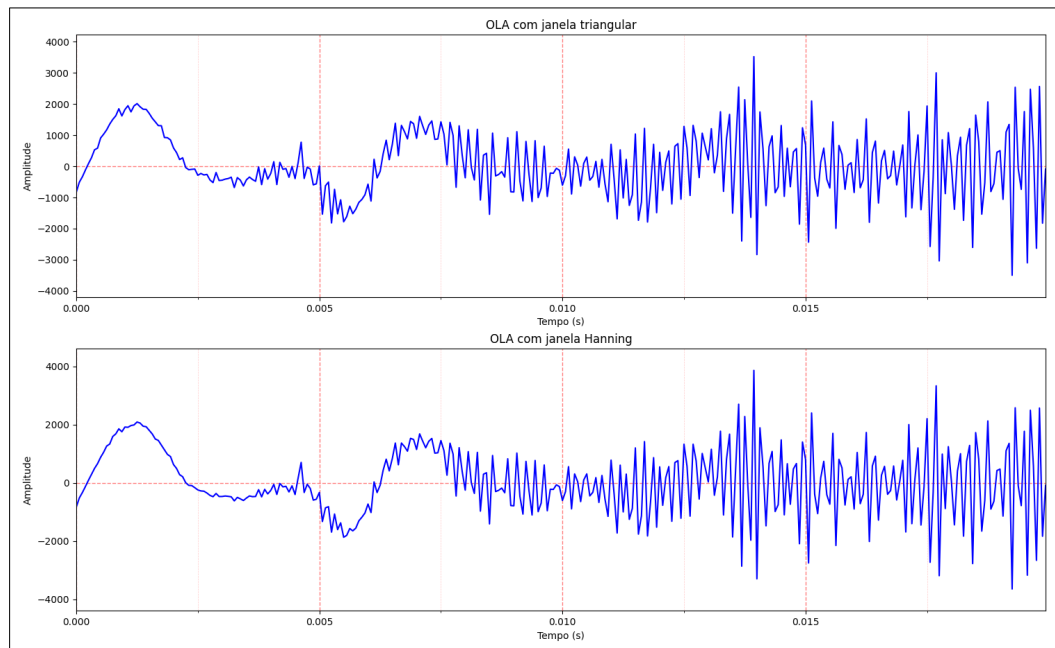
FIGURA 26 – Aplicação de janelas triangular e Hanning sobre segmentos interseccionantes dos sinais 1 e 2



Fonte: Autoria própria (2023).

Ao se executar a operação descrita na Equação 10, obtém-se os resultados vistos na Figura 27. O gráfico superior apresenta a OLA com janela triangular, enquanto o inferior utiliza a janela Hanning. É possível notar que as duas formas de onda são similares e compartilham a característica buscada anteriormente, da transição suave, sem discontinuidades, de formato remetente ao sinal 1 e sinal 2 entre o começo e final do resultado, respectivamente. Em oposição ao visto na Figura 23, percebe-se diminuta presença de altas frequências no começo dos sinais e baixas frequências no final.

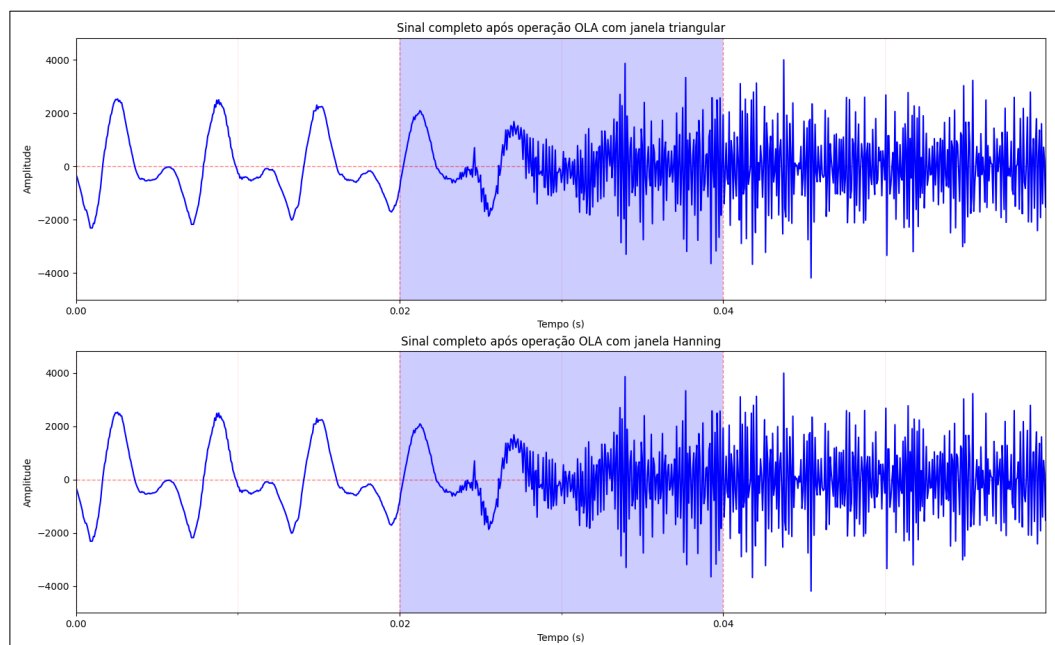
FIGURA 27 – Operação de OLA sobre sinais com janelamento triangular e Hanning



Fonte: Autoria própria (2023).

Para simbolizar a transição entre dois sinais distintos, a Figura 28 apresenta gráficos com o primeiro *frame* do sinal 1 e segundo *frame* do sinal 2 unidos pelos segmentos calculados por OLA, demarcados pelas regiões em azul, com janela triangular no gráfico superior e Hanning no inferior.

FIGURA 28 – Transições entre primeiro *frame* do sinal 1 e segundo *frame* do sinal 2 com segmentos calculados por OLA com janelamento triangular e Hanning



Fonte: Autoria própria (2023).

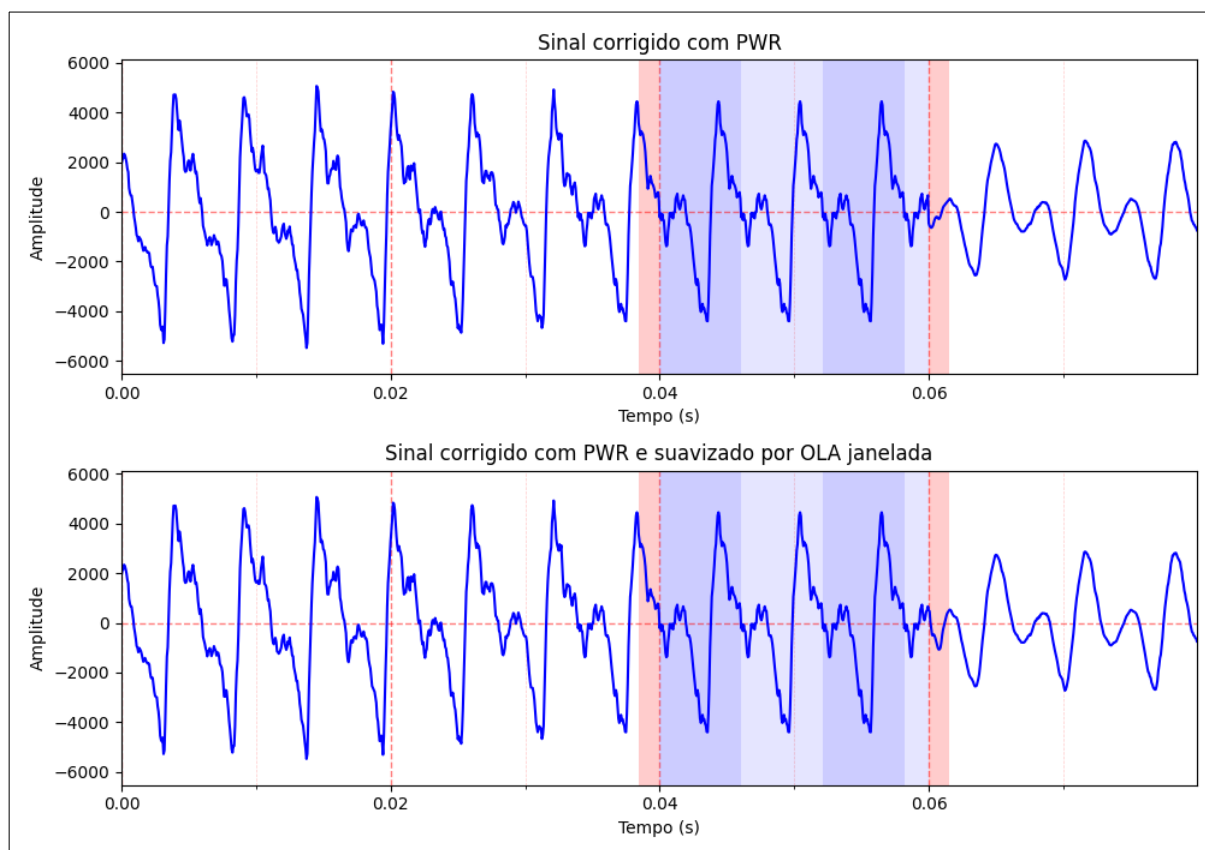
Com base nas ondas demonstradas na Figura 28, é possível destacar a modificação de comportamento dos sinais de forma suave e sem descontinuidades entre o primeiro e último *frame*.

O algoritmo de PLC apresentado no Anexo A do *codec* G.711 (ITU-T, 1999) é baseado na técnica de PWR para preencher o espaço deixado pelo picote; porém, como discutido no tópico anterior, esse método traz o revés de gerar uma descontinuidade entre o fim de um *frame* sintetizado e o início do próximo *frame* recebido corretamente. ITU-T descreve a aplicação de uma etapa de OLA no algoritmo para contornar esse problema, além de também corrigir alguma possível deformidade na transição entre o *frame* anterior e o gerado sinteticamente.

De acordo com ITU-T (1999), o método de PLC apresentado aplica uma operação de OLA janelada nas regiões imediatamente anterior e posterior ao sinal gerado por PWR com comprimento igual a um quarto do período de *pitch*. A operação utiliza como primeiro sinal entrante (sinal 1) o segmento do sinal original, recebido devidamente, referente ao intervalo de tempo em questão. O sinal 2 é proveniente de uma extensão do sinal gerado por PWR. Para isso, o processo descrito no tópico 2.4.1.2 é realizado considerando não apenas a extensão do *frame* perdido, mas essa medida somada a um quarto de período de *pitch*. Tal fragmento adicional não é levado ao *frame* picotado, mas mantido salvo para aplicação na OLA do segmento posterior ao picote. Em conjunto, salva-se também o quarto final do período de *pitch* como uma extensão retroativa ao começo do sinal sintético, o que é utilizado como segundo sinal entrante para a OLA janelada da região anterior ao picote. Os dois sinais descritos são gerados de tais formas para que estejam em fase com o sinal inserido na região do picote.

A Figura 29 ilustra tal modificação sobre o algoritmo de PWR apresentado anteriormente, utilizando o mesmo sinal de demonstração da Figura 21. O gráfico superior mostra o mesmo resultado visto na figura citada, sem a suavização das bordas, enquanto o inferior conta com o processo de OLA janelada – nessa etapa, utilizou-se janela do tipo Hanning. As regiões em vermelho são as áreas onde foi aplicada individualmente a OLA janelada.

FIGURA 29 – Sinais com PLC por método de PWR puro e por PWR com processo de OLA nas fronteiras anterior e posterior ao picote



Fonte: Autoria própria (2023).

Como é possível perceber na Figura 29, cada janela vermelha possui 22 amostras de comprimento, o que se refere a um quarto do período de *pitch* visto no tópico anterior deste trabalho. A primeira janela, à esquerda, não apresenta mudanças visíveis, pois, como visto anteriormente, a transição entre um *frame* recebido corretamente e um sintético se dá de maneira suave com o emprego de PWR. No período da segunda janela, no entanto, é possível notar diferença entre os dois sinais apresentados, aproximando o formato de onda da região dos períodos seguintes.

4.1.3 Linear Prediction Coding – LPC

Elsabrouty, Bouchard e Aboulnasr (2004) citam a existência de um método de PLC presente no Apêndice B do padrão ANSI T1-521-2000 baseado em modelo de predição linear para estimar o sinal de voz faltante. De acordo com Rabiner e Schafer (1978), a análise de predição linear – ou codificação por predição

linear (*Linear Prediction Coding* – LPC) – é uma técnica voltada a extrair parâmetros básicos de voz, como *pitch*, frequências formantes, espectro e funções de área do trato vocal, além de representar sinais de voz em transmissões ou armazenamento que requeiram *bit rate* baixo. Ainda, conforme estes autores (1978, p. 396, tradução nossa), “a ideia básica por trás da análise de predição linear é de que uma amostra de voz pode ser aproximada como uma combinação linear de amostras anteriores.”.

O trabalho com predição linear envolve a determinação de coeficientes para um filtro linear, o que requer a solução de uma série de equações lineares com simetrias específicas (PROAKIS; MANOLAKIS, 1996). São métodos reconhecidos de solução desse tipo de sistema os algoritmos *Levinson-Durbin* e *Schur*. A biblioteca *Python Librosa* (MCFEE *et al.*, 2015) oferece uma função para estimação de coeficientes de predição linear, “*librosa.lpc*”, a qual é baseada no método de *Burg* para a obtenção dos valores; segundo os autores, esse método é uma extensão do método *Yule-Walker*. De acordo com Marple (1980, p. 441, tradução nossa), o algoritmo de *Burg* “[...] utiliza uma estimação de mínimos quadrados restrita para obter os M parâmetros de autorregressão estimados de N amostras de dados. A restrição requer que os parâmetros de autorregressão estimados satisfaçam a recursão de Levinson.”.

A função fornecida pela biblioteca *Python Librosa* tem como parâmetros a ordem para o filtro, simbolizada na declaração de Marple pela variável M , e uma sequência de amostras a se usar de base para o cálculo, cujo comprimento é representado pela variável N . Para este estudo, utilizou-se como janela de cálculo o *frame* anterior completo, que consiste em 320 amostras para um sinal dividido em frames de 20ms com frequência de amostragem de 16 kHz. Elsabrouty, Bouchard e Aboulnasr (2004) manifestam que para a análise de sinais do tamanho deste *frame* o uso de 50 coeficientes se mostra suficiente.

A determinação de uma amostra predita do sinal $s[n]$ a partir de um sinal de excitação $u[n]$ ocorre conforme a Equação 11 (RABINER; SCHAFER, 1978, p. 398), em que a_k e G são, respectivamente, os coeficientes e o ganho do filtro de predição linear, adquiridos por meio de algoritmos citados anteriormente sobre as amostras anteriores à predita. A variável M indica a ordem do filtro, ou seja, a quantidade de coeficientes.

$$s[n] = \sum_{k=1}^M a_k \cdot s[n-k] + G \cdot u[n] \quad (11)$$

Segundo Rabiner e Schafer (1978), tal sistema é excitado por um trem de impulsos para fonemas vocalizados e por uma sequência de ruído aleatório para fonemas mudos. Elsabrouty, Bouchard e Aboulnasr (2004) sugerem, no entanto, uma outra abordagem. Os autores citam os elementos $s[n-k]$ e $u[n]$ da Equação X(anterior) como termos previsível e desconhecido, respectivamente. O primeiro se trata do sinal recebido antes do picote. O segundo termo é referenciado também como sinal residual.

Considerando que o sinal residual não é conhecido no *codec* receptor, Elsabrouty, Bouchard e Aboulnasr (2004) afirmam que uma boa alternativa para o sinal de excitação pode ser utilizar uma pequena porcentagem de sinal obtido por PWR do *frame* anterior, citando a técnica abordada no tópico 2.4.1 deste trabalho. Dessa forma, Elsabrouty, Bouchard e Aboulnasr (2004, p. 665) reescrevem a Equação 11 conforme a Equação 12, sendo $S[n]$ o sinal previsto, $\hat{S}[n]$ o sinal replicado do frame anterior por PWR e a_k os coeficientes de predição linear. Os autores sugerem aplicar o valor de 0,01 a G nesse cenário.

$$S[n] = \sum_{k=1}^P a_k \cdot S[n-k] + \hat{S}[n] \cdot G \quad (12)$$

Por fim, os mesmos autores propõem mais uma modificação para a geração de sinal com uso de LPC e indicam uma soma ponderada entre a amostra calculada por LPC, vista na Equação 12, e uma amostra do sinal obtido por PWR do *frame* anterior no respectivo tempo. O sinal visto na Equação 12 passa a ser chamado de $S_1[n]$, enquanto repete-se $\hat{S}[n]$ como o sinal replicado por PWR. Para a soma ponderada, utilizam-se as variáveis α e β , que devem somar 1 e multiplicam $S_1[n]$ e $\hat{S}[n]$, respectivamente. Obtém-se o sinal $S[n]$, a ser inserido no lugar do picote, como na Equação 13 (ELSABROUTY; BOUCHARD; ABOULNASR, 2004, p. 665). Os autores recomendam a α e β os valores de 0,7 e 0,3 respectivamente.

$$S[n] = \alpha \cdot S_1[n] + \beta \cdot \hat{S}[n] \quad (13)$$

Apesar de Elsabrouty, Bouchard e Aboulnasr (2004) e Rabiner e Schafer (1978) apontarem as formas de se calcular um sinal por LPC conforme as equações dadas até aqui, outros autores oferecem uma mudança que se mostrou significativa no decorrer do projeto. Toyoshima e Shimamura (2014), por exemplo, descrevem a predição de uma amostra a partir das conhecidas anteriores conforme a Equação 14, sendo $\hat{x}_n$ a amostra n prevista, M a ordem da predição linear, a_k os coeficientes de predição linear e $x[n-k]$ as amostras anteriores a n .

$$\hat{x}_n = - \sum_{k=1}^M a_k \cdot x[n-k] \quad (14)$$

Observa-se, na Equação 14, a negatividade do termo somatório. Durante a implementação da PLC com predição linear em *Python*, notou-se a necessidade de agregar essa característica às equações anteriores. Essa visão é corroborada por Kondo e Nakagawa (2006) e Collomb (2009). As Equações 13 e 14 foram adaptadas conforme as Equações 15 e 16, respectivamente, em que $\hat{S}[n]$ é o sinal replicado por PWR do frame anterior, $S_1[n]$ é a amostra prevista puramente por LPC, M a ordem da predição linear, $S[n-k]$ é o sinal anterior ao previsto, G , α e β são constantes com valores recomendados pelos autores e $S[n]$ é o valor final da amostra a se inserir no sinal de áudio.

$$S_1[n] = - \sum_{k=1}^M a_k \cdot S[n-k] + \hat{S}[n] \cdot G \quad (15)$$

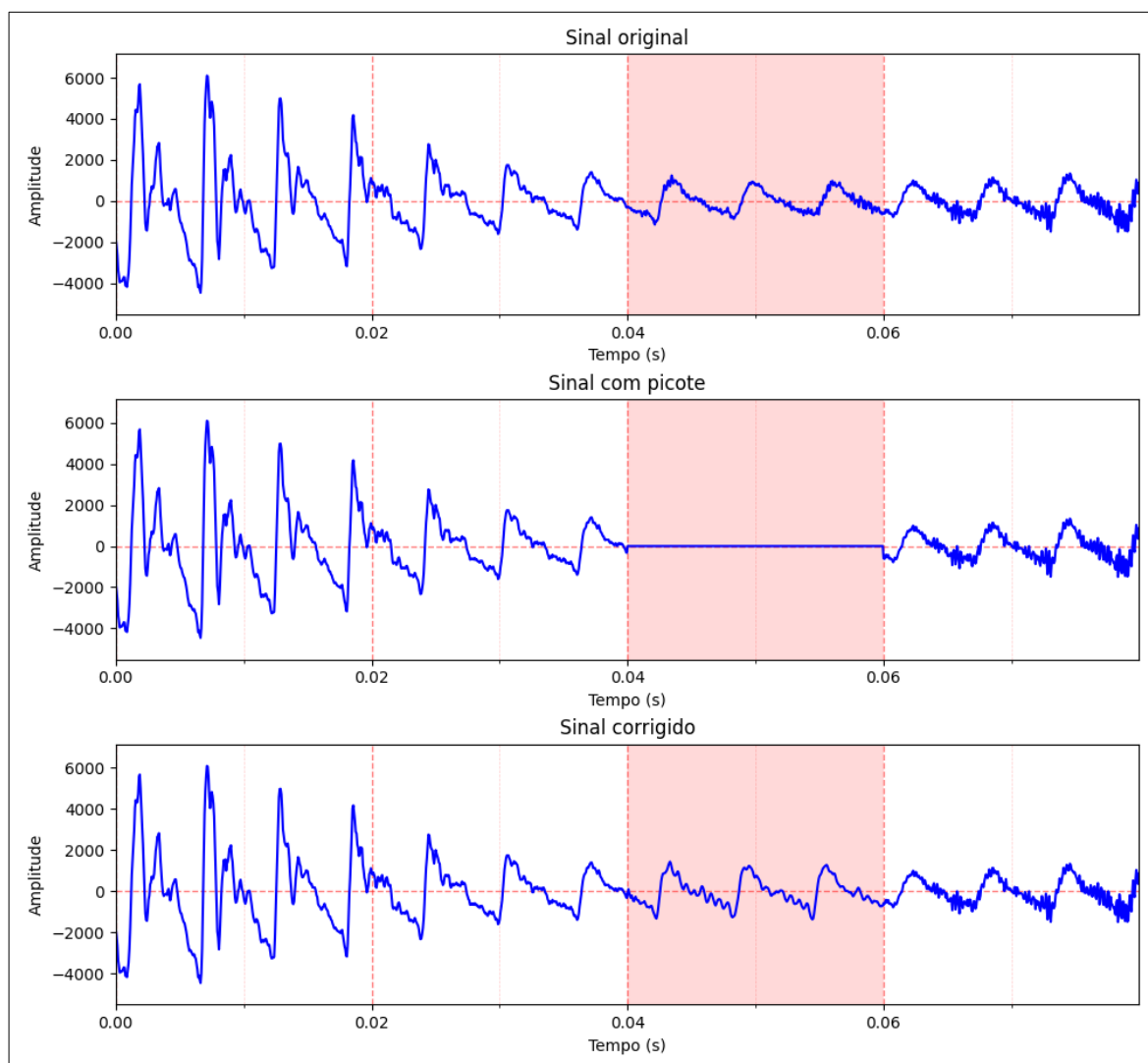
$$S[n] = \alpha \cdot S_1[n] + \beta \cdot \hat{S}[n] \quad (16)$$

A geração de amostras para substituição do picote parte da primeira após a perda de sinal, ou seja, nessa situação todas as anteriores devem ser provenientes do sinal recebido corretamente. As amostras seguintes à primeira invariavelmente teriam algum espaço entre elas e a última amostra real, intervalo este que é preenchido pelas amostras previstas anteriormente. Kondo e Nakagawa (2006) esclarecem que se deve utilizar as próprias amostras previstas como parte do sinal anterior, mantendo os mesmos coeficientes a_i .

Demonstra-se, a seguir, a geração de sinais com o método discutido. São utilizados segmentos de sinal do mesmo arquivo de áudio dos exemplos anteriores.

A Figura 30 apresenta o resultado de LPC para a perda de um *frame* com sinal vocalizado, cujo segmento apresentado se encontra de 5,72s a 5,8s no arquivo de áudio. O gráfico superior apresenta o segmento original, enquanto o do meio tem o terceiro *frame* levado a zero, representando o picote, e o terceiro conta com a correção desse problema. As áreas em vermelho destacam a região do picote.

FIGURA 30 – Reconstrução de *frame* perdido de sinal vocalizado com método de LPC discutido

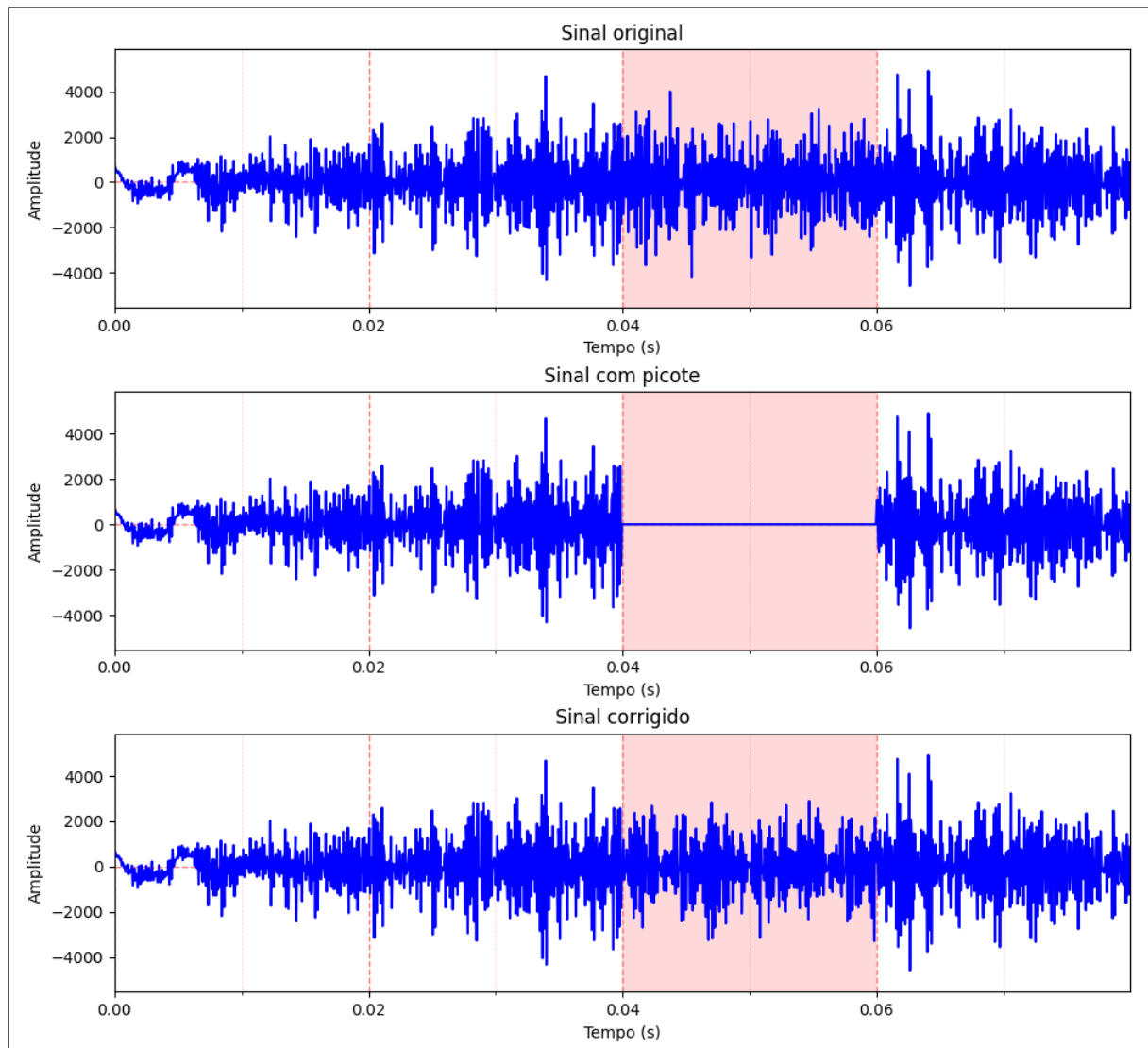


Fonte: Autoria própria (2023).

Em contraste ao método de PWR, observa-se na Figura 30 que o LPC apresenta repetições do período de *pitch* não idênticas.

A Figura 31 conta com a mesma estrutura da Figura 30. Neste caso, é apresentado o resultado de correção para um segmento de fala não-vocalizada, retirado do trecho de 6,26s a 6,34s do arquivo utilizado.

FIGURA 31 – Reconstrução de *frame* perdido de sinal não-vocalizado com método de LPC discutido



Fonte: Autoria própria (2023).

4.1.4 LPC com sentido inverso

Um algoritmo adicional surge com mesma base de processamento de sinal que o anterior. Trata-se da implementação da equação 16 que reconstrói um sinal com a mescla de sinais sintetizados por LPC e PWR (ELSABROUTY; BOUCHARD; ABOULNASR, 2004). Altera-se nesse caso o sentido temporal em que

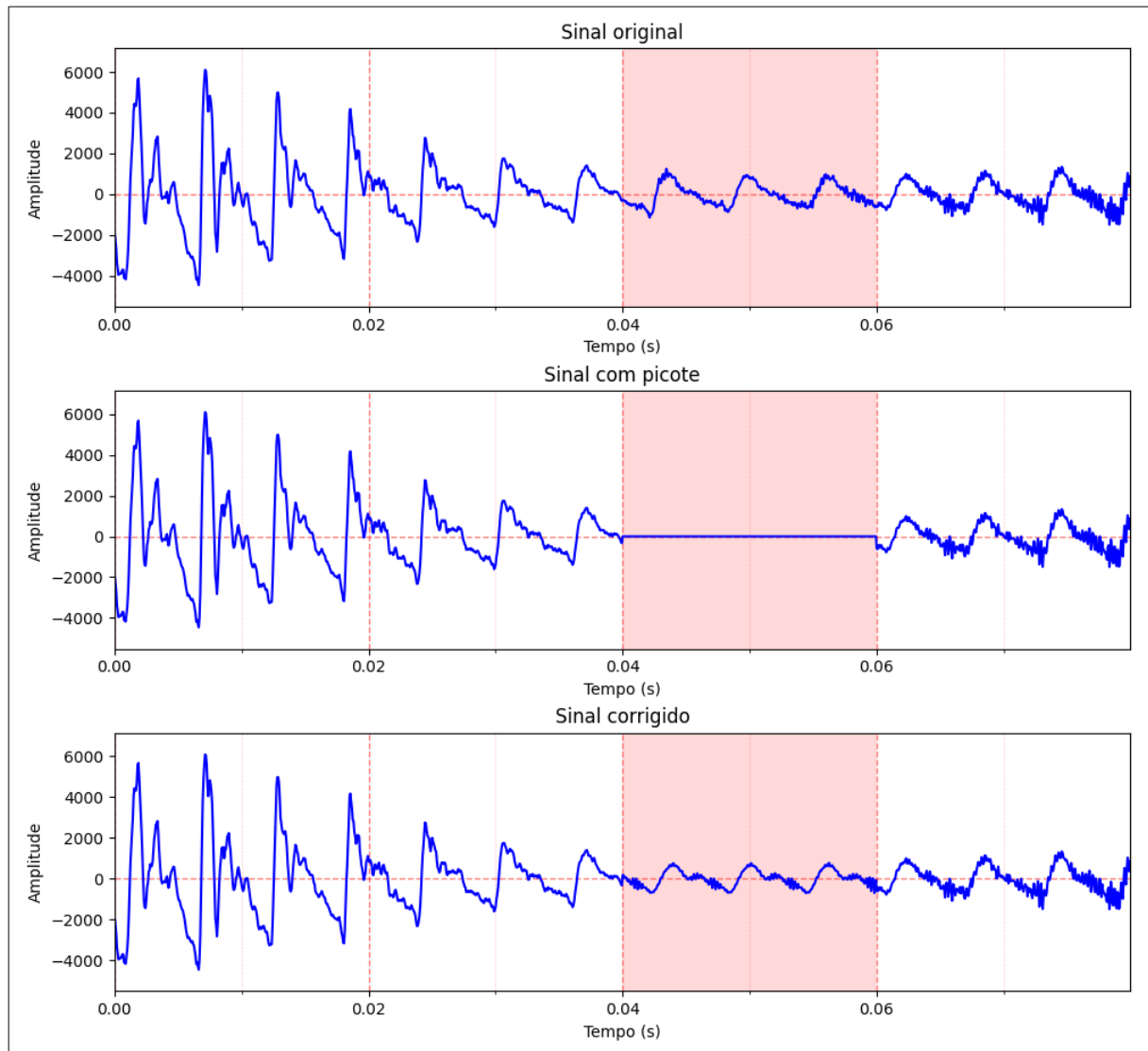
a função de PLC gera um sinal novo a partir de *frame* vizinho: a reconstrução é baseada no *frame* seguinte ao picote, ao invés do anterior como tem sido feito até esse ponto.

Esta estratégia é proposta por diversos autores no contexto de algoritmos de PLC com LPC bilateral, que é tratado no subcapítulo seguinte. Xie, Wei e Chen (2016) comentam que métodos de reconstrução de sinal clássicos como PWR e LPC desconsideram a informação contida no pacote subsequente ao perdido. Segundo os autores, no caso do processamento em tempo real de dispositivos VoIP, o armazenamento do sinal recebido em um *buffer* permite que se aguarde a chegada de pacotes posteriores a um picote para realizar a correção. A ideia pode ser estendida ao caso do processamento de um arquivo de áudio inteiro, visto que todo o sinal está disponível ao tempo de execução da PLC.

Métodos de predição linear bilateral como o proposto por Kondo e Nakagawa (2006) contam com uma seção de processamento “para frente” (*forward*) e uma “para trás” (*backward*). O primeiro método, *forward*, se refere ao utilizado no subcapítulo anterior, que gera sinal “para frente” a partir do *frame* anterior. O segundo, *backward*, retratado nesse subcapítulo, obtém o frame posterior ao picote e gera o sinal reconstruído da última amostra do picote em direção à primeira.

De acordo com Xie, Wei e Chen (2016), o LPC “*backward*” tem a mesma base matemática que o “*forward*”, alterando-se apenas as funções de detecção de *pitch* e LPC para partirem da amostra final do *frame* perdido e considerarem amostras anteriores como as posicionadas na direção positiva do eixo do tempo. Observa-se na Figura 32 uma situação semelhante à da Figura 30, com mesmo intervalo de sinal vocalizado e organização gráfica. O terceiro *frame* do gráfico inferior conta com a geração por meio de LPC *backward*, sendo notável a semelhança do trecho com o *frame* seguinte.

FIGURA 32 – Reconstrução de *frame* perdido de sinal vocalizado com método de LPC *backward*



Fonte: Autoria própria (2023).

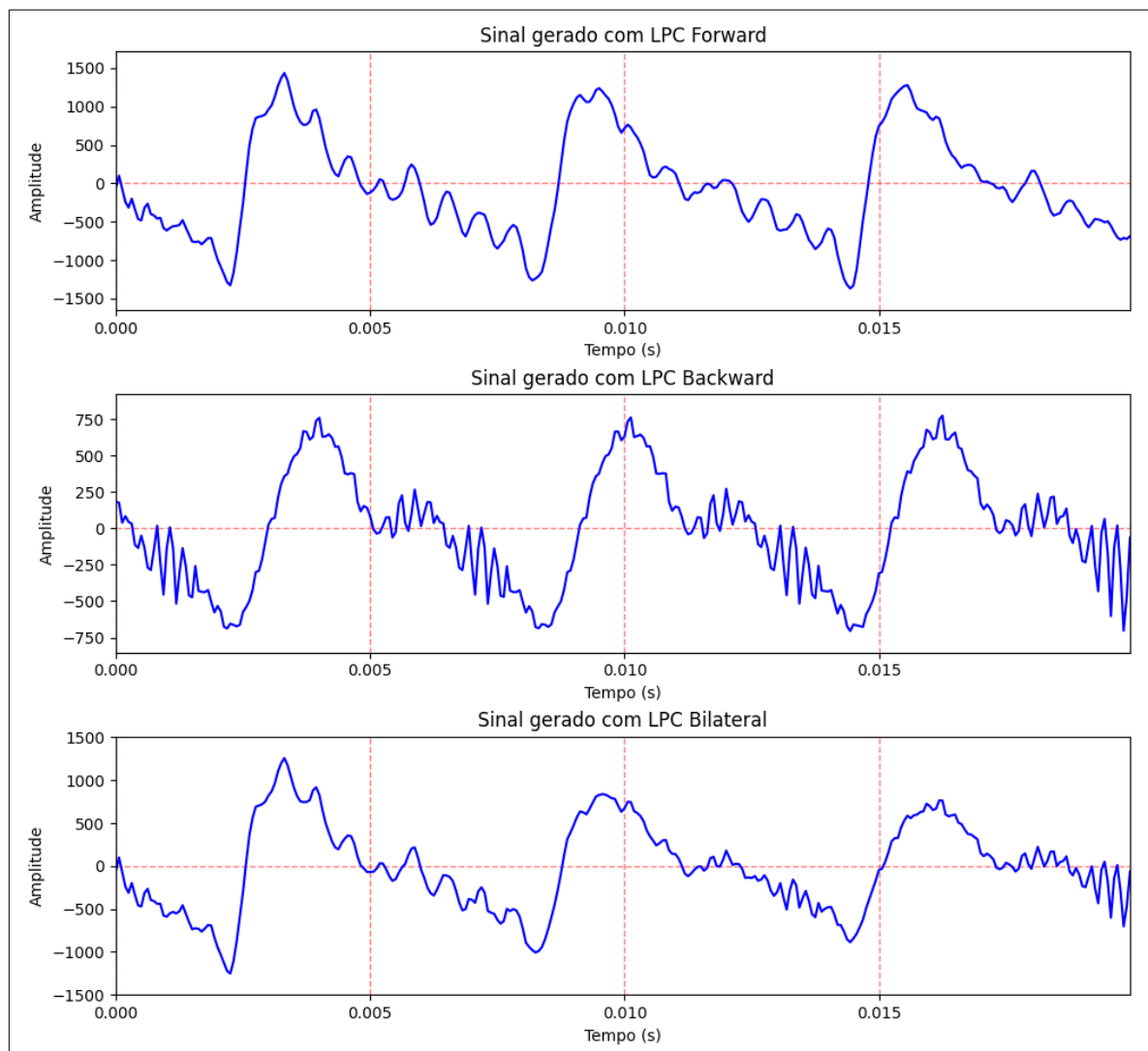
4.1.5 LPC bilateral

Como visto no tópico anterior, o modelo de LPC *backward* surge para complementar o funcionamento do LPC *forward*, que utiliza apenas a informação anterior ao picote para gerar sinal sintético. Kondo e Nakagawa (2006) propõem a união entre os dois processamentos, *forward* e *backward*, em que dois sinais são sintetizados para o intervalo do picote e interpolados, para que ocorra a transição suave entre o começo do *frame* com sinal *forward* e final com sinal *backward*. Xie, Wei e Chen (2016) levantam um método similar, esclarecendo que esta transição é

realizada por meio de uma OLA janelada e sugerindo que a janela utilizada seja triangular.

Tendo em vista que a concepção de um *frame* por LPC bilateral é dada pela OLA janelada entre o sinal gerado por LPC *forward* e o gerado por LPC *backward* no frame perdido, a Figura 33 compara os trechos sintetizados em cada caso exemplificado anteriormente, vistos nas regiões marcadas em vermelho dos gráficos inferiores das Figuras 30 e 32. Os gráficos superior e central apresentam os sinais gerados por LPC *forward* e *backward*, respectivamente, enquanto o inferior ilustra o resultado de uma operação de OLA janelada entre eles, o que caracteriza a sintetização do sinal por método de LPC bilateral.

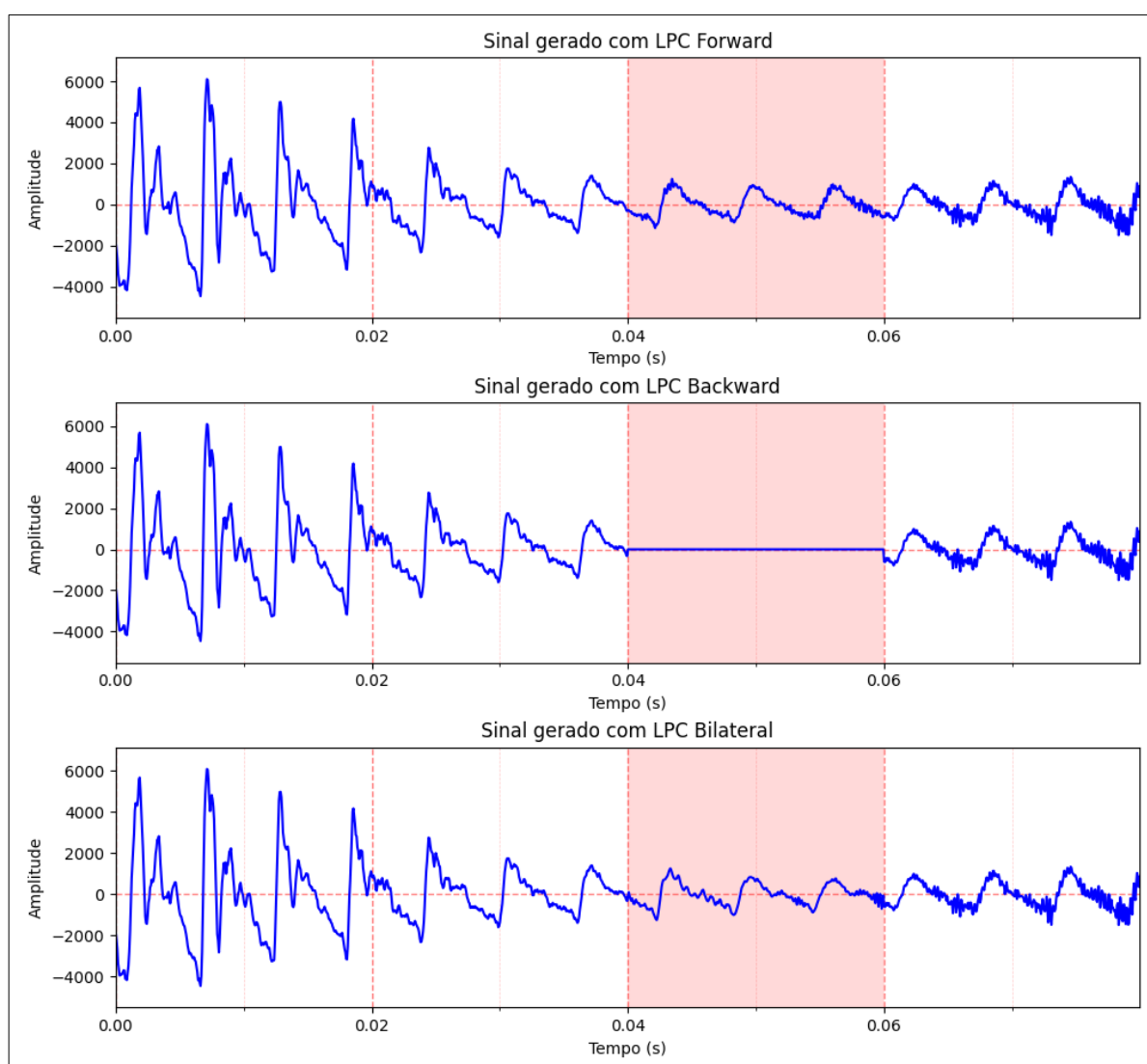
FIGURA 33 – Comparação entre *frames* sintetizados por meio de LPC *forward*, *backward* e bilateral



Fonte: Autoria própria (2023).

O sinal resultante visto na Figura 33 pode ser aplicado em um *frame* perdido, de forma a reconstituir o trecho silencioso, com características sonoras que transitam da similaridade com o *frame* anterior à com o posterior. A Figura 34 exibe a reconstrução de sinal por LPC bilateral no sinal de demonstração dos exemplos anteriores, com sinal original no gráfico superior, sinal acometido por picote no gráfico central e corrigido no gráfico inferior.

FIGURA 34 – Reconstrução de *frame* perdido de sinal vocalizado com método de LPC bilateral



Fonte: Autoria própria (2023).

4.2 Ferramentas de análise de áudio

A primeira ferramenta de análise de áudio determinada para o uso nesta pesquisa foi a API de biometria de voz da empresa Khomp, por contemplar uma

aplicação prática ao tema, fornecer uma estrutura de testes rica e já estar envolvida com os correntes trabalhos do autor. Esse instrumento é utilizado nos produtos de controle de acesso da empresa, como centrais de interfonos VoIP que abrem portas ao comando de voz do usuário (KHOMP INDÚSTRIA E COMÉRCIO LTDA., 2022).

A API de biometria foi acessada por meio de rotinas de código próprias da empresa. Seu uso é compreendido por duas etapas, sendo elas o treinamento de perfis e a verificação de identidade. Para cada teste desta pesquisa, a API foi reiniciada (para exclusão de qualquer perfil previamente registrado) e executou-se a criação e treinamento de perfil para cada um dos quatro locutores. Cada treinamento utilizou os três primeiros arquivos de áudio do respectivo locutor. Em seguida, cada um dos vinte e um áudios restantes de cada locutor foi checado com a API para verificar a confiança de identidade com cada um dos quatro perfis criados, tanto do próprio locutor quanto dos demais. Cada verificação resultou em um índice de confiança de 0 a 1, sendo 0 a confiança mínima, com menor probabilidade de tal áudio pertencer ao locutor indicado, e 1 a confiança máxima, que indicava alta probabilidade de a voz pertencer a tal locutor.

Conforme documentação interna da API de biometria de voz, verificou-se a necessidade de se operar com arquivos de áudio de extensão “wav” monocanais, com compressão *PCM 16-bit* e frequência de amostragem igual a 16 kHz.

A ferramenta de transcrição de áudio utilizada não é diretamente relacionada a alguma aplicação profissional como a anteriormente vista. Fez-se uso de bibliotecas *Python* específicas para transcrição e medição dos resultados. Foi utilizada a biblioteca *SpeechRecognition*, que oferece acesso a diversas APIs e métodos, ou *engines*, de transcrição, tanto *online* quanto *offline* (SPEECHRECOGNITION, 2022). Foi escolhida a *engine* de reconhecimento de fala CMU *Sphinx/PocketSphinx*, por funcionar de maneira *offline*, baseada em linguagem C e oferecer um conjunto de modelos acústicos prontos para a transcrição (CMUSPHINX/POCKETSPHINX, 2022).

A *engine* utilizada demanda arquivos de áudio de extensão “wav” monocanais, com compressão *PCM 16-bit* (CMUSPHINX/POCKETSPHINX, 2022) e, conforme o código em *Python* da biblioteca *SpeechRecognition* (SPEECHRECOGNITION, 2022), com taxa de amostragem de 16 kHz, o que vai ao encontro dos requisitos da ferramenta de biometria de voz.

Para os testes relativos à transcrição de texto, foi utilizada a medição de taxa de erro de caracteres (*Character Error Rate* – CER), uma variação da taxa de erro de palavras (*Word Error Rate* – WER). Segundo Young e Chase (1998), havendo uma sequência escrita *REF* contendo n palavras e uma sequência escrita hipotética *HYP*, proveniente de transcrição de áudio, a WER é encontrada ao se alinhar as duas sequências e se contar a quantidade de modificações entre elas. Tais modificações seguem o cálculo de distância de Levenshtein (SEYFARTH; ZHAO, 2020), que considera a substituição (*S*), deleção (*D*) ou inserção (*I*) de uma palavra na sequência hipotética em relação à de referência. A Equação 17 apresenta o cálculo de WER nesse cenário (YOUNG; CHASE, 1998).

$$WER = \frac{S+D+I}{n} * 100\% \quad (17)$$

A porcentagem de WER indica a quantidade de diferenças entre a frase hipotética e a de referência, ou seja, um número mais alto demonstra que a transcrição conteve mais erros. A medida de CER ocorre da mesma forma que a WER; porém, deve-se considerar os caracteres individuais ao invés de palavras inteiras separadas por espaços (YOUNG; CHASE, 1998).

Considerou-se como frase de referência para o cálculo a expressão obtida pela transcrição de cada áudio do *dataset* de referência. Fez-se a medida para todos os demais *datasets*, com picote e os corrigidos, para analisar como se modifica a transcrição com cada técnica. Foi utilizada a biblioteca *Python JiWER* para o rápido cálculo das taxas de erro, que integra um módulo de cálculo de distância de Levenshtein em Cython (JIWER, 2022).

4.3 Conjunto de arquivos de áudio

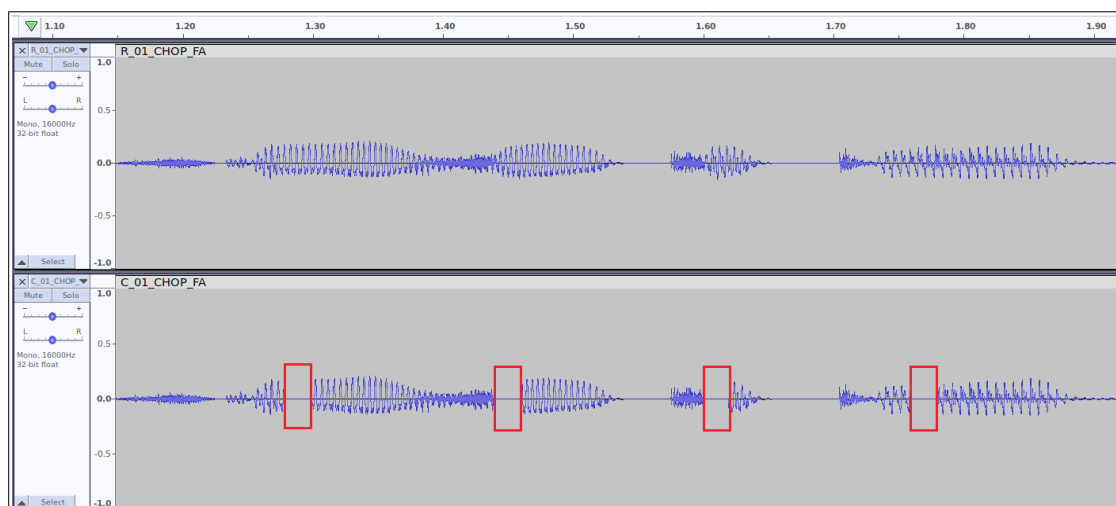
Um teste rigoroso das ferramentas e algoritmos utilizados requer uma base de dados coesa, representativa e bem documentada. Para isso, buscou-se um *dataset* com diversas características em relação ao formato dos arquivos e ao conteúdo audível, capaz de favorecer os testes pretendidos com o mínimo de desvio – além de estar disponível livremente.

Neste trabalho, o formato de um arquivo de áudio se refere a um tipo de compressão do sinal, extensão de arquivo, quantidade de canais e frequência de amostragem. As restrições de formato são determinadas pela possibilidade de uso dos áudios nas ferramentas de análise de voz, sobretudo a biometria de voz da empresa Khomp, por ser mais restritiva. Esperava-se que o *dataset* escolhido fornecesse sinais com 1 canal, frequência de amostragem igual a 16 kHz e compressão *PCM 16-bit* em arquivos “wav”.

Em termos do conteúdo dos áudios, as ferramentas de análise se dividem. Para retornar um resultado, o *software* de transcrição de áudio necessita apenas da gravação de um locutor pronunciando algum conjunto de palavras. De outra forma, a biometria de voz trabalha com a identidade dos locutores; portanto, é necessário haver múltiplas gravações de locutores diversos e previamente identificados, para que seja feito o treinamento do perfil de cada um e obtenção das porcentagens de confiança de locutor igual ou diferente.

Por questão de definição de projeto, as duas ferramentas compartilham uma característica de conteúdo imprescindível nos arquivos: foi necessário haver, para cada gravação de voz, uma versão original (sem adição de distorção) e uma acometida por picotes, cujas regiões não modificadas fossem idênticas, para não haver diferenças nos processamentos de áudio, além dos planejados para o trabalho. A Figura 35 ilustra um exemplo de par de sinais buscados, cujo gráfico superior mostra um segmento de arquivo de áudio original e o inferior apresenta o mesmo trecho com a presença de picotes nas regiões destacadas em vermelho.

FIGURA 35 – Trechos de áudios buscados em versão original e com presença de picotes



Fonte: Autoria própria (2023).

Foi obtido o conjunto TCD-VoIP, desenvolvido pelos autores Harte, Gillen e Hines (2015). O *dataset* foi criado com foco em testes de identificação de qualidade de voz, fornecendo uma base rica em diferentes problemas como picote, saturação e eco, além de ter passado por um processo de avaliação por ouvintes e estabelecimento de um placar de pontuação média de opinião (*Mean Opinion Score* – MOS) – esse último quesito é essencial no desenvolvimento de aplicações de predição automática de MOS em arquivos de áudio, porém não foi utilizado neste trabalho. O conjunto de arquivos foi disponibilizado pelo Grupo de Pesquisa *Sigmedia* da *Trinity College Dublin* no endereço eletrônico “https://sigmedia.github.io/resources/dataset/tcd_voip_ltd/”.

Os principais pontos considerados na escolha deste *dataset* foram os de conteúdo dos áudios, que eram compostos de gravações curtas de quatro locutores identificados (dois homens e duas mulheres), em que cada falante possuía 24 pares de áudios (de referência e picotados) (HARTE; GILLEN; HINES, 2015), garantindo variedade suficiente para realizar os testes com a ferramenta de biometria de voz. Os arquivos picotados contaram com características de deformação diferentes, em questão de duração de cada picote, intervalo entre eventos e tipo de picotamento (substituição por zero, substituição pela porção anterior ou apagamento total do pacote). Eram utilizadas frases curtas em língua inglesa, variadas e randomizadas, que não se repetiam dentro dos arquivos utilizados.

Todas as características de formato de arquivo estiveram de acordo com o esperado, com exceção da frequência de amostragem. O valor original nos arquivos foi de 48 kHz, mas foi possível contornar esse problema com uma etapa de *downsampling*, descrita no item 4.2.2.

4.3.1 Descrição dos arquivos componentes do *dataset*

De acordo com Harte, Gillen e Hines (2015), o *dataset* foi gerado com base no documento ITU-T Rec. P.800 de métodos para determinação subjetiva de qualidade de transmissão, que fornece esquemas aprovados para se determinar quão satisfatórias devem ser as conexões telefônicas em determinados sistemas de comunicação. As gravações de voz em estúdio contaram com 23 locutores, dos quais apenas quatro, dois homens e duas mulheres, tiveram suas vozes escolhidas

para compor o conjunto final de áudios. Esses receberam a identificação de FA, FG, MK e ML, sendo as letras iniciais indicativas do sexo – *Female* e *Male*, feminino e masculino respectivamente. O Quadro 1 resume essas informações.

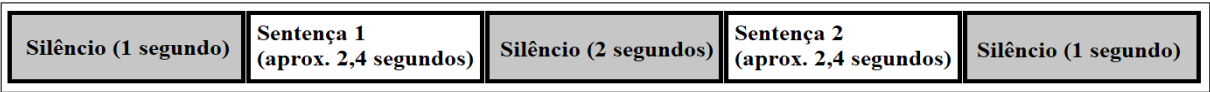
QUADRO 1 – Locutores envolvidos com a gravação do *dataset* utilizado

Locutor	Sexo	Áudios de referência (unidades)	Áudios picotados (unidades)
FA	Feminino	24	24
FG	Feminino	24	24
MK	Masculino	24	24
ML	Masculino	24	24

Fonte: Autoria própria (2023).

Cada um deles gravou em uma câmara anecóica 60 sentenças com duração média de 2,4s cada (HARTE; GILLEN; HINES, 2015). Os arquivos de áudio foram gerados unindo duas sentenças aleatórias do mesmo locutor, separadas por um silêncio de 2 segundos e com silêncio de 1 segundo no começo e fim do áudio. A Figura 36 ilustra o conteúdo de cada arquivo.

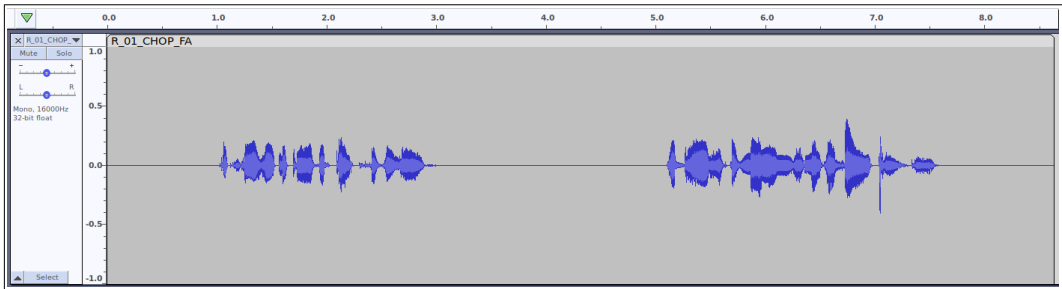
FIGURA 36 – Conteúdo de cada arquivo de áudio do *dataset*



Fonte: Adaptado de Harte, Gillen e Hines (2015, p. 2).

A Figura 37 mostra a forma de onda completa de um exemplo de arquivo utilizado, em que é notável a separação de dois aglomerados de fala, nos intervalos aproximados de 1,0s a 3,0s e de 5,0s a 7,5s.

FIGURA 37 – Exemplo de forma de onda completa em arquivo de áudio utilizado



Fonte: Autoria própria (2023).

Para o grupo de áudios utilizado nesta pesquisa, foram gerados 24 arquivos para cada locutor, totalizando 96 exemplos. Os arquivos originais, além de serem disponibilizados no *dataset* como referência, serviram de base para a geração dos acometidos por picote. A degradação dos sinais foi feita em função de três parâmetros (HARTE; GILLEN; HINES, 2015):

a) modo de picote: “determina se as amostras serão substituídas por zeros (modo 1), completamente apagadas (modo 2) ou substituídas por porção anterior de amostras (modo 3).” (HARTE; GILLEN; HINES, p. 2, 2015, tradução nossa);

b) período de picote: “determina o comprimento (em milissegundos) de cada porção de amostras a ser picotada.” (HARTE; GILLEN; HINES, p. 2, 2015, tradução nossa);

c) taxa de picotamento: “determina com que intervalo de tempo (em milissegundos) uma porção de amostras deve ser picotada.” (HARTE; GILLEN; HINES, p. 2, 2015, tradução nossa).

O Quadro 2 informa as faixas de valor dos dois parâmetros numéricos citados anteriormente, período e taxa de picotamento. Salienta-se, no entanto, que quatro arquivos foram mantidos sem alteração. Os autores (HARTE; GILLEN; HINES, 2015) declaram que a taxa de picotamento mínima é zero; o quadro a seguir ignora esse valor.

QUADRO 2 – Faixa de valores utilizados na geração de picotes do *dataset*

Parâmetro	Valor mínimo	Valor máximo
Período de picote (s)	0,02	0,04
Taxa de picotamento (picotes/s)	0,15	6

Fonte: Adaptado de Harte, Gillen e Hines (2015, p. 2).

4.3.2 Adaptação de frequência de amostragem

Como visto anteriormente, as ferramentas de análise de voz trabalham com arquivos com frequência de amostragem de 16 kHz. O processo de diminuição da frequência de amostragem (*downsampling*) de um arquivo de 48 kHz para 16 kHz

compreende eliminar duas a cada três amostras do sinal, além de adequar seu cabeçalho.

Para o *downsampling* dos conjuntos de áudios originais e picotados, foi utilizada a ferramenta SoX, um *software* de processamento de som em linha de comando capaz de realizar diversas conversões e leituras (SOX, 2015). Nesta pesquisa, escreveu-se uma rotina em *Python* para a varredura dos arquivos de cada *dataset* e submissão de cada um ao SoX em modo de *resampling*, com auxílio da biblioteca *Subprocess*.

4.3.3 Insuficiência dos arquivos acometidos por picote e geração de *dataset* adequado

No início da pesquisa, notou-se que os arquivos acometidos por picote do *dataset* TCD-VoIP apresentavam resultados muito próximos dos arquivos de referência nas medições da ferramenta de biometria de voz. A documentação do conjunto escrita por Harte, Gillen e Hines (2015) deixa claro o uso de diversos valores entre os vistos no Quadro 2, mas testes locais apontaram que apenas os áudios com taxa de picotamento perto do valor máximo, de seis picotes por segundo, causavam um detrimento significativo no resultado. Sendo a maioria dos arquivos distorcidos acometida por uma taxa de picotamento baixa demais, foi necessário gerar um novo conjunto a partir do de referência, de modo a suportar os testes pretendidos.

A geração de novos arquivos foi feita com uma rotina em *Python*. Cada arquivo foi lido e teve seu sinal dividido em *frames* de 20ms cada, mesmo valor de um dos utilizados por Harte, Gillen e Hines. Estabeleceu-se a neutralização – levar os valores de todas as amostras a zero – de um *frame* a cada oito, o que acarreta uma taxa de picotamento de 6,25 picotes por segundo ou taxa de perda de pacotes de 12,5%, que foram distribuídos de maneira equidistante ao longo de cada arquivo, assim como nos áudios picotados de Harte, Gillen e Hines. O conjunto de arquivos gerados dessa forma substituiu completamente o *dataset* de áudios picotados fornecido por Harte, Gillen e Hines.

4.4 Testes com arquivos de referência

Os testes com arquivos de referência, assim como os com arquivos picotados, servem como base de comparação para os com arquivos corrigidos. São apresentados a seguir os resultados obtidos com os áudios de referência, que não contam com nenhuma inserção de picote.

O Quadro 3 apresenta os resultados obtidos com a ferramenta de biometria de voz. Os dados são apresentados da seguinte forma, que também é utilizada nas apresentações de resultados da biometria dos subcapítulos seguintes:

- a) a coluna “Locutor original” informa o locutor conhecido previamente dono dos determinados arquivos de áudio;
- b) a coluna “Perfil testado” indica para qual perfil cadastrado na API os áudios do locutor foram verificados;
- c) as colunas “4” a “24” indicam a taxa de reconhecimento resultante da verificação para dados áudios, visto que os três primeiros de cada locutor são usados para treinamento;
- d) a coluna “Média” apresenta a média dos 21 resultados para cada combinação de locutor original e perfil testado.

QUADRO 3 – Resultados com ferramenta de biometria de voz para *dataset* de referência

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,840	-1	0,930	0,920	0,820	0,800	0,820	0,840	0,890	0,870	0,670
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0,880	0,950	0,920	0,930	0,910	0,890	0,830	0,910	0,950	0,930	0,910
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,950	0,950	0,927	0,930	0,950	0,980	0,870	0,930	0,950	0,970	0,930
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,870	0,950	0,927	0,900	0,910	0,923	0,950	0,940	0,950	0,960	0,890

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,860	0,820	0,850	0,890	0,800	0,910	0,830	0,850	0,840	0,850	0,845
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,870	0,920	0,860	0,880	0,880	0,940	0,920	0,880	0,950	0,950	0,908
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,940	0,910	0,950	0,960	0,920	0,920	0,930	0,980	0,950	1,000	0,943
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,830	0,920	0,880	0,950	0,910	0,930	0,950	0,970	0,960	0,780	0,917

Fonte: Autoria própria (2023).

É possível obter informações importantes a partir da tabela anterior. Primeiramente, todas as verificações de áudios para o perfil errado tiveram resultado zero, ou seja, a API de biometria foi categórica ao informar que tal voz não pertence àquele perfil. Além disso, durante os testes foi exposto que o quinto áudio da locutora FA não pôde ser utilizado, devido ao seu tempo total de voz ser inferior ao mínimo aceitável pela API. Representa-se este resultado nas tabelas anterior e posteriores como um valor -1, que não participa do cálculo de média. Por fim, denota-se os valores 0,845, 0,908, 0,943 e 0,917 como as metas para médias obtidas para os locutores FA, FG, MK e ML respectivamente após a aplicação de algoritmos de correção sobre os áudios picotados.

Os resultados obtidos com a ferramenta de transcrição de áudio, como discutido anteriormente, devem conter uma lista de frases escritas para o *dataset* de referência, uma lista de frases para o *dataset* testado e os referentes valores de CER. As transcrições de referência sozinhas, portanto, não compreendem um resultado completo. As frases transcritas a partir dos áudios de referência são expostas no Apêndice A.

4.5 Testes com arquivos acometidos por picotes

Repetiu-se os testes com os áudios descritos no tópico 4.3.3, gerados a partir do *dataset* de referência com frequência de amostragem de 16 kHz e com a inserção de picotes de 20ms a cada 160ms de sinal. O Quadro 4 apresenta os resultados obtidos com a ferramenta de biometria de voz.

QUADRO 4 – Resultados com ferramenta de biometria de voz para *dataset* com picotes

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,790	-1	0,830	0,890	0,790	0,650	0,680	0,790	0,840	0,770	0,620
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0,840	0,900	0,840	0,870	0,860	0,840	0,830	0,820	0,890	0,830	0,850
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,880	0,840	0,880	0,870	0,890	0,940	0,800	0,900	0,910	0,880	0,860
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,830	0,910	0,910	0,870	0,910	0,900	0,960	0,840	0,900	0,790	0,880

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,800	0,740	0,820	0,810	0,767	0,830	0,760	0,740	0,840	0,740	0,775
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,743	0,840	0,820	0,870	0,830	0,880	0,870	0,780	0,910	0,910	0,849
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,900	0,840	0,850	0,910	0,820	0,850	0,860	0,890	0,870	0,930	0,875
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,770	0,860	0,800	0,930	0,830	0,920	0,940	0,940	0,960	0,770	0,877

Fonte: Autoria própria (2023).

A partir das colunas de médias dos dois quadros anteriores, é possível traçar uma comparação entre o resultado de referência e o com picotes, para se analisar a queda de confiança do reconhecimento dos áudios de cada locutor com seu próprio perfil. Observa-se esta análise no Quadro 5.

QUADRO 5 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* de referência e acometido por picotes

Locutor	Média de confiança de referência	Média de confiança com picotes	Variação de confiança
FA	0,845	0,775	-0,070
FG	0,908	0,849	-0,059
MK	0,943	0,875	-0,068
ML	0,917	0,877	-0,040

Fonte: Autoria própria (2023).

As transcrições de áudio deste *dataset* foram obtidas e a lista de frases escritas está disposta no Apêndice B. Tendo posse das frases de referência e das desse caso, é possível obter os valores de CER considerando respectivamente estas listas como frases de referência e hipotéticas para cálculo de distância de Levenshtein, como indicado no subcapítulo 4.1.

O Quadro 6 apresenta o resultado de cálculos de CER para os pares de transcrições dos *datasets* de referência e com picotes. As colunas se organizam da forma a seguir, que também é utilizada na apresentação de dados similares dos subcapítulos posteriores:

- a) a coluna “Locutor” separa os resultados por áudios de cada locutor do dataset;
- b) as colunas “1” a “24” apresentam os resultados de CER para os pares referência-hipótese das frases transcritas dos áudios 1 a 24 dos *datasets* de referência e acometido por picotes;
- c) a coluna “Média” apresenta a média aritmética dos 24 resultados exibidos para cada locutor.

QUADRO 6 – Valores de CER para comparação entre transcrições de áudio dos *datasets* de referência e acometido por picotes

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese											
	1	2	3	4	5	6	7	8	9	10	11	12
FA	0,115	0,055	0,111	0,106	0,057	0,171	0,182	0,048	0,000	0,083	0,172	0,068
FG	0,154	0,365	0,108	0,621	0,158	0,247	0,106	0,155	0,000	0,190	0,049	0,068
MK	0,103	0,282	0,035	0,091	0,072	0,045	0,241	0,053	0,197	0,205	0,000	0,114
ML	0,364	0,406	0,107	0,208	0,025	0,188	0,089	0,159	0,155	0,141	0,041	0,307

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese												
	13	14	15	16	17	18	19	20	21	22	23	24	Média
FA	0,090	0,122	0,000	0,176	0,000	0,083	0,271	0,247	0,068	0,127	0,123	0,311	0,116
FG	0,362	0,203	0,000	0,326	0,056	0,070	0,055	0,104	0,131	0,302	0,104	0,433	0,186
MK	0,159	0,114	0,226	0,079	0,159	0,136	0,167	0,186	0,000	0,090	0,111	0,092	0,123
ML	0,122	0,235	0,044	0,103	0,410	0,365	0,292	0,076	0,063	0,055	0,130	0,189	0,178

Fonte: Autoria própria (2023).

As médias obtidas para os resultados de CER dos locutores FA, FG, MK e ML foram 0,116, 0,186, 0,123 e 0,178 respectivamente. Visto que as ferramentas de PLC tem como função reparar os sinais de áudio picotados e trazê-los a formatos similares aos de referência, espera-se que os testes com áudios corrigidos tenham valores mais próximos de zero, ou seja, que haja menos diferenças nas transcrições de cada par de arquivos referência-corrigido.

4.6 Testes com arquivos corrigidos

Diversas formas de PLC foram implementadas em Python e aplicadas sobre os arquivos do *dataset* de áudios picotados, gerando um *dataset* novo para cada caso. A seguir discute-se cada algoritmo de PLC e apresenta-se os resultados obtidos com as ferramentas de biometria de voz e transcrição de áudio.

Devido ao cenário proposto de arquivos provenientes de comunicação VoIP, assume-se que os sinais de áudio foram gerados a partir de frames de 20ms e que os acometidos por perda de pacote têm em seu conteúdo todas as amostras em zero, como discutido no subcapítulo 4.2.3.

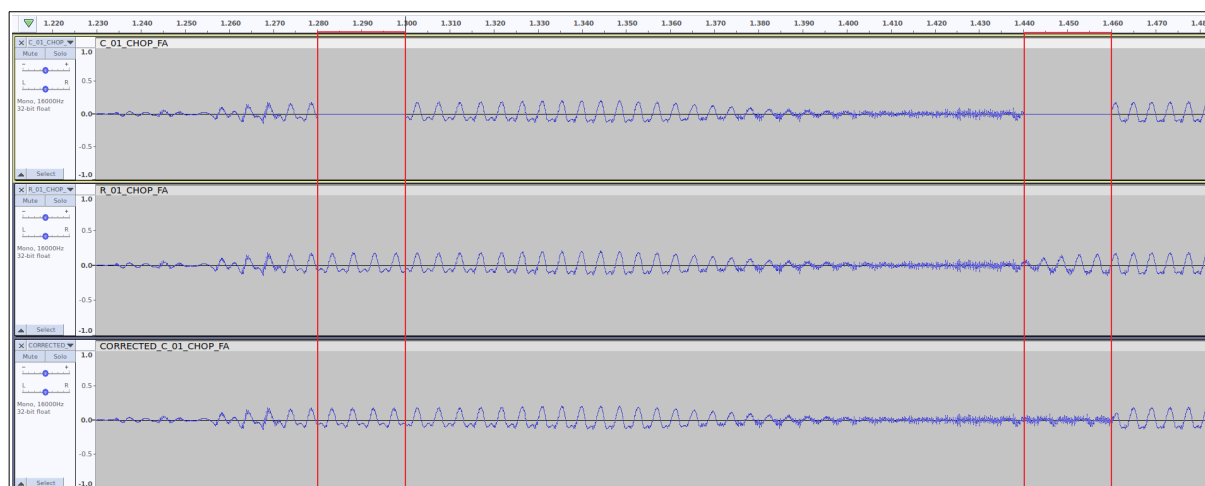
4.6.1 Algoritmo I – Pitch Wave Replication baseado no Anexo A do codec ITU-T G.711 (1999)

O primeiro algoritmo é baseado na técnica exposta no Anexo A do *codec* G.711. Implementa-se em conjunto os métodos de PWR e OLA, das mesmas formas vistas no subcapítulo 2.4.1 e no final do subcapítulo 2.4.2 respectivamente.

Como discutido anteriormente, o documento recomenda a geração de um sinal sintético com a técnica de PWR para o intervalo do picote (ITU-T, 1999). Produzem-se também extensões deste sinal para zonas adicionais com comprimento de um quarto de período de *pitch* antes e depois do picote, utilizadas para aplicar-se OLA janelada com o final e início dos frames anterior e posterior respectivamente, de forma a suavizar a transição entre sinais naturais e sintético.

A Figura 38 apresenta um exemplo de trecho do sinal corrigido pelo algoritmo I. O gráfico superior exhibe o áudio com picotes, enquanto o do meio e inferior mostram os sinais de referência e corrigido respectivamente. Com base nas regiões notadas na forma de onda superior, os retângulos vermelhos demarcam os períodos de picote.

FIGURA 38 – Trecho de áudios com picotes, de referência e corrigido por algoritmo I

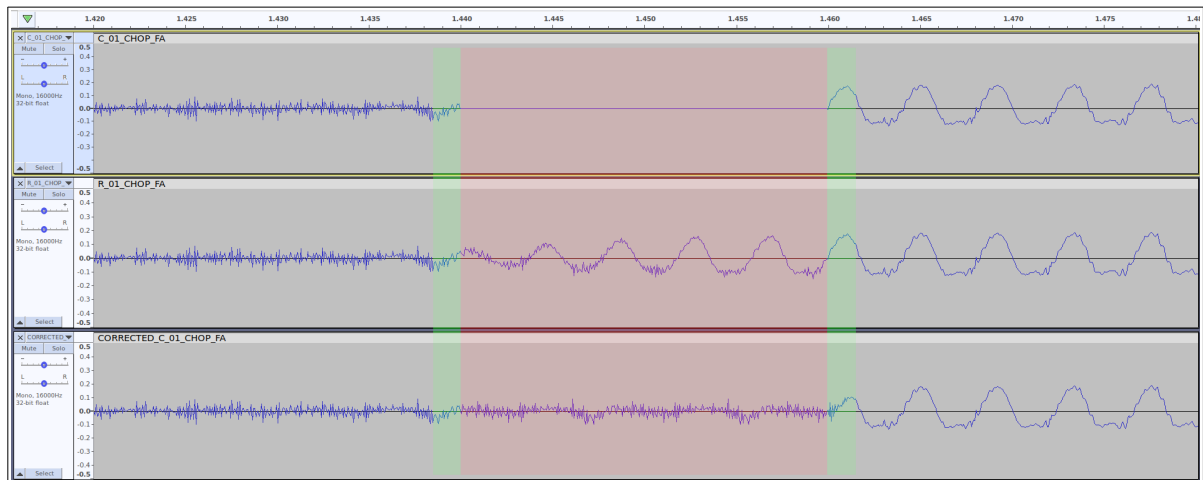


Fonte: Autoria própria (2023).

A região marcada à esquerda compreende um sinal vocalizado, mas a à direita está situada exatamente na transição entre um sinal não-vocalizado e vocalizado. É possível notar em ambas as regiões a replicação padronizada de formatos imediatamente anteriores aos picotes, como espera-se de uma ferramenta

de PWR. A região à direita evidencia o efeito da OLA janelada, na transição entre o picote e o *frame* seguinte. A Figura 39 detalha esse caso. Conforme recomendação dos autores (ITU-T, 1999) para casos em que a complexidade operacional não é um fator limitante, utiliza-se janela de Hanning para a OLA.

FIGURA 39 – Detalhe de replicação de período de *pitch* de sinal de voz não-vocalizado com PWR e transição entre *frames* com OLA janelada



Fonte: Autoria própria (2023).

Tendo aplicado o algoritmo de correção I sobre o *dataset* picotado, obtém-se o primeiro *dataset* corrigido. Dando continuidade aos testes com a ferramenta de biometria de voz, o Quadro 7 apresenta os resultados obtidos com este conjunto de arquivos.

QUADRO 7 – Resultados com ferramenta de biometria de voz para *dataset* corrigido pelo algoritmo I

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,837	-1	0,900	0,900	0,830	0,710	0,780	0,810	0,850	0,870	0,640
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0,850	0,950	0,900	0,890	0,860	0,890	0,830	0,900	0,940	0,920	0,870
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,900	0,950	0,910	0,890	0,950	0,970	0,850	0,930	0,930	0,970	0,910
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,850	0,910	0,920	0,910	0,897	0,890	0,950	0,893	0,930	0,950	0,870

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,820	0,790	0,840	0,860	0,790	0,900	0,840	0,810	0,830	0,750	0,818
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,860	0,900	0,820	0,900	0,830	0,920	0,920	0,850	0,920	0,940	0,889
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,957	0,880	0,950	0,970	0,840	0,840	0,920	0,950	0,893	0,970	0,920
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,840	0,880	0,900	0,940	0,910	0,930	0,950	0,960	0,970	0,790	0,907

Fonte: Autoria própria (2023).

Ao se comparar os resultados de médias de confiança do caso anterior com os do *dataset* picotado, fica evidente uma melhora no uso da ferramenta de biometria de voz. O Quadro 8 exibe essa comparação, além da variação entre os dois conjuntos de resultados. Estes valores positivos são aquilo que se espera dos algoritmos, pois é a medida que aproxima o resultado corrigido do de referência.

QUADRO 8 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* acometido por picotes e corrigido por algoritmo de PLC I e variação entre os valores

Locutor	Média com <i>d. picotado</i>	Média com <i>d. corrigido</i> por algoritmo I	Variação com algoritmo I em relação a <i>d. picotado</i>
FA	0,775	0,818	0,043
FG	0,849	0,889	0,040
MK	0,875	0,920	0,045
ML	0,877	0,907	0,030

Fonte: Autoria própria (2023).

Executando a ferramenta de transcrição de áudio sobre o *dataset* em questão, obtém-se a lista de frases expostas no Apêndice C. O Quadro 9 apresenta os valores de CER calculados com as transcrições de referência e as deste conjunto.

QUADRO 9 – Valores de CER para comparação entre transcrições de áudio dos *datasets* de referência e com picotes corrigidos por algoritmo de PLC I

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese											
	1	2	3	4	5	6	7	8	9	10	11	12
FA	0,064	0,000	0,074	0,071	0,115	0,066	0,026	0,024	0,000	0,048	0,031	0,068
FG	0,185	0,063	0,000	0,500	0,000	0,130	0,061	0,282	0,000	0,052	0,000	0,258
MK	0,000	0,192	0,000	0,000	0,205	0,000	0,139	0,000	0,000	0,000	0,000	0,057
ML	0,307	0,341	0,143	0,000	0,175	0,116	0,000	0,000	0,042	0,167	0,137	0,053

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese												
	13	14	15	16	17	18	19	20	21	22	23	24	Média
FA	0,060	0,000	0,000	0,059	0,000	0,000	0,157	0,026	0,068	0,025	0,096	0,081	0,048
FG	0,435	0,169	0,000	0,326	0,366	0,113	0,055	0,078	0,082	0,429	0,039	0,463	0,170
MK	0,203	0,051	0,036	0,000	0,000	0,057	0,000	0,163	0,000	0,077	0,037	0,132	0,056
ML	0,110	0,173	0,000	0,000	0,346	0,447	0,347	0,038	0,114	0,000	0,247	0,230	0,147

Fonte: Autoria própria (2023).

O Quadro 10 compara as médias de CER por locutor para os *datasets* com picote e corrigido com este algoritmo de PLC. Apresenta-se na última coluna a variação entre os valores dos dois casos. Esta distância denota quanto a taxa de erro de caracteres aumentou ou diminuiu, ou seja, um número negativo aponta uma melhora dos resultados e uma transcrição mais fiel à de referência.

QUADRO 10 – Comparação entre resultados entre médias de CER obtidas para *datasets* acometido por picotes e corrigido por algoritmo de PLC I e variação entre os valores

Locutor	Médias com <i>d. picotado</i>	Médias com <i>d. corrigido</i> por algoritmo I	Variação com algoritmo I em relação a <i>d. picotado</i>
FA	0,116	0,048	-0,070
FG	0,186	0,170	-0,016
MK	0,123	0,056	-0,067
ML	0,178	0,147	-0,031

Fonte: Autoria própria (2023).

4.6.2 Algoritmo II – *Linear Prediction Coding* baseado em Elsabrouty, Bouchard e Aboulnasr (2004)

O segundo algoritmo deste projeto é uma implementação de *Linear Prediction Coding* baseada nas equações dadas por Elsabrouty, Bouchard e Aboulnasr (2004), além de todas as descrições dos autores citados no capítulo 2.4.3. Empreendeu-se em *Python* um código para gerar sinal sintético conforme a Equação 16, com uso da biblioteca *Librosa* para cálculo de coeficientes de predição linear e variáveis detalhadas no Quadro 11. Estes valores são baseados em recomendações dos autores consultados e alterados durante a execução para se adaptar à frequência de amostragem deste projeto.

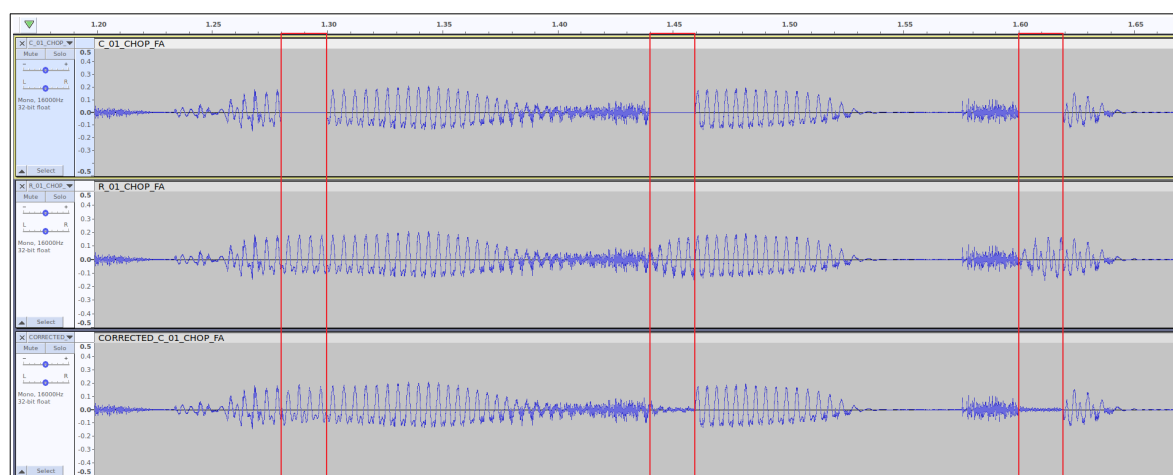
QUADRO 11 – Valores de variáveis em implementação de PLC II com LPC em *Python*

Variável	Valor
G	0,01
α	0,7
β	0,3
Ordem do filtro de predição linear	50
Comprimento de janela de correlação (amostras)	200
Período de <i>pitch</i> máximo (amostras)	238
Frequência de <i>pitch</i> mínima (Hz)	68
Período de <i>pitch</i> mínimo (amostras)	40
Frequência de <i>pitch</i> máxima (Hz)	400

Fonte: Autoria própria (2023).

A Figura 40 ilustra o mesmo segmento de três arquivos de áudio correspondentes à mesma gravação, na versão com picotes no gráfico superior, de referência no do meio e corrigida pelo segundo algoritmo de PLC no gráfico inferior. Visualizam-se três regiões de picote, marcadas em vermelho. Em relação ao *frame* anterior de cada uma, elas se tratam de sinal vocalizado no primeiro picote e não-vocalizado nos dois últimos, o que se demonstra no sinal sintético gerado para cada caso. Nota-se que o algoritmo reconstrói sinais vocalizados com maior fidelidade que não-vocalizados.

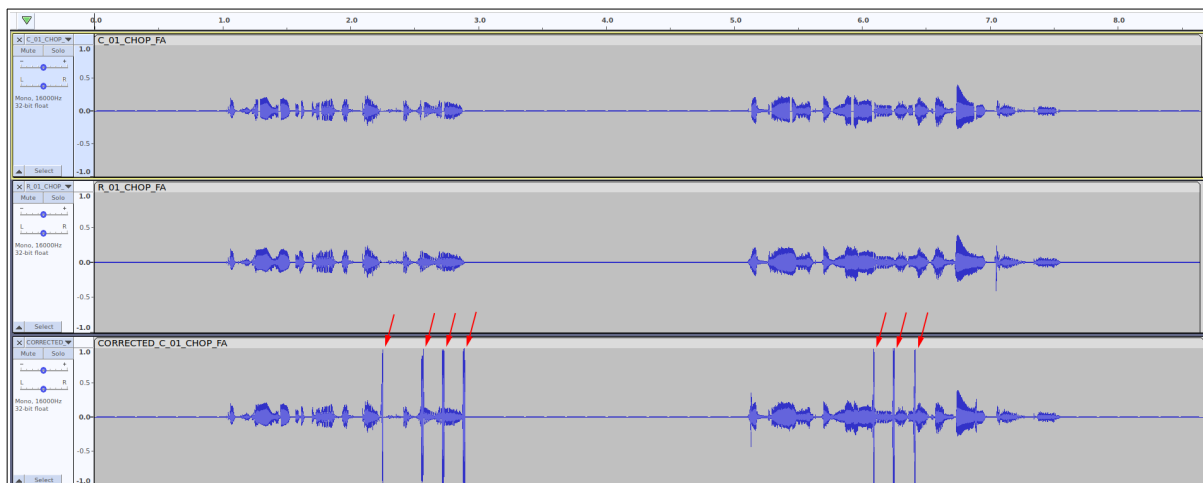
FIGURA 40 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC II



Fonte: Autoria própria (2023).

Durante as execuções de testes foi possível notar também a presença de artefatos sonoros gerados pelo algoritmo de LPC em alguns intervalos referentes a picotes em diversos arquivos de áudio. A Figura 41 ilustra no gráfico inferior um exemplo de arquivo de áudio corrigido completo, com os arquivos picotado e de referência nos gráficos superior e central respectivamente. As sete setas vermelhas destacam o fenômeno em foco.

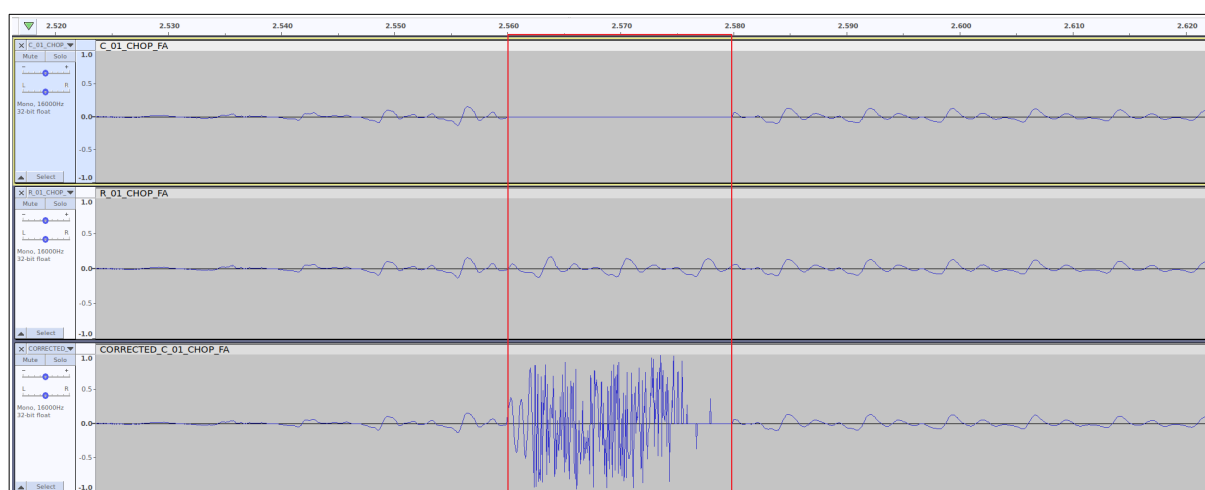
FIGURA 41 – Formas de onda completa de arquivos de áudio picotado, de referência e corrigido por algoritmo de PLC II com destaque para artefatos sonoros



Fonte: Autoria própria (2023).

A Figura 42 apresenta um destes momentos mais detalhadamente, com a região exata do picote demarcada em vermelho. Percebe-se o comportamento instável do sinal reconstruído neste trecho, que se inicia por alguns semiciclos com amplitude condizente com o resto do sinal e logo atinge um volume muito superior. Este fenômeno é percebido auditivamente como um curto estouro.

FIGURA 42 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC II



Fonte: Autoria própria (2023).

A ferramenta de biometria de voz foi aplicada sobre o dataset corrigido pelo algoritmo de PLC II e os resultados de confiança obtidos são apresentados no Quadro 12.

QUADRO 12 – Resultados com ferramenta de biometria de voz para *dataset* corrigido pelo algoritmo II

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,780	-1	0,890	0,900	0,820	0,760	0,750	0,780	0,830	0,830	0,630
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0,810	0,860	0,870	0,860	0,820	0,847	0,820	0,880	0,910	0,900	0,870
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,910	0,860	0,890	0,860	0,950	0,950	0,850	0,920	0,930	0,920	0,880
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	-1	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,810	0,900	0,930	0,890	0,900	0,890	0,940	0,870	0,940	0,930	0,850

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,850	0,770	0,820	0,820	0,770	0,820	0,820	0,790	0,800	0,830	0,803
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,870	0,900	0,860	0,840	0,847	0,900	0,870	0,860	0,890	0,860	0,864
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,910	0,850	0,910	0,940	0,810	0,870	0,840	0,940	0,880	0,920	0,895
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,840	0,890	0,863	0,890	0,840	0,920	0,940	0,930	0,940	0,760	0,889

Fonte: Autoria própria (2023).

Repetindo a comparação do resultado de algoritmo de correção com os valores obtidos para *dataset* picotado, observa-se no Quadro 13 que o algoritmo de PLC II também obteve uma melhoria de desempenho aplicado à ferramenta de biometria de voz. Apesar disso, as variações de confiança com algoritmo de PLC II em relação ao *dataset* picotado são menores que no caso do algoritmo I.

QUADRO 13 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* acometido por picotes e corrigido por algoritmo de PLC II e variação entre os valores

Locutor	Média com <i>d.</i> picotado	Média com <i>d.</i> corrigido por algoritmo II	Variação com algoritmo II em relação a <i>d.</i> picotado
FA	0,775	0,803	0,028
FG	0,849	0,864	0,015
MK	0,875	0,895	0,020
ML	0,877	0,889	0,012

Fonte: Autoria própria (2023).

Em seguida, realizou-se o teste com a ferramenta de transcrição de texto. As frases identificadas pela ferramenta são expostas no Apêndice D. Os valores de CER obtidos são apresentados no Quadro 14.

QUADRO 14 – Valores de CER para comparação entre transcrições de áudio dos *datasets* de referência e com picotes corrigidos por algoritmo de PLC II

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese											
	1	2	3	4	5	6	7	8	9	10	11	12
FA	0,218	0,233	0,099	0,094	0,241	0,132	0,506	0,253	0,064	0,214	0,203	0,000
FG	0,200	0,556	0,311	0,845	0,421	0,494	0,667	0,521	0,145	0,707	0,213	0,677
MK	0,552	0,474	0,349	0,338	0,277	0,227	0,241	0,329	0,250	0,233	0,202	0,250
ML	0,341	0,451	0,345	0,264	0,175	0,420	0,506	0,333	0,507	0,308	0,219	0,373

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese												
	13	14	15	16	17	18	19	20	21	22	23	24	Média
FA	0,119	0,257	0,093	0,000	0,000	0,181	0,157	0,195	0,162	0,177	0,205	0,216	0,168
FG	0,580	0,407	0,000	0,198	0,493	0,183	0,562	0,143	0,426	0,317	0,286	0,552	0,413
MK	0,232	0,316	0,226	0,118	0,330	0,375	0,417	0,233	0,082	0,500	0,309	0,316	0,299
ML	0,317	0,407	0,176	0,385	0,449	0,529	0,528	0,076	0,418	0,342	0,351	0,500	0,363

Fonte: Autoria própria (2023).

O resultado de médias de CER obtido para o algoritmo II se mostrou problemático. O Quadro 15 expõe a comparação entre este cenário e os resultados com *dataset* picotado. Ao contrário do que se obteve com o PLC I, a última coluna do quadro mostra valores positivos, ou seja, as médias de taxa de erros de caractere para cada locutor aumentaram com a geração de áudios por esse algoritmo.

QUADRO 15 – Comparação entre resultados entre médias de CER obtidas para *datasets* com picote e corrigido por algoritmo de PLC II e variação entre os valores

Locutor	Médias com <i>d.</i> picotado	Médias com <i>d.</i> corrigido por algoritmo II	Variação com algoritmo II em relação a <i>d.</i> picotado
FA	0,116	0,168	0,050
FG	0,186	0,413	0,227
MK	0,123	0,299	0,176
ML	0,178	0,363	0,185

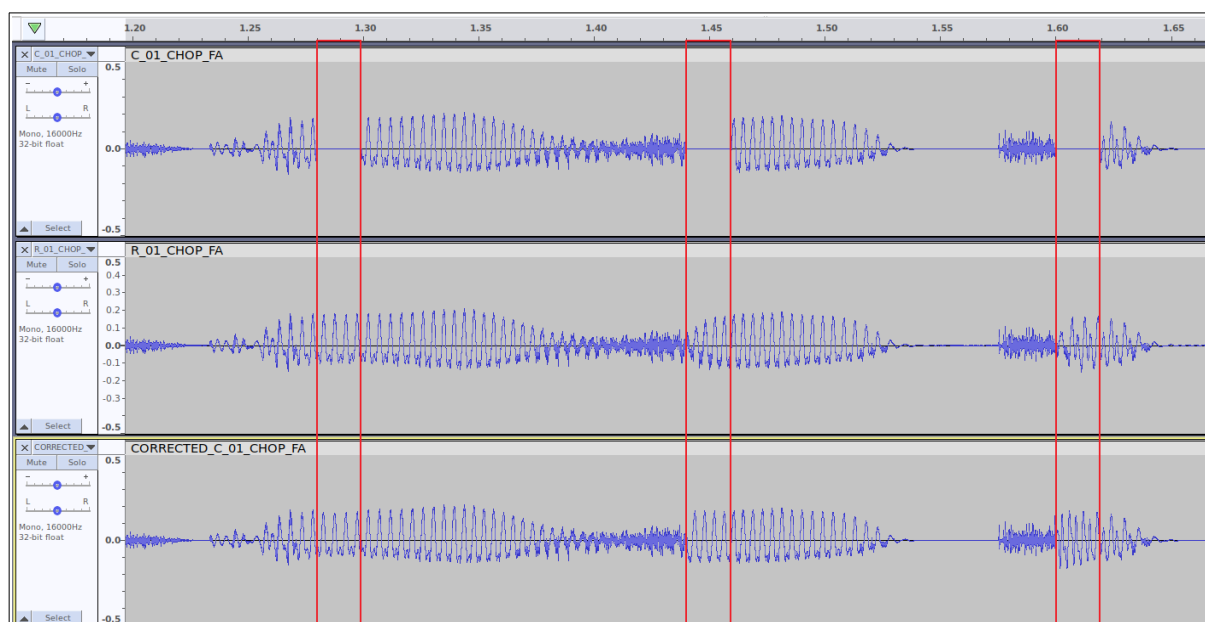
Fonte: Autoria própria (2023).

4.6.3 Algoritmo III – *Linear Prediction Coding* com sentido inverso

O terceiro algoritmo de PLC executa o método de LPC *backward*, similar ao do algoritmo II, diferindo no sentido de sintetização do novo sinal, utilizando como base de análise o frame seguinte ao picote e gerando amostras do final ao começo do *frame* perdido.

Com estas modificações, aplica-se no *dataset* picotado o algoritmo de correção por LPC *backward* e obtém-se o terceiro *dataset* de áudios corrigidos. A Figura 43 apresenta um exemplo de trecho de arquivo gerado nesse cenário, sendo do mesmo arquivo e intervalo de tempo retratados na Figura 40, exposta no gráfico inferior. Os gráficos superior e central retratam os sinais picotado e de referência respectivamente.

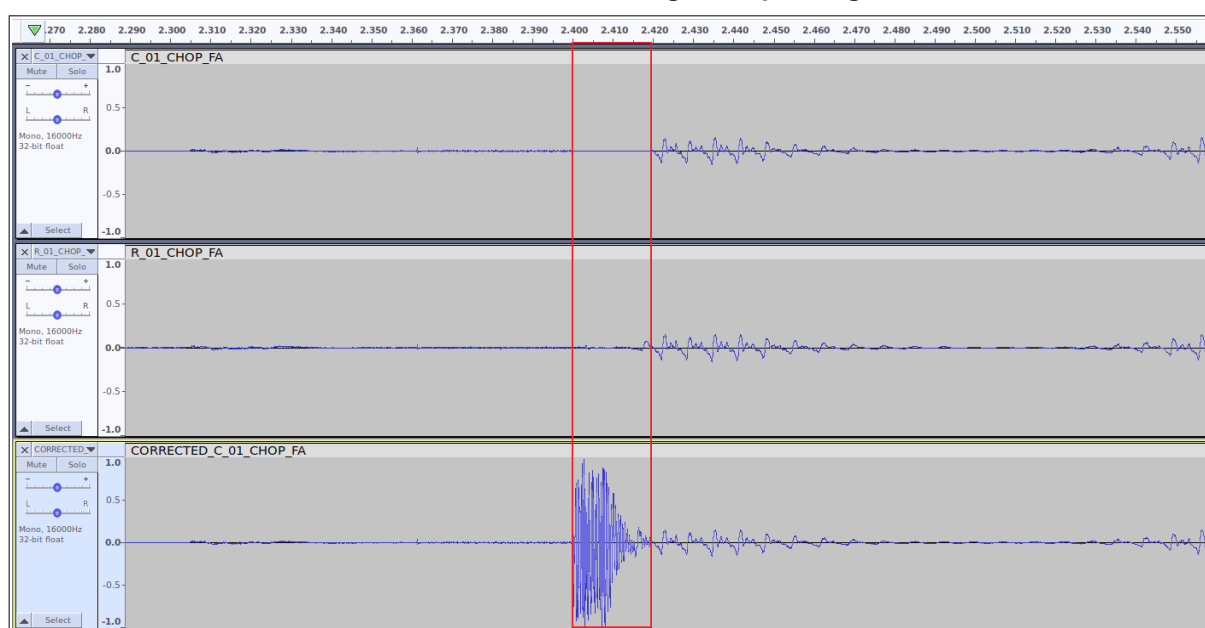
FIGURA 43 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC III



Fonte: Autoria própria (2023).

É notável na figura anterior a geração de sinal que se encaixa com a região posterior a si. Na figura 40, que ilustra o mesmo trecho corrigido pelo algoritmo de LPC *forward*, observa-se que no segundo e terceiro picotes o sinal utilizado para basear a correção é não-vocalizado, o que faz com que os sinais sintetizados nestes intervalos tenham a mesma característica. O fenômeno de geração de artefatos sonoros também foi observado no *dataset* desse caso, embora em menor quantidade em relação ao *dataset* corrigido pelo algoritmo de PLC II. A Figura 44 detalha uma das ocorrências.

FIGURA 44 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC III



Fonte: Autoria própria (2023).

É visível na região marcada em vermelho da figura anterior, referente ao frame picotado, que o sinal gerado tem um comportamento instável, assim como vê-se na ilustração de artefato sonoro do algoritmo anterior, mas no sentido direita-esquerda.

A aplicação da ferramenta de biometria de voz sobre o *dataset* gerado por meio do algoritmo de correção III gera os resultados vistos no Quadro 16. Uma mudança destacável ocorrida nesse conjunto de resultados de biometria de voz em relação aos anteriores é a adequação do quinto áudio da locutora FA, que obteve resultados numéricos normais ao invés de erro na API como em todos os testes anteriores.

QUADRO 16 – Resultados com ferramenta de biometria de voz para *dataset* corrigido pelo algoritmo III

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,780	0,930	0,880	0,880	0,750	0,790	0,770	0,780	0,830	0,790	0,620
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,870	0,910	0,870	0,890	0,880	0,830	0,790	0,850	0,890	0,900	0,860
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,920	0,850	0,880	0,830	0,950	0,920	0,810	0,900	0,917	0,950	0,880
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,827	0,900	0,930	0,880	0,900	0,890	0,940	0,890	0,900	0,930	0,880

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,890	0,760	0,800	0,830	0,730	0,800	0,807	0,790	0,770	0,820	0,800
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,870	0,880	0,860	0,800	0,840	0,890	0,900	0,860	0,880	0,880	0,867
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,910	0,820	0,900	0,960	0,810	0,870	0,840	0,940	0,920	0,930	0,891
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,830	0,870	0,850	0,837	0,850	0,920	0,950	0,930	0,900	0,780	0,885

Fonte: Autoria própria (2023).

Os resultados obtidos no teste de ferramenta de biometria de voz deste cenário foram similares aos com o algoritmo II, com valores em média ligeiramente inferiores. O Quadro 17 apresenta a comparação entre resultados picotados e com algoritmo de correção III.

QUADRO 17 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* acometido por picotes e corrigido por algoritmo de PLC III e variação entre os valores

Locutor	Média com <i>d.</i> picotado	Média com <i>d.</i> corrigido por algoritmo III	Variação com algoritmo III em relação a <i>d.</i> picotado
FA	0,775	0,800	0,025
FG	0,849	0,867	0,018
MK	0,875	0,891	0,016
ML	0,877	0,885	0,008

Fonte: Autoria própria (2023).

Prosseguindo ao teste com ferramenta de transcrição de áudio, as frases obtidas com o *dataset* corrigido pelo algoritmo III são expostas no Apêndice E e os resultados de CER para cada locutor vistos no Quadro 18.

QUADRO 18 – Valores de CER para comparação entre transcrições de áudio dos *datasets* de referência e com picotes corrigidos por algoritmo de PLC III

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese											
	1	2	3	4	5	6	7	8	9	10	11	12
FA	0,179	0,274	0,099	0,176	0,184	0,184	0,442	0,108	0,167	0,250	0,094	0,081
FG	0,231	0,333	0,432	0,724	0,289	0,416	0,364	0,310	0,032	0,310	0,426	0,387
MK	0,356	0,449	0,488	0,325	0,193	0,386	0,532	0,513	0,224	0,233	0,106	0,273
ML	0,420	0,440	0,274	0,361	0,250	0,362	0,203	0,348	0,310	0,218	0,315	0,440

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese												
	13	14	15	16	17	18	19	20	21	22	23	24	Média
FA	0,104	0,149	0,360	0,103	0,135	0,208	0,300	0,195	0,122	0,342	0,178	0,459	0,204
FG	0,493	0,593	0,282	0,360	0,197	0,239	0,548	0,416	0,459	0,460	0,403	0,463	0,382
MK	0,232	0,190	0,190	0,368	0,159	0,318	0,236	0,314	0,035	0,487	0,148	0,250	0,292
ML	0,402	0,370	0,250	0,372	0,487	0,329	0,444	0,114	0,291	0,192	0,364	0,419	0,332

Fonte: Autoria própria (2023).

A comparação com os valores de CER obtidos com *dataset* picotado é exposta no Quadro 19.

QUADRO 19 – Comparação entre resultados entre médias de CER obtidas para *datasets* com picote e corrigido por algoritmo de PLC II e variação entre os valores

Locutor	Médias com <i>d. picotado</i>	Médias com <i>d. corrigido</i> por algoritmo III	Variação com algoritmo III em relação a <i>d. picotado</i>
FA	0,116	0,204	0,090
FG	0,186	0,382	0,196
MK	0,123	0,292	0,169
ML	0,178	0,332	0,154

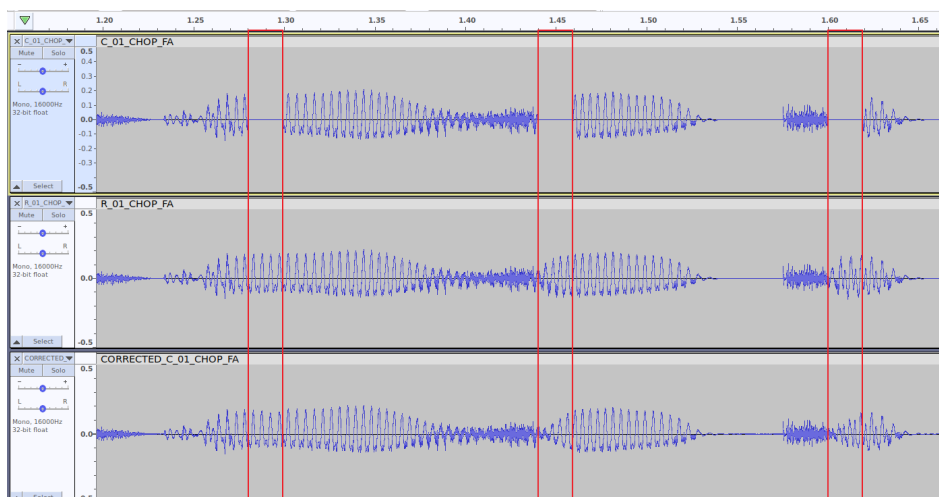
Fonte: Autoria própria (2023).

Novamente obtém-se resultados muito próximos dos alcançados com o algoritmo II. Os valores expostos na última coluna do quadro anterior diferem pouco dos mesmos no quadro 15, mantendo-se numa faixa similar e estando dois valores abaixo e dois acima.

4.6.4 Algoritmo IV – *Linear Prediction Coding* bilateral

Como discutido no tópico 4.1.5, o LPC bilateral é composto de etapas de LPC *forward*, *backward* e operação de OLA janelada com os dois sinais gerados. A técnica proposta é implementada em *Python* e executada sobre os arquivos do *dataset* picotado. Examina-se novamente o trecho do primeiro áudio da locutora FA, visto na primeira imagem de cada um dos dois últimos subcapítulos. A Figura 45 expõe o sinal picotado no gráfico superior, de referência no central e corrigido no inferior, tendo as regiões de picote circuladas em vermelho.

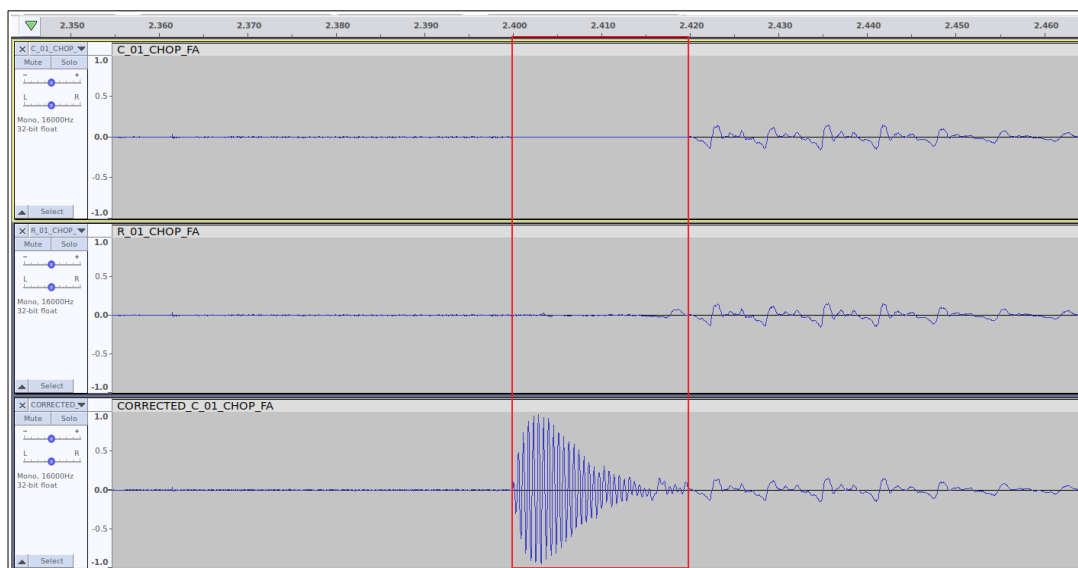
FIGURA 45 – Trecho de áudios picotado, de referência e corrigido por algoritmo de PLC IV



Fonte: Autoria própria (2023).

É perceptível na Figura 45 que o sinal reconstruído transita de características não-vocalizadas a vocalizadas nos dois trechos de picotes à direita, o que é o que se busca ao agregar os métodos de LPC *forward* e *backward*. Ao mesmo tempo, a reconstrução do picote à esquerda, com ambos os frames vizinhos sendo compostos de sinal vocalizado, ocorre sem desordens. Nota-se também que em todos os casos os sinais iniciam e terminam sem haver desencontro com os frames adjacentes. Apesar disso, assim como nos dois algoritmos anteriores, é notada a presença de artefatos sonoros nos arquivos. A Figura 46 ilustra um caso.

FIGURA 46 – Detalhe de artefato sonoro gerado pelo algoritmo de PLC IV



Fonte: Autoria própria (2023).

O Quadro 20 apresenta os resultados obtidos em testes de biometria de voz com o *dataset* gerado pelo algoritmo de PLC IV.

QUADRO 20 – Resultados com ferramenta de biometria de voz para *dataset* corrigido pelo algoritmo III

Locutor original	Perfil testado	4	5	6	7	8	9	10	11	12	13	14
FA	FA	0,800	0,920	0,890	0,890	0,800	0,797	0,750	0,780	0,840	0,830	0,640
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,810	0,850	0,880	0,880	0,860	0,860	0,810	0,890	0,910	0,880	0,880
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,910	0,850	0,900	0,840	0,950	0,910	0,850	0,920	0,920	0,960	0,870
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,793	0,880	0,940	0,890	0,880	0,900	0,950	0,870	0,920	0,930	0,853

Locutor original	Perfil testado	15	16	17	18	19	20	21	22	23	24	Média
FA	FA	0,880	0,780	0,820	0,840	0,760	0,810	0,810	0,750	0,770	0,850	0,810
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
FG	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0,880	0,900	0,860	0,880	0,820	0,900	0,880	0,810	0,900	0,880	0,868
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0	0	0	0	0	0	0	0	0	0	0
MK	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0,930	0,817	0,890	0,910	0,810	0,880	0,840	0,943	0,910	0,930	0,892
	ML	0	0	0	0	0	0	0	0	0	0	0
ML	FA	0	0	0	0	0	0	0	0	0	0	0
	FG	0	0	0	0	0	0	0	0	0	0	0
	MK	0	0	0	0	0	0	0	0	0	0	0
	ML	0,800	0,880	0,880	0,890	0,850	0,890	0,930	0,910	0,930	0,750	0,882

Fonte: Autoria própria (2023).

As médias de confiança por locutor com seu próprio perfil nesse cenário são comparadas com os resultados com *dataset* picotado no Quadro 21. Observa-se um conjunto de resultados próximo dos algoritmos II e III.

QUADRO 21 – Comparação entre médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* acometido por picotes e corrigido por algoritmo de PLC IV e variação entre os valores

Locutor	Média com <i>d.</i> picotado	Média com <i>d.</i> corrigido por algoritmo IV	Variação com algoritmo IV em relação a <i>d.</i> picotado
FA	0,775	0,810	0,035
FG	0,849	0,868	0,019
MK	0,875	0,892	0,017
ML	0,877	0,882	0,005

Fonte: Autoria própria (2023).

Por fim, executa-se o teste de transcrição sobre o *dataset* corrigido pelo algoritmo IV. Os resultados textuais são expostos no Apêndice F, enquanto os valores de CER são apresentados no Quadro 22.

QUADRO 22 – Valores de CER para comparação entre transcrições de áudio dos *datasets* de referência e com picotes corrigidos por algoritmo de PLC IV

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese											
	1	2	3	4	5	6	7	8	9	10	11	12
FA	0,218	0,288	0,099	0,141	0,253	0,039	0,390	0,241	0,115	0,167	0,156	0,095
FG	0,246	0,571	0,432	0,707	0,276	0,390	0,545	0,535	0,145	0,500	0,459	0,516
MK	0,483	0,513	0,535	0,208	0,337	0,205	0,405	0,395	0,276	0,425	0,255	0,284
ML	0,284	0,549	0,286	0,375	0,288	0,304	0,519	0,391	0,408	0,513	0,315	0,453

Locutor	Taxa de erro de caracteres (CER) para cada par de arquivos referência-hipótese												
	13	14	15	16	17	18	19	20	21	22	23	24	Média
FA	0,179	0,230	0,093	0,044	0,000	0,139	0,257	0,247	0,122	0,316	0,151	0,284	0,178
FG	0,406	0,593	0,296	0,360	0,493	0,268	0,438	0,325	0,361	0,333	0,364	0,567	0,422
MK	0,319	0,266	0,202	0,355	0,341	0,455	0,389	0,360	0,082	0,500	0,247	0,316	0,340
ML	0,341	0,469	0,191	0,385	0,551	0,353	0,681	0,165	0,291	0,411	0,299	0,419	0,385

Fonte: Autoria própria (2023).

A comparação das médias obtidas nesse cenário de transcrição com as alcançadas nos testes com *dataset* picotado é exposta no Quadro 23.

QUADRO 23 – Comparação entre resultados entre médias de CER obtidas para *datasets* com picote e corrigido por algoritmo de PLC IV e variação entre os valores

Locutor	Médias com <i>d. picotado</i>	Médias com <i>d. corrigido por algoritmo IV</i>	Variação com algoritmo IV em relação a <i>d. picotado</i>
FA	0,116	0,178	0,062
FG	0,186	0,422	0,236
MK	0,123	0,340	0,217
ML	0,178	0,385	0,207

Fonte: Autoria própria (2023).

4.7 Resultados e discussões

O objetivo da implementação de algoritmos de correção de picote é a obtenção de resultados mais próximos dos de referência (não picotados), partindo dos resultados com *dataset* acometido por picotes. Nos termos de testes com ferramenta de biometria de voz, isso significa obter valores de confiança mais altos, que se aproximem dos de referência. Em relação a testes com ferramenta de transcrição de texto, o objetivo é alcançado obtendo valores de CER tendentes a zero, o que configura a aquisição de frases escritas com menos diferenças de caracteres em relação às de referência. Os conjuntos de resultados obtidos no decorrer deste trabalho podem ser analisados separadamente entre os testes de cada ferramenta.

Visto que, nos testes de biometria de voz, a ação de um algoritmo de PLC pode ser mensurada pela variação entre médias de confiança de testes com *datasets* corrigido e acometido por picotes, agrupa-se no Quadro 24 as colunas referentes a tal medida dos Quadros 8, 13, 17 e 21, relativas aos testes dos algoritmos I, II, III e IV, respectivamente.

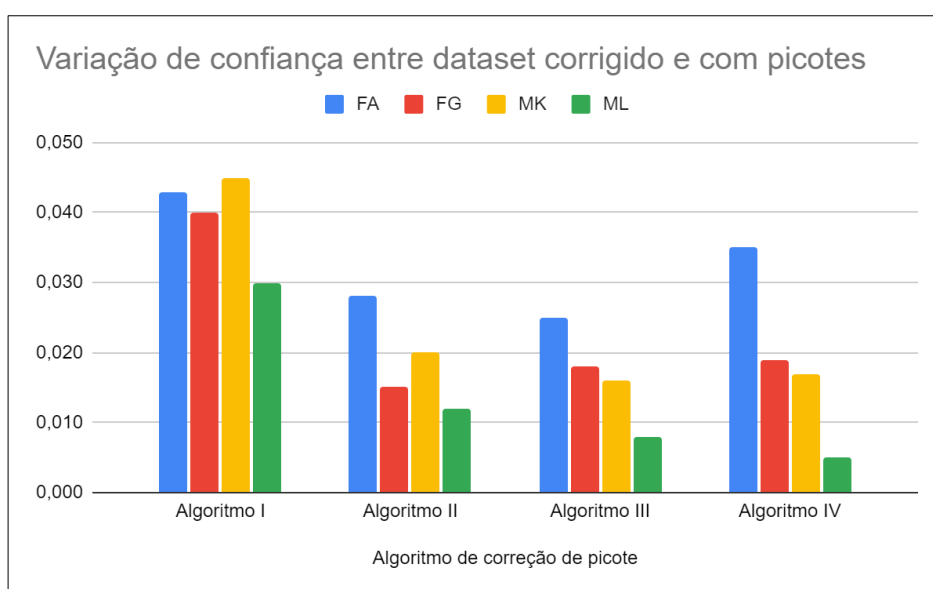
QUADRO 24 – Comparação entre variações de médias de confiança obtidas para cada locutor com seu próprio perfil com API de biometria entre *datasets* corrigidos por algoritmos de PLC I, II, III e IV e *dataset* acometido por picotes

Locutor	Variação com algoritmo I	Variação com algoritmo II	Variação com algoritmo III	Variação com algoritmo IV
FA	0,043	0,028	0,025	0,035
FG	0,040	0,015	0,018	0,019
MK	0,045	0,020	0,016	0,017
ML	0,030	0,012	0,008	0,005

Fonte: Autoria própria (2023).

Todas as ferramentas implementadas tiveram resultados positivos nos testes de biometria de voz. Ainda assim, o Quadro 24 expõe uma superioridade numérica clara do algoritmo de correção I, no que diz respeito às variações de média de confiança para todos os locutores. Todos os outros cenários envolvidos com a implementação de métodos de LPC obtiveram resultados inferiores para os quatro locutores. A Figura 47 representa os dados do Quadro 24 graficamente, para melhor compreensão da proporção dos resultados descritos.

FIGURA 47 – Representação gráfica de resultados obtidos em testes de algoritmos com ferramenta de biometria de voz



Fonte: Autoria própria (2023).

Novamente, no sentido de comparar resultados, o Quadro 25 agrupa valores obtidos pelos quatro algoritmos de PLC durante os testes com ferramenta de transcrição de áudio. São vistas, para cada locutor, as variações entre taxas de erro de caracteres com respectivos *datasets* corrigidos e com *dataset* picotado.

QUADRO 25 – Comparação entre variações de médias de CER para *datasets* corrigidos por algoritmos de PLC I, II, III e IV e *dataset* picotado

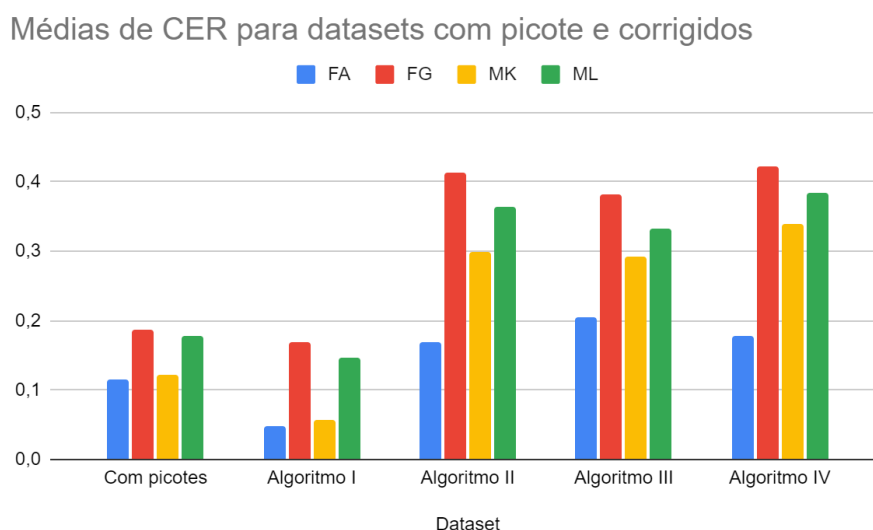
Locutor	Variação com algoritmo I	Variação com algoritmo II	Variação com algoritmo III	Variação com algoritmo IV
FA	-0,070	0,050	0,090	0,062
FG	-0,016	0,227	0,196	0,236
MK	-0,067	0,176	0,169	0,217
ML	-0,031	0,185	0,154	0,207

Fonte: Autoria própria (2023).

Em contraste com os resultados de biometria de voz, que foram proveitosos com todos os algoritmos, os testes com ferramenta de transcrição de áudio apresentaram melhora em apenas um dos quatro cenários. Considerando que uma variação positiva de CER, após a geração de um dataset corrigido a partir do picotado, indica o aumento de erros nos textos transcritos, as três formas de correção baseadas em LPC, os algoritmos II, III e IV, obtiveram resultados inutilizáveis. Em contrapartida, o algoritmo I ofereceu redução significativa de CER.

Para se visualizar graficamente o panorama mostrado no Quadro 25, cabe exibir as médias de CER obtidas em casa caso. A Figura 48 agrupa esses dados, possibilitando que se observe o aumento expressivo da taxa de erro de caracteres nos três últimos casos.

FIGURA 48 – Representação gráfica de resultados obtidos em testes de algoritmos com ferramenta de transcrição de áudio



Fonte: Autoria própria (2023).

Respalando-se nas observações anteriores, é possível concluir que o algoritmo de PLC I, baseado na estratégia fornecida por ITU-T (1999), é o que oferece melhores perspectivas para uma possível implementação em ambientes práticos. Sua aplicação nos contextos de biometria de voz e de transcrição de áudio se mostrou efetiva, uma vez que, no primeiro caso, o algoritmo apresentou os melhores valores numéricos com uma margem ampla e, no segundo, foi o único a auxiliar a ferramenta a obter resultados melhores após a correção de picotes.

REFERÊNCIAS

- BAKRI, Adil *et al.* *Automatic Recognition for VoIP with Packet Loss Concealment*. In: *FIRST INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE AND SPEECH PROCESSING*, 1, 2015, Algiers. **Anais...** [S.l.]: Procedia Computer Science, 2018, 88p. p.72-78.
- CARLE, Georg; BIERACK, Ernst W. *Survey of error recovery techniques for IP-based audio-visual multicast applications*. **IEEE Network**, [S.l.], v.11, n.6, p.24-36, nov.-dez. 1997.
- CISCO SYSTEMS INC. **IP Telephony/Voice over IP (VoIP)**. [S.l.]. 2017. Disponível em: <https://www.cisco.com/c/en/us/tech/voice/ip-telephony-voice-over-ip-voip/index.html?dtid=ossdc 000283>. Acesso em: 15 set. 2022.
- CMUSPHINX/POCKETSPHINX. **Repositório oficial do PocketSphinx 5.0.0**. [S.l.]. 2022. Disponível em: <https://github.com/cmusphinx/pocketsphinx/releases/tag/v5.0.0>. Acesso em: 15 nov. 2022.
- COLLOMB, Cedrick. **Linear Prediction and Levinson-Durbin Algorithm**. [S.l.]. 2009.
- ELSABROUTY, Maha; BOUCHARD, Martin; ABOUNASR, Tyseer. *Receiver-based packet loss concealment for pulse codemodulation (PCM G.711) coder*. **Signal Processing**, [S.l.], v.83, n.3, p.663-667, mar. 2004.
- GIL, Antônio C. **Como elaborar projetos de pesquisa**. 4ª ed. São Paulo: Atlas, 2002.
- GIL, Antônio C. **Métodos e técnicas de pesquisa social**. 6ª ed. São Paulo: Atlas, 2008.
- GOLD, Bernard. Digital Speech Networks. **Proceedings of the IEEE**, [S.l.], v.65, n.12, p.1636-1658, dez. 1977.
- GÜNDÜZHAN, Emre; MOMTAHAN, Kathryn. *A Linear Prediction Based Packet Loss Concealment Algorithm for PCM Coded Speech*. *IEEE Transactions on Speech and Audio Processing*, [S.l.], v.9, n.8, p.778-785, nov. 2001.
- HARTE, Naomi; GILLEN, Eoin; HINES, Andrew. *TCD-VoIP, a Research Database of Degraded Speech for Assessing Quality in VoIP Applications*. In: *2015 SEVENTH INTERNATIONAL WORKSHOP ON QUALITY OF MULTIMEDIA EXPERIENCE (QOMEX)*, 7, 2015, Pílos. **Anais...** [S.l.]: IEEE, 2015.
- HARTPENCE, Bruce. **Packet Guide to Voice over IP**. 1ª ed. Sebastopol: O'Reilly Media, 2013.

HATFIELD, Trevor. **A Guide On the VoIP Market, Its Share and Market Size**. [S.l.]. 2022. Disponível em: <https://www.ring4.com/blog/voip-market-guide>. Acesso em: 30 jan. 2023.

ITU-T. **Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711**. [S.l.]. 1999. Disponível em: <https://www.itu.int/rec/T-REC-G.711-199909-!!Appl/en>. Acesso em: 20 out. 2022.

JIWER. **Repositório oficial**. [S.l.]. 2022. Disponível em: <https://pypi.org/project/jiwer/>. Acesso em: 30 dez. 2022.

KARAPANTAZIS, Stylianos; PAVLIDOU, Fotini-Niovi. *VoIP: A comprehensive survey on a promising technology*. **Computer Networks**. [S.l.], v.53, n.12, p.2050-2090, ago. 2009.

KHOMP INDÚSTRIA E COMÉRCIO LTDA. **Telefones IP e DECT**. Florianópolis, 2022. Disponível em: <https://www.khomp.com/pt/categoria-de-produto/telefones-ip-inteligentes/>. Acesso em: 15 set. 2022.

KONDO, Kazuhiro; NAKAGAWA, Kiyoshi. A Speech Packet Loss Concealment Method Using Linear Prediction. **IEICE Transactions on Information and Systems**. [S.l.], v.E89-D, n.2, p.806-813, fev. 2006.

LATHI, Bhagawandas P. **Linear Systems and Signals**. 2ª ed. Nova Iorque: Oxford University Press, 2005.

LATHI, Bhagawandas P.; GREEN, Roger. **Essentials of Digital Signal Processing**. 1ª ed. Nova Iorque: Cambridge University Press, 2014.

LECOMTE, Jérémie *et al.* *Enhanced time domain packet loss concealment in switched speech/audio codec*. In: 2015 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP). 1, 2015, South Brisbane. **Anais...** [S.l.]: IEEE, 2015.

LEE, Minkyu; ZITOUNI, Imed; ZHOU, Qiru. *On a N-gram model approach for packet loss concealment*. In: 8TH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (ICSLP). 8, 2004, Jeju Island. **Anais...** Murray Hill: ISCA, 2004.

MCFEE, Brian *et al.* *Librosa: Audio and music signal analysis in Python*. **PROCEEDINGS OF THE 14TH PYTHON IN SCIENCE CONFERENCE**. **Anais...** [S.l.]: SciPy, 2015.

MARPLE, Larry. A new autoregressive spectrum analysis algorithm. **IEEE Transactions on Acoustics, Speech, and Signal Processing**. [S.l.], v.28, n.4, p.441-454, ago. 1980.

MENEZES, Afonso H. N. *et al.* **Metodologia científica: teoria e aplicação na educação a distância**. Petrolina: Universidade Federal do Vale do São Francisco, 2019.

OPPENHEIM, Alan V.; SCHAFER, Ronald W. **Discrete-Time Signal Processing**. 3ª ed. Upper Saddle River: Pearson Higher Education, 2010.

PERKINS, Colin; HODSON, Orion; HARDMAN, Vicky. *A survey of packet loss recovery techniques for streaming audio*. **IEEE Network**, [S.l.], v.12, n.5, p.40-48, set.-out. 1998.

PROAKIS, John G.; MANOLAKIS, Dimitris G. **Digital Signal Processing: principles, algorithms and applications**. 3ª ed. Upper Saddle River: Prentice-Hall, 1996.

PRODANOV, Cleber C.; FREITAS, Ernani C. **Metodologia do trabalho científico: Métodos e técnicas da pesquisa e do trabalho acadêmico**. 2ª ed. Novo Hamburgo: Feevale, 2013.

RABINER, Lawrence R.; SCHAFER, Ronald W. **Digital processing of speech signals**. 1ª ed. Eaglewood Cliffs: Prentice Hall, 1978.

RASHID, Rozeha A. *et al.* *Security system using biometric technology: Design and implementation of Voice Recognition System (VRS)*. In: 2008 INTERNATIONAL CONFERENCE ON COMPUTER AND COMMUNICATION ENGINEERING. 1, 2008, Kuala Lumpur. **Anais...** [S.l.]: IEEE, 2008.

SALES, Stanley A.; SALES, Nevilde M. R.; SILVA, Marco Antônio L. R. Biometria: Reconhecimento de voz. **Revista de Sistemas e Computação**, Salvador, v.11, n.1, p.11-18, jan.-abr. 2021.

SANNECK, H. *et al.* *A new technique for audio packet loss concealment*. In: GLOBECOM'96. 1996 IEEE GLOBAL TELECOMMUNICATIONS CONFERENCE. 1, 1996, Londres. **Anais...** [S.l.]: IEEE, 1996.

SEYFARTH, Scott; ZHAO, Paul. **Evaluating an automatic speech recognition service**. [S.l.]. 2020. Disponível em: <https://aws.amazon.com/pt/blogs/machine-learning/evaluating-an-automatic-speech-recognition-service/>. Acesso em: 10 jan. 2023.

SHAW, Urjashee; SHARMA, Bobby. *A Survey Paper on Voice over Internet Protocol (VOIP)*. **International Journal of Computer Applications**, [S.l.], v.139, n.2, p.16-22, abr. 2016.

SILVA, Sabrina C. *et al.* Métodos computacionais aplicados em reconhecimento sonoro e biometria por voz. **Colloquium Exactarum**, [S.l.], v.11, n.2, p.78–90, 2019.

SMITH, Steven W. **The Scientist and Engineer's Guide to Digital Signal Processing**. 2ª ed. San Diego: California Technical Publishing, 1999.

SOX. **Página inicial**. [S.l.]. 2015. Disponível em: <https://sox.sourceforge.net/>. Acesso em: 07 nov. 2022.

SPEECHRECOGNITION. **Repositório oficial**. [S.l.]. 2022. Disponível em: <https://pypi.org/project/SpeechRecognition/>. Acesso em: 15 out. 2022.

TOYOSHIMA, Masahiro; SHIMAMURA, Tetsuya. *Packet Loss Concealment for VoIP Based on Pitch Waveform Replication and Linear Predictive Coding*. In: *2014 IEEE ASIA PACIFIC CONFERENCE ON CIRCUITS AND SYSTEMS (APCCAS)*. 2014, Okinawa. **Anais...** [S.l.]: IEEE, 2014.

UNIVERSIDADE FEDERAL DE SANTA CATARINA. **Metodologia da Pesquisa e Elaboração de Dissertação**. 4^a ed. Florianópolis: UFSC, 2005.

UNIVERSITY OF CALIFORNIA. **Voiced and Unvoiced Speech Overview**. Disponível em: <http://www.seas.ucla.edu/dsplab/vus/over.html>. 2015. Acesso em: 01 dez. 2022.

VOIP DO BRASIL. **Estatísticas sobre a tecnologia VoIP**. [S.l.]. 2017. Disponível em: <https://www.voipdobrasil.com.br/blog/estatisticas-voip-telefonip/>. Acesso em: 15 set. 2022.

XIE, Qingsong; WEI, Wei; CHEN, Quji. A new packet loss concealment algorithm in VoIP. In: *2016 IEEE ADVANCED INFORMATION MANAGEMENT, COMMUNICATES, ELECTRONIC AND AUTOMATION CONTROL CONFERENCE (IMCEC)*. 2016, Xi'an. **Anais...** [S.l.]: IEEE, 2016.

YOUNG, Steve J.; CHASE, Lin L. *Speech recognition evaluation: a review of the U.S. CSR and LVCSR programmes*. **Computer Speech & Language**, [S.l.], v.12, n.4, p.263-279, out. 1998.

APÊNDICE A – Frases transcritas a partir dos áudios de referência

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of the way up the juice of lemons makes fine hunch</i>
	2	<i>police to load your left shoulder to sell breeze came across from the sea</i>
	3	<i>the crooked maize field to fool the mouse the show was a flop from the very start</i>
	4	<i>the source of the huge river is the clear spring upon empty helps to pass the evening</i>
	5	<i>a large size and stockings is hard to sell it snowed rained and hailed the same morning</i>
	6	<i>arrivals use to catch pink salmon the ship was torn apart on the sharp grief</i>
	7	<i>the rope pull binders seven books at once race is often served in round balls</i>
	8	<i>the fish twisted and turned on the bench hook the hogs perfect shot corn in garbage</i>
	9	<i>march the soldiers pass the next he'll take the winding path to reach the lake</i>
	10	<i>the king ruled the state in the early days both lost their lives in the raging storm</i>
	11	<i>adding fast leads to wrong sums me in the coat before you go out</i>
	12	<i>restores badly strained and hung limp to fishy to the dark blue background</i>
	13	<i>the frost era passed through the coat mesh wire keeps chicks inside</i>
	14	<i>reid first out loud for pleasure kick the ball straight and follow through</i>
	15	<i>the white road shimmered in the hot sun the boy was there when the sun rose</i>
	16	<i>wipe the grease off his dirty face to the blue fish swam in the tank</i>
	17	<i>help the woman get back to her feet the swan dive was far short of perfect</i>
	18	<i>four hours of steady work faced s, leave the square stone over the fence</i>
	19	<i>stray cat gave birth to kittens assad is a tool used for making boards</i>
	20	<i>the birch canoes lid on the smooth planks no closely the size of the gas tank</i>
	21	<i>a cup of sugar make sweet fledged the box was thrown beside the park track</i>
	22	<i>press the pants and so button on the vest the small pop not a hole in the stock</i>
	23	<i>koppel for the fence and plunge in the girl at the booth sold fifty bonds</i>
	24	<i>the soft cushion broke the man's fall the lacy cow late in the cool cracks</i>
FG	1	<i>no hides it seemed to keep and sad grape juice and water mix well</i>
	2	<i>the fat cats goes by many names he crawled with caroline allege</i>
	3	<i>a clone costs five cents on mondays he takes the oath of office each march</i>
	4	<i>a chink in law allowed aghast to blow raising seven mccabe</i>
	5	<i>the desk as firm on a shaky floor the drop of the rain made a pleasant</i>

		<i>sound</i>
	6	<i>the point of this deal pena's bent and twisted their houses made of rare beef</i>
	7	<i>the lake sparkled in the red hot sun elaine cuts and chan's anyway</i>
	8	<i>it takes you to bring out the order this dale smell of old year lingers</i>
	9	<i>once we stood beside the shore smoke poured out of every crack</i>
	10	<i>no cement awful hard would see the line over the port side</i>
	11	<i>this plank was made for walking on go now and come here later</i>
	12	<i>so can wash most scared away thinks does shine the cost little</i>
	13	<i>does she bore them home by agog the new step down into restless minds</i>
	14	<i>nudged gently but we can now apply this appease I, is going</i>
	15	<i>feed the white man's sunflower seeds much of the story makes good sense</i>
	16	<i>there's a lag between thought and and he took the lead in camp that the whole distance</i>
	17	<i>rhoda paved with sticky tire the sanders over the sill of the old house</i>
	18	<i>it takes a lot of help to finish these take to sarah's is a fair profit</i>
	19	<i>we now have a new pays for shipping raise assailants dealership northward</i>
	20	<i>the thought came early and freed the stream jack the dogs in the court target</i>
	21	<i>the fis guess in some lamb he wrote down a long list of items</i>
	22	<i>the key you designed both did a lot fastened to pin senate side</i>
	23	<i>palestinians is carved around the base the sun came up to lay the eastern sky</i>
	24	<i>and as she plotted out wanders the cold dead person's health incest</i>
MK	1	<i>a whiff of it will cure the most stubborn called to send it now in large amounts is bad</i>
	2	<i>the street to collar random left edge there is a fine hard paying in salty air</i>
	3	<i>the square peg will settle in the round hole the knife was hot inside the bread sheath</i>
	4	<i>poached eggs and he must suffice the farmer swapped his horse for brown locks</i>
	5	<i>on the frost is common is time for tricky the mule trod the treadmill day and night</i>
	6	<i>clean with the lawyer to drop the lost cause the roof should be tilted and a sharp slant</i>
	7	<i>as a matter of french is worse than on the latch on the back gate needed a nail</i>
	8	<i>green moss grows on the northern side greet the new guests and leave quickly</i>
	9	<i>the ram scare the school children all of this deal biscuits and strong drink</i>
	10	<i>dusty trail is painful for feet crochet before you jump for miss the mark</i>
	11	<i>the sink is the thing in which we prohibitions the loss of the cruiser was a blow to the fleet</i>
	12	<i>sit on the porch and tell the others what to do with the break to the left</i>

		<i>and then over</i>
	13	<i>but a kid's and don't forget the salt t, into china has a sweet taste</i>
	14	<i>essex comes up more often than attend this lab was hewn from heavy box of slate</i>
	15	<i>pitched the straw through the door of the stable and gold vases both rare and costly</i>
	16	<i>the team with the best timing looks good to them before the boy could see it</i>
	17	<i>that nerves are jangle by the door slam the aim of the contest is to raise the great fun</i>
	18	<i>there's a strong chance it will happen once more the duke left the park in a sober coach</i>
	19	<i>post no bills on this office wall is saying the same teams in each purdy</i>
	20	<i>accrues in warm water isn't asleep yacht is fond terror thin sheet from the yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight from the old market</i>
	22	<i>the pods appease ferment in their fields pile the call high in the shed corner</i>
	23	<i>he lent his coach to the tall gaunt stranger bush for and roll on the lofty rocks</i>
	24	<i>cod is the main business of the north shore times shower down from all sides</i>
ML	1	<i>it was about error on the part of the new judge the night shift man who rejects repaired</i>
	2	<i>take some action strike against to shoot the puppet jerked the leash as he saw feline shape</i>
	3	<i>the steady drip is worse than drenching rain kept the juror with a tight brass cover</i>
	4	<i>the corner store was robbed lost on the big red apple fell to the ground</i>
	5	<i>the red paper brightened the dim stage that no powdered on the brink brown grass</i>
	6	<i>many hands help to the job done see the player scooped the third base</i>
	7	<i>very nice frosted the punch bowl the streets are narrow and full of sharp turns</i>
	8	<i>and tinsel from both branches we don't like to admit her small faults</i>
	9	<i>a severe storm tore down the barn the bombs left also the town in ruins</i>
	10	<i>the good book and forms of what we ought to know open a book to the first page</i>
	11	<i>the big show was found on the sandy beach fish evade the net and swim all</i>
	12	<i>the pot boiled of the contents fail to jail the poor boy who stumbled again</i>
	13	<i>the baby puts his right foot in his mouth the gloss on top they didn't fit to read</i>
	14	<i>she saw it happen in the neighbor's house put the tournaments all the target them</i>
	15	<i>no doubt about the way the wind blows small children came to see you</i>
	16	<i>dig deep in the earth for pirate's gold gold ring will please most of the</i>

	<i>girl</i>
17	<i>you cannot brewed tea in a cold pot bust of terror slipped from the moving van</i>
18	<i>the stitch full surf that needs to be shorn to be sure to support from when the whole</i>
19	<i>leave now and you will arrive on time that the bill wants and what it so</i>
20	<i>stop and stare at the hard working man one step more on the board will collapse</i>
21	<i>the young prince became heir to the throne the first part of the plan exchanges</i>
22	<i>we please answer that phone the store was jammed before the silk to start</i>
23	<i>the grass and bushes wrote what was to the milk comes in three bachus per day</i>
24	<i>slide the bill between the two leaves thoughts like betrayed the black cat</i>

APÊNDICE B – Frases transcritas a partir dos áudios acometidos por picotes

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of the way up the juice lemon snakes fine punched</i>
	2	<i>police to load your left shoulder to sell priest came across from the sea</i>
	3	<i>the crooked nice failed to fool the mouse the shell was a flop from the very start</i>
	4	<i>the source of the huge river is that clear spring upon after he helps to pass the evening</i>
	5	<i>a large size and stockings is hard to sell it snowed rain and hail of the same morning</i>
	6	<i>rogers to catch pink salmon the ship was torn apart on the sharp relief</i>
	7	<i>they're hopeful binders seven books at once graced is often served in the round balls</i>
	8	<i>the fish twisted and turned on the bench hope the hogs perfect shot corn and garbage</i>
	9	<i>march the soldiers pass the next he'll take the winding path to reach the lake</i>
	10	<i>the king ruled the state in your early days both lost their lives in the region storm</i>
	11	<i>i think fast leads to wrong sums in the club before you go out</i>
	12	<i>reso's badly strained and hung limp the fishy to the dark blue background</i>
	13	<i>the frost they are passed through the coat mesh wire keeps chicks inside</i>
	14	<i>reverse and loud for pleasure kick the ball straight and follow through</i>
	15	<i>the white road shimmered in the hot sun the boy was there when the sun rose</i>
	16	<i>wipe the grease off his dirty face to unfinished swam in the tank</i>
	17	<i>help the woman get back to her feet the swan dive was far short of perfect</i>
	18	<i>four hours of steady work faced that's leave the square stone over the fence</i>
	19	<i>stray cat keeper it hit once i saw as a tool used for making boards</i>
	20	<i>the birch canoe slid on the smooth planks no look closely decisively gas tank</i>
	21	<i>a cup of sugar make sweet french the box was thrown beside the park track</i>
	22	<i>press the penance and so a button on the vest the small talk not a hole in the stock</i>
	23	<i>hop over the fence and plunge and the girl at the booth sold fifty bonds</i>
	24	<i>the soft cushion broke demands for all the lacy kowal eighteen to coop kraft</i>
FG	1	<i>no hides it seemed to keep and sad bridges and water mix well</i>
	2	<i>the fat cats goes by many names the chronic care along on the edge</i>
	3	<i>the club costs five cents on monday as he takes the oath of office each march</i>
	4	<i>a chink in the law unleaded gas to blow her as a separate lack of beauty</i>

	5	<i>the desk is feminist shaky floor the drop of the rain in a pleasant sound</i>
	6	<i>the point of this deal pena's bent and twisted and hatches may never be</i>
	7	<i>the lakes sparkled in a red hot sun elaine cuts in chen's anyway</i>
	8	<i>it takes you to bring up the odor is still smell of old dear lingers</i>
	9	<i>once we stood beside the shore smoke poured out of every crack</i>
	10	<i>no segment will tired would see in the line over the port side</i>
	11	<i>this plant is made for walking on go now and come here later</i>
	12	<i>so can wash most steered away fakes don't shine the cost little</i>
	13	<i>the cheaper than home by adopt the news stockdale interest was minds</i>
	14	<i>nudged gently but we can now apply his lucky zealots going</i>
	15	<i>feed the white man's sunflower seeds much of the story makes good sense</i>
	16	<i>there's a lag between thought and that he didn't believe in capital distance</i>
	17	<i>rhoda paved with sticky tire the center's over the sill of the old house</i>
	18	<i>it takes a lot of help to finish these take to share is a fair profit</i>
	19	<i>we now have a new base for shipping raised assailants dealership northward</i>
	20	<i>the thought came early and freed the stream jack the back from the court target</i>
	21	<i>the fist guess in some lamb you know down a long list of items</i>
	22	<i>it's you designed both a block fastened to consented side</i>
	23	<i>palestinians is carter on the base the sun came up to let the eastern sky</i>
	24	<i>tend sheep what about wanders the cold dead mistress happens next</i>
MK	1	<i>but with a bit will cure the most stubborn called to send it now in large amounts is bad</i>
	2	<i>a streak of color random left edge is a fine hard paying insult either</i>
	3	<i>the square peg will settle in the round hole the knife was hot inside the bag sheath</i>
	4	<i>poached eggs in teams must suffice the farmers swapped his horse for brown locks</i>
	5	<i>on the frost as calm as time for tricky the mule trod the treadmill day and night</i>
	6	<i>clean with the lawyer to drop the lost cause the roof should be tilted at a sharp slut</i>
	7	<i>asked whether a french is worse than on the left i'm a back gate needed a nail</i>
	8	<i>green moss grows on the northern side greet the new guests leave quickly</i>
	9	<i>the ram scare the school children off the hook the steel biscuits and strong drink</i>
	10	<i>that's the trail was painful for feet kosher before you jump or miss the mark</i>
	11	<i>the sink is the thing in which we prohibitions the loss of the cruiser was a blow to the fleet</i>
	12	<i>sit on the purchase to lead others what to do with the break to the left and then over</i>
	13	<i>pack the kids and don't forget the salt t, into china has the sweet taste</i>

	14	<i>six comes up more often than attend the slab with hewn from heavy box of slate</i>
	15	<i>picked the strove to the door of the stable and cold cases both very costly</i>
	16	<i>the team with the best timing looks good the gun before the boy could see it</i>
	17	<i>and there's a jungle by the door slam the game of the contest is to raise the great fun</i>
	18	<i>there's a strong chance it will happen once more that you quit the park in a silver coach</i>
	19	<i>post no bills on this office wall is setting this thing teams in each party</i>
	20	<i>cruising warm waters and asleep yacht is fun to thin sheet from the yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight from the old market</i>
	22	<i>the pods appease ferment in their fields political high in the shed corner</i>
	23	<i>you and his coach to the tall got stranger bush for and roll on the lofty rocks</i>
	24	<i>cod is in the business of the north shore times shower down from all sides</i>
ML	1	<i>it was about it or on the part of the new judge an adjustment rid occupied</i>
	2	<i>take some action and strict begins to shoot the pop jerked bleach that you saw a few aren't cheap</i>
	3	<i>the study group is worse than drenching rain kept the juror was a tight brats cover</i>
	4	<i>the corner store was robbed laughs but the big medical fall to the ground</i>
	5	<i>the red paper brightened the dumb stage that no powdered on the brink brown grass</i>
	6	<i>many hands out of the job don't see the player scooped up to third base</i>
	7	<i>the next frosted the punch bowl the streets are narrow and full of sharp turns</i>
	8	<i>and tinsel from both directions we don't like to admit her small folds</i>
	9	<i>a severe storm tore down the bar and the bombs but most of the town in ruins</i>
	10	<i>the good book and forms of what we ought to know what the book to the first pitch</i>
	11	<i>the big show was found on the sandy beach fish evade the net and swim off</i>
	12	<i>the pot boiled of the continents fiddle to tell the poor boy witnessed the book again</i>
	13	<i>the baby puts his right foot in his book the gloss on top the gun fit to read</i>
	14	<i>she saw a cap in the neighbor's house put the turn of an adult and target them</i>
	15	<i>no doubt about the way the wind blows small children came to see it</i>
	16	<i>dig deep in the earth for parts goal of gold ring will please most of the girl</i>

17	<i>you cannot bertucci and called on the stuff to slip from a moving van</i>
18	<i>the stitch will serve but needs to be shortened assured of support from legal</i>
19	<i>leave now and you'll arrive and i took the pill once and let it snow</i>
20	<i>stop and stare at the hardworking and one step more than the board will collapse</i>
21	<i>the young prince became heir to the throne the first part of the poor and exchanges</i>
22	<i>we please answer that phone the store was jammed before the civil to start</i>
23	<i>the grass and bushes were what what was to the milk comes in three bunches per day</i>
24	<i>slide the bill between the two dudes dots of light betrayed the black cat</i>

APÊNDICE C – Frases transcritas a partir dos áudios corrigidos pelo algoritmo I

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of the whale the juice of lemons makes fine hunch</i>
	2	<i>police to load your left shoulder to sell breeze came across from the sea</i>
	3	<i>the crooked nice failed to fool the mouse the show was a flop from the very start</i>
	4	<i>the source of the huge river is the clear spring upon a team helps to pass the evening</i>
	5	<i>a large size and stockings is hard to sell it snowed rain and hail does seem morning</i>
	6	<i>arrivals use to catch pink salmon the ship was torn apart on the sharp reach</i>
	7	<i>the rope pull binders seven books at once graced is often served in round balls</i>
	8	<i>the fish twisted and turned on the bench hope the hogs perfect shot corn in garbage</i>
	9	<i>march the soldiers pass the next he'll take the winding path to reach the lake</i>
	10	<i>the king ruled the stake in the early days both lost their lives in the region storm</i>
	11	<i>adding fast leads to wrong sums me in the coma before you go out</i>
	12	<i>arrest was badly strained and hung limp to fishy to the dark blue background</i>
	13	<i>the frosty air passed through the coat mesh wire keeps chicks inside</i>
	14	<i>reid first out loud for pleasure kick the ball straight and follow through</i>
	15	<i>the white road shimmered in the hot sun the boy was there when the sun rose</i>
	16	<i>wipe the grease off his dirty face to blue fish swam in the tank</i>
	17	<i>help the woman get back to her feet the swan dive was far short of perfect</i>
	18	<i>four hours of steady work faced s, leave the square stone over the fence</i>
	19	<i>stray cat keeps kittens assad is a tool used for making boards</i>
	20	<i>the birch canoe slid on the smooth planks no closely the size of the gas tank</i>
	21	<i>a cup of sugar make sweet french the box was thrown beside the park track</i>
	22	<i>press the pants and so a button on the vest the small pop not a hole in the stock</i>
	23	<i>hop over the fence and plunge in the girl at the booth sold fifty bonds</i>
	24	<i>the soft crushing broke the man's fall the lacy cow late in the cool grass</i>
FG	1	<i>no hides it seemed to keep and sad bridges and water maxwell</i>
	2	<i>the fat cats goes by many names the crowd with caroline allege</i>
	3	<i>a clone costs five cents on mondays he takes the oath of office each march</i>
	4	<i>it's taken along allowed a gas to blow racing seven lack of beauty</i>
	5	<i>the desk as firm on a shaky floor the drop of the rain made a pleasant</i>

		<i>sound</i>
	6	<i>the point is this still pena's bent and twisted their houses made of rare piece</i>
	7	<i>the lake sparkled in the red hot sun laid cuts and chen's anyway</i>
	8	<i>it takes me to bring out the order of the scales now a cold beer lingers</i>
	9	<i>once we stood beside the shore smoke poured out of every crack</i>
	10	<i>no cement awful hard would see in the line over the port side</i>
	11	<i>this plank was made for walking on go now and come here later</i>
	12	<i>so pugwash most scared away extends shine the cost lil</i>
	13	<i>the sheep one at home by about the news that getting tresses minds</i>
	14	<i>nudged gently but we can now apply this wimpy zealous going</i>
	15	<i>feed the white man's sunflower seeds much of the story makes good sense</i>
	16	<i>there's a lag between thought and that he didn't believe in capital distance</i>
	17	<i>no they're paid twisted guitar the standards over the sill of the old house</i>
	18	<i>it takes a lot of help to fantasies take to sarah's is a fair profit</i>
	19	<i>we now have a new base for shipping raised assailants dealership northward</i>
	20	<i>the thought came early and frida stream jack the dogs in the court target</i>
	21	<i>the fiske is in some lamb he wrote down a long list of items</i>
	22	<i>it seemed designed both et de la fastened to cannes and each side</i>
	23	<i>palestinians is carved around the base the sun came up to like the eastern sky</i>
	24	<i>tend sheep what about wanders a cloth diapers stars have incest</i>
MK	1	<i>a whiff of it will cure the most stubborn called to send it now in large amounts is bad</i>
	2	<i>a streak of color random left edge is a fine hard paying in salty air</i>
	3	<i>the square peg will settle in the round hole the knife was hot inside the bread sheath</i>
	4	<i>poached eggs and he must suffice the farmer swapped his horse for brown locks</i>
	5	<i>on the frost as calm as time pritchard keep the mule trod the treadmill day and night</i>
	6	<i>clean with the lawyer to drop the lost cause the roof should be tilted and a sharp slant</i>
	7	<i>as a matter of french is worse than on the latin a back gate need to nail</i>
	8	<i>green moss grows on the northern side greet the new guests and leave quickly</i>
	9	<i>the ram scare the school children all of this deal biscuits and strong drink</i>
	10	<i>dusty trail is painful for feet crochet before you jump for miss the mark</i>
	11	<i>the sink is the thing in which we prohibitions the loss of the cruiser was a blow to the fleet</i>
	12	<i>sit on the persian tell the others what to do with the break to the left and then over</i>
	13	<i>how the kids and don't forget the salt team thing china has a sweet taste</i>
	14	<i>at six comes up more often than attend this lab was hewn from heavy</i>

		<i>box of slate</i>
	15	<i>pitched the straw through the door of the stable and golden bases both rare and costly</i>
	16	<i>the team with the best timing looks good to them before the boy could see it</i>
	17	<i>that nerves are jangle by the door slam the aim of the contest is to raise the great fun</i>
	18	<i>there is strong chance it will happen once more the duke let the park in a sober coach</i>
	19	<i>post no bills on this office wall is saying the same teams in each purdy</i>
	20	<i>cruzan warm waters in the sleep yacht is fine terror thin sheet from the yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight from the old market</i>
	22	<i>the pods appease ferment in their fields pilot cool high in the shed corner</i>
	23	<i>he lent his coach to the tall got stranger bush for and roll on the lofty rocks</i>
	24	<i>ca doesn't mean business of the north shore times shower down from all sides</i>
ML	1	<i>it was about error on the part of the new judge than adjustment rid occupied</i>
	2	<i>take some action strict adherence to issue the pop jerked felicia aziz saw feeling shape</i>
	3	<i>the study group is worse than drenching rain kept the juror was a type brass cover</i>
	4	<i>the corner store was robbed lost on the big red apple fell to the ground</i>
	5	<i>the red paper pregnant for the stage that no powdered on the brick brown grass</i>
	6	<i>many hands out of the job done see the player scooped to third base</i>
	7	<i>very nice frosted the punch bowl the streets are narrow and full of sharp turns</i>
	8	<i>and tinsel from both branches we don't like to admit her small faults</i>
	9	<i>a severe storm tore down the barn the bombs but also the town in ruins</i>
	10	<i>a good book and forms of what we ought to know what the new book to the first page</i>
	11	<i>pumping show was found on the sandy beach fish evade the net and swivel</i>
	12	<i>the pot boiled of the contents fail to jail the poor boy stumbled again</i>
	13	<i>the baby puts his right foot in his mouth because our top they didn't fit to read</i>
	14	<i>she saw it happen in the neighbor's house put the tournaments will attack it though</i>
	15	<i>no doubt about the way the wind blows small children came to see you</i>
	16	<i>dig deep in the earth for pirate's gold gold ring will please most of the girl</i>
	17	<i>you cannot grucci and called on the staff terror slipped from the move in</i>
	18	<i>the stifel surf that needs to be sure and assured us that the weapon firmly in the hole</i>

19	<i>leave now than you will arrive and i took the pill once and levitz hello</i>
20	<i>stop and stare at the hard working man one step more than the board will collapse</i>
21	<i>the young prince became heir to the throne the first part of the plan needs to</i>
22	<i>we please answer that phone the store was jammed before the silk to start</i>
23	<i>the grass and bushes were wet wet with two the malcolm's in three bunches per day</i>
24	<i>slide that bill between the two leads but slight betrayed the blackout</i>

APÊNDICE D – Frases transcritas a partir dos áudios corrigidos pelo algoritmo II

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of the juice lemon snakes fine hunch</i>
	2	<i>police that she your left shoulder that's all these came across from the sea</i>
	3	<i>the crooked nice failed to fool the mouse the show was a flop for the very start</i>
	4	<i>the source of the huge river is clear spring upon of t, helps to pass the evening</i>
	5	<i>a large size and stockings is hard to sell its knew it would rain and hail the singing morning</i>
	6	<i>rod was used to catch pink salmon the ship was torn apart on the shore brief</i>
	7	<i>the rope by the seven books at once graced is off in search of the rich and bought with</i>
	8	<i>the fish twisted and turned on the band who took the hots with fed chopped court in garbage</i>
	9	<i>march the soldiers pass the next hill to the winding path to reach the lake</i>
	10	<i>the king unfurled the state in the early days for flustered lies in the region storm</i>
	11	<i>adding fast leads to a wrong saddam's into the club before you go out</i>
	12	<i>restores badly strained and hung limp to fishy to the dark blue background</i>
	13	<i>the frost a year passed through the coat mesh water keeps chicks inside</i>
	14	<i>reverse atlanta for pleasure kick the ball sturdy and follow through</i>
	15	<i>the white road she emerged in the hot sun in the boy was there when the sun rose</i>
	16	<i>wipe the grease off his dirty face to the blue fish swam in the tank</i>
	17	<i>help the woman get back to her feet the swan dive was far short of perfect</i>
	18	<i>four hours of steady work faced us leave the square stoughton or defense</i>
	19	<i>stray cat keeper kittens assad is a tool used for making boards</i>
	20	<i>the birch canoes laid out the smooth planks no closely decisively gas tank</i>
	21	<i>a cup of sugar to make sweet fledged the box was thrown out beside the track</i>
	22	<i>press the parents and so a button on the vest this notebook not a hole in the stock</i>
	23	<i>hop over the fence a plunge and the girl at the booth sold fifty buttons</i>
	24	<i>the sock cushion for the man's fault the lacy couch lady in the cool grass</i>
FG	1	<i>no hides it seemed to keep and bridges and water mix well</i>
	2	<i>the fat cats goes by maintaining its peak robert k, along on the edge</i>
	3	<i>the club costs by sense and monday's to the oath of office each march</i>
	4	<i>it's a good while it just a little risley said that the d,</i>
	5	<i>the guests from a shaky flirt the drop of the three and a presence that</i>

		<i>on</i>
	6	<i>the point is to stay up and husbands and twisted ashes make it better be</i>
	7	<i>blake's article dunham at hudson constituents and he would</i>
	8	<i>take the divert the older this does not all of the ringgit</i>
	9	<i>what he stood beside a short smoke poured out of every crack</i>
	10	<i>president cold hearted you know i know that upwards i</i>
	11	<i>this plant is made for walking out on the lead come here later</i>
	12	<i>so you can watch closed it lets a fixed and shot and a costly</i>
	13	<i>he put that whoa dead i used a cadet interest was minds</i>
	14	<i>and instantly but we can now apply the key is i was and</i>
	15	<i>feed the white man's sunflower seeds much of the story makes good sense</i>
	16	<i>there's a lag between art and i he took the lead in camp but the club distance</i>
	17	<i>the die paid twisted the tyre this and if so that is still the old house</i>
	18	<i>it takes a lot of help defend us these take to share is a fair profit</i>
	19	<i>now into bits per second which is this a balance to shed northward</i>
	20	<i>the thought came early and freed the stream jack back from the court target</i>
	21	<i>the fiske is said and some like you took out a long list items</i>
	22	<i>that you designed both the de la fastened to cannes any side</i>
	23	<i>this sydney mrs carter on the base the sun came up to like the eastern sky</i>
	24	<i>and the sheep what about wanders conte star as happens next</i>
MK	1	<i>with good look you to the most urgent call to send it out and has died</i>
	2	<i>the street to college or threatened unless damage is a fine art hanging saw peter</i>
	3	<i>the scooter pettengill settling the round hole than typhus hong inside the bag she</i>
	4	<i>poached eggs in th must suffice the farmers to scores for brought an ox</i>
	5	<i>the frost has come to his temperature to keep the mule trod the treadmill day and night</i>
	6	<i>clean with the loiter to draw that was caught the two should be tilted and sharp slut</i>
	7	<i>this matter to french is worse than on the lot on the back he didn't nail</i>
	8	<i>it was gross on that and so deep the new guest and leave quickly</i>
	9	<i>the bran scare the school children are acoustical biscuits and strong</i>
	10	<i>the trail has been for feet crush before you jump for miss the mark</i>
	11	<i>the sink is the thing in which we have to shift the loss of the crew there was a blow to the fleet</i>
	12	<i>sit on the projectile each other's what to do with the body to the left and bent over</i>
	13	<i>how the kids and don't forget the salt teaching china has the sweet taste</i>
	14	<i>six comes out the door opened and attend this was huge from heavy box of slate</i>
	15	<i>pitched the straw through the door of the stable of gold of the sun boat returned costly</i>

	16	<i>the team with the best time it looks good to them before the public could see it</i>
	17	<i>that nurses and ankle biter doors for the love the contest is to raise agreed on</i>
	18	<i>this is a strong chance of what happened once before the g, clef the initial coach</i>
	19	<i>postal bills on the sock was all besetting to think and teams in each party</i>
	20	<i>cruising towards waters and asleep yacht is fun to thin sheet from the yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight for the old guard</i>
	22	<i>the pods a piece for men to interview political hind should grow</i>
	23	<i>he and his coach to the tall dark street under which further and ropes on the lofty rocks</i>
ML	24	<i>condescending to business the north shore and shot or down from all sides</i>
	1	<i>it was about her on the part of the new judge and adjustment rejection and</i>
	2	<i>take some action strictly construed to the jerked polish as sought you can shape</i>
	3	<i>the study group is worse than trudging and kept the daughter with that ross cutter</i>
	4	<i>the corner store was robbed last but the big apple folk took it out</i>
	5	<i>the red paper brightened the dim stage that no patent on the bird progress</i>
	6	<i>the cans of the job in a supporter stooped to third base</i>
	7	<i>putting snakes frogs to the punch the streets are undertow of her parents</i>
	8	<i>and to insult from both directions we don't like to admit sport folks</i>
	9	<i>a severe storm poured out about the bombs which will salute and newton's</i>
	10	<i>good good fortune so what we ought to go to your book to the first page</i>
	11	<i>the big show was found on a sandy beach fish is good the net and scuttle</i>
	12	<i>the pot boy told of the contents fitted to the poor boy witnessed the book again</i>
	13	<i>the beta puts his right foot in his book because our top good fit to read</i>
	14	<i>she saw the cat in the neighbor's house with the chart and gentle and tie it up</i>
	15	<i>what about the way the wind blows small children to see you</i>
	16	<i>the deep in the earth for that school of gold and what please note the bigger</i>
	17	<i>you cannot bertucci and called on a stop to slip from a moving there</i>
	18	<i>the stitch pulsar tough but nice to be sure to push it to wear it for legal</i>
	19	<i>we've got all the right i'm john carter did the billboards and let it snow</i>
	20	<i>stop and stare at the hardworking and one step more than the board will collapse</i>
	21	<i>the can speak it or to throw the first part of the poor and the exchanges</i>

	22	<i>would you please answer that phone the store was dead and both of us up to start</i>
	23	<i>the grass and bushes were witness to do that all comes in three to spurted</i>
	24	<i>slotted the battle between the two leagues but to the country but</i>

APÊNDICE E – Frases transcritas a partir dos áudios corrigidos pelo algoritmo III

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of whale the juice lemon snakes fine hunch</i>
	2	<i>place to go to your left shoulder subtle priest came across from the city</i>
	3	<i>the crooked nice failed to fool the mouse the shower was a flop from the very start</i>
	4	<i>the source of the huge character is a clear spring upon a team helps to pass the evening</i>
	5	<i>large size and stockings is hard to settle it snowed and inhale the scene morning</i>
	6	<i>or auditions to catch pink salmon the ship was torn apart on the shore brief</i>
	7	<i>the rope about five to seven books at once graced pays off in search of the round</i>
	8	<i>the fish twisted and turned on the bench who took the han's perfect shot court in garbage</i>
	9	<i>march the soldiers as the next hill to winding path to reach the leak</i>
	10	<i>a kangaroo the state in the early days of lost their lives in the region still term</i>
	11	<i>adding fast leads to wrong saddam's in the coat before you go out</i>
	12	<i>reso's badly strained and hung limp can fishy to the dark blue background</i>
	13	<i>the frost error passed through the cloaked mesh wire keeps chicks inside</i>
	14	<i>the first atlanta for pleasure kick the ball straight and follow through</i>
	15	<i>the book i drop would shimmered in the hot sun the boy was a dear one send arose</i>
	16	<i>what the grease off his dirty face to blue fish swam in the tank</i>
	17	<i>help the one in the back to her feet the swan dive is far short of perfect</i>
	18	<i>four hours of steady work faced death if the squares to knock over the fence</i>
	19	<i>st khaki priest kittens us our is a tool used for making boards</i>
	20	<i>the birch canoe slid out the smooth planks york closely decisive the gas tank</i>
	21	<i>a cup of sugar to make sweet fledged the box was thrown beside the truck</i>
	22	<i>press the pants and celibate and addressed this blood not a hole in the sock</i>
	23	<i>over the fence and plunge and the girl at the booth sold fifty buttons</i>
	24	<i>the stock cushion per demands for all lacey cattle taken the cruel press</i>
FG	1	<i>no hides it seemed to keep his ass and bridges and water mix well</i>
	2	<i>the fat cats goes by many names the crowd and carry out on the edge</i>
	3	<i>cotton costs by sand and monday's he takes the oath at the top fifty to march</i>
	4	<i>a chink in a letter drafted little prison and said that the beauty</i>
	5	<i>the desk is from a shaky floor the drop of the race and it doesn't stand on</i>
	6	<i>the point of this stupendous and twisted catches me every week</i>

MK	7	<i>the lake sparkled and matt hudson and it cuts in cannes anyway</i>
	8	<i>it takes to bring up the older this does not colby and lingers</i>
	9	<i>once he stood beside the short smoke poured out of every crack</i>
	10	<i>president awful hard would you delight and over the port side</i>
	11	<i>this plank as they are talking i'm not one competitor</i>
	12	<i>subplots most way thinks does shot and the cost a little</i>
	13	<i>if she bled at home and to the news that get inter isis minds</i>
	14	<i>next we can apply to the peace with jordan</i>
	15	<i>ito claim as sunflower seeds but to the story makes good sense</i>
	16	<i>there's a lag between art and add to believe in capital distance</i>
	17	<i>so the paved with sticky tire this and it's over the skill of the old house</i>
	18	<i>it takes a lot of help defend us these two shares is a fair profit</i>
	19	<i>when i'm in today's disturbing resistance to ship northward</i>
	20	<i>the bucking the the humphrey de stream jack the back from the court target</i>
	21	<i>the escape as it turns out here don't get on a long list of items</i>
	22	<i>that you designed both the block thousand two cans on each side</i>
	23	<i>listen to me and says carter on the base this on teamed up to like the eastern sky</i>
	24	<i>and the sheep the top honors the club that the stars have incest</i>
	1	<i>with a bit will cure the most open called to send in only lot judgment is bad</i>
	2	<i>the streak of color random to lift it is a fine art hanging thought peter</i>
	3	<i>the square pettengill saddle and rat on hold and typhus inside but she</i>
	4	<i>poached eggs and he must suffice of partners bought his horse for brought to my looks</i>
	5	<i>the frost has come to this time for tricky the ultra the treadmill day and night</i>
	6	<i>to it though goiter to drop the lost cause that would the tilted the sharp's lot</i>
	7	<i>astrologer trenches were stunned and locked on the back even if nato</i>
	8	<i>the most ruthless and assad greeting guests and quickly</i>
	9	<i>the red and scare the school children the steel biscuits and strong drink</i>
	10	<i>a speech it was painful for feet pushed before you jump or miss the mark</i>
	11	<i>the stink is the thinking which we prohibitions the loss of the crew's it was a blow to the fleet</i>
	12	<i>sit on the project we don't just what to do the break to the left and then over</i>
	13	<i>the tickets and don't forget the salt t, into the china has the sweet taste</i>
	14	<i>essex comes up more often than attend islam was here to aunt hetty box of slate</i>
	15	<i>picked the straw through the door of the stable golden cases both the turn costly</i>
	16	<i>the team with the best time looks good and unaffordable to concede</i>
	17	<i>anders and angle by the door slam the of the contest is to raise the great fun</i>
	18	<i>that is a strong chance it will happen once more that you quit the party</i>

		<i>miss it up</i>
	19	<i>post no bills on this office or besetting the thing teams in each party</i>
	20	<i>crews and warm waters and asleep yacht is fun to a thing she comes a yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight for the gold market</i>
	22	<i>the pods apiece for demented and airfields tell the court china should</i>
	23	<i>he went his coach to the tall gaunt street under bush for drop on the lofty rocks</i>
	24	<i>what is in the business of the north shore and shot or down from all sides</i>
ML	1	<i>it was about an utter on the part of the new judge an adjustment reattach it</i>
	2	<i>to commit actions picnics and trish to the pot jerked leash as you saw a few in shape</i>
	3	<i>the study group is worse to drenching rain kept the jar with the broad scott</i>
	4	<i>the corner store was robbed last month the grammatical filled it out</i>
	5	<i>the red paper brightened the dumb stage than a hundred in the brig russians</i>
	6	<i>the hands up to the job done supporter stooped to thirty days</i>
	7	<i>critics frosted the punch the streets are turtle and full of sharp turns</i>
	8	<i>thanks and stole from both directions we don't like to deter small faults</i>
	9	<i>a severe storm tore it apart and the bombs but also the total tunes</i>
	10	<i>a good book and forms of what we ought to know what to do go to the first pitch</i>
	11	<i>the big show was found on us and each fish diverted and bands what'll</i>
	12	<i>the pot boy told of the contents of the dough or balkans the book again</i>
	13	<i>the data that says right foot and is now the boss on top that hit the</i>
	14	<i>she saw it happen in that first house with the chart and so on target that</i>
	15	<i>not in about the way the wind blows small children can kiss the cut</i>
	16	<i>dig deep in the earth for battered school of gold and all please most amigo</i>
	17	<i>you cannot bertucci been called the stuff to turn slipped from linda</i>
	18	<i>the stitch certify that needs to be sure to be sure to support efforts would go</i>
	19	<i>we've got all the old arrived on time to cut a old ones and limits</i>
	20	<i>stop and stare at the hardworking and once at or on the board will collapse</i>
	21	<i>the young prince to cater to the throat on the first part of the plan needs to do</i>
	22	<i>we please answer that phone the storm to stand before the senate to start</i>
	23	<i>the grass and bushes were witness to the tokens and three back to spurted</i>
	24	<i>slide the battle between the two dudes that's like drinkable arquette</i>

APÊNDICE F – Frases transcritas a partir dos áudios corrigidos pelo algoritmo IV

Locutor	Nº	Frase transcrita
FA	1	<i>it's easy to tell the depth of the juice lemon snakes fine hunch</i>
	2	<i>place the tier left shoulder the subtle priest came across from the sea</i>
	3	<i>the crooked nice failed to fool the mouse the shower was a flop from the very start</i>
	4	<i>the source of the huge crater and clear spring upon hefty helps to pass the evening</i>
	5	<i>large size and stockings is hard to say how it's new to it rained and hail the scene morning</i>
	6	<i>arrivals used to catch pink salmon the ship was torn apart on the sharp reef</i>
	7	<i>the rich or poor guy to seven books at once graced is often served in renewed call for us</i>
	8	<i>the fish twisted and turned on the bench who took the hans what fed chopped court in garbage</i>
	9	<i>march the soldiers pass the next hill to winding path to reach the lake</i>
	10	<i>a kangaroo to stake in the early days both lost their lives and the raging storm</i>
	11	<i>adding fast leads to wrong saddam's into the code before you go out</i>
	12	<i>this is badly strained and hung limp to fishy to the dark blue background</i>
	13	<i>the frost error passed through the coat measure letter keeps chicks inside</i>
	14	<i>reid first atlanta for pleasure kick the ball to stir the follow through</i>
	15	<i>the white road shimmered in the hot sun the boy was a gator when the sun rose</i>
	16	<i>what the grease off his dirty face to the blue fish swam in the tank</i>
	17	<i>help the woman get back to her feet the swan dive was far short of perfect</i>
	18	<i>four hours of state work faced us leave the square starting over the fence</i>
	19	<i>stray cat keeper kittens assault on as a tool used for making boards</i>
	20	<i>the birch canoes laid out the smooth planks that look closely decisive the gas tank</i>
	21	<i>a cup of sugar to make sweet fledged the box was thrown beside the truck</i>
	22	<i>press the parents and so abundant addressed this blood not a hole in the sock</i>
	23	<i>hop over the fence and plunge in the girl at the booth sold fifty buttons</i>
	24	<i>the stock cushion brooke demands for all of the sea cow late in the cool grass</i>
FG	1	<i>no hides it seemed to keep hands and bridges and water to mix well</i>
	2	<i>the fat cats close but maintains economists care so long to it</i>
	3	<i>the club costs by sunset monday's to the and opposite each march</i>
	4	<i>she and i let out a draft of the reason except i could be</i>
	5	<i>the desk as for the shaky floor the drop of the rain presence that on</i>
	6	<i>the point is this the apprentice and twisted ashes made a rare piece</i>
	7	<i>blake's article and odds and huntington's anyway</i>

MK	8	<i>did he do but now the older this does not all be earning it</i>
	9	<i>what he stood beside a short smoke poured out of every crack</i>
	10	<i>president will heart would you do on the ports i</i>
	11	<i>this plant is a robot keen on an island under later</i>
	12	<i>so the snow stared way it stands shined a costume at all</i>
	13	<i>she put at home and on the step down interest was minds</i>
	14	<i>and and standing we can apply the pizza with and</i>
	15	<i>but why does it sunflower seeds but to the story makes good sense</i>
	16	<i>there's a lag between art and add to believe in capital distance</i>
	17	<i>don't pay twisted the tire this and if so that is still the old house</i>
	18	<i>it takes a lot of help defend us these dictators is a fair profit</i>
	19	<i>now and who pays the shipping business the stena ship northward</i>
	20	<i>the thought he had an frida stream jack back from the court target</i>
	21	<i>the disgusted and sunlight here took down a long list of items</i>
	22	<i>it seemed designed both the block fastened to pin sunnyside</i>
	23	<i>listen to and says carter on the base to send him up to like the eastern sky</i>
	24	<i>and she bought the top honors club that the stars have incest</i>
	1	<i>with a bit look you to the most urgent call to send it to go into my judgment is bad</i>
	2	<i>a streak of color are and unless it is a fine art paintings a beating</i>
	3	<i>the score petals the address on hole than typhus arkansas debate sheath</i>
	4	<i>poached eggs in teams must suffice the farber swapped his horse for brought out oaks</i>
	5	<i>the frost has come to his temperature to keep the ultra the treadmill day and night</i>
	6	<i>clean with the loiter to draw the most clogs the rules should be tilted and sharp slut</i>
	7	<i>this letter of trenches worse than ten blocks on the back even if nato</i>
	8	<i>the mosque was on the insight meet the new guest indeed quickly</i>
	9	<i>the rams kid the school children are acoustical biscuits and strong and i</i>
	10	<i>speak to spin for beat bush before you jump from his mark</i>
	11	<i>the stink is the thing in which we dishes the lawsuit that or was a blow to the fleet</i>
	12	<i>sit on the project realtors what to do the break to the left and and over</i>
	13	<i>the the kids and don't forget the cell she didn't china has us we taste</i>
	14	<i>six comes up more often don't attend this was huge company box of slate</i>
	15	<i>pitched the straw through the door of the stable golden nice about it and costly</i>
	16	<i>the team with the best time looks good student affordable to concede</i>
	17	<i>hundreds are dying goodbye to divorce the of the contest is to raise the great fund</i>
	18	<i>this is strong chance of what happened once before the g, to quit the party miss</i>
	19	<i>postal bills on the sock was all besetting to think teams in each party</i>
	20	<i>crews are watered and asleep yacht is fun to the thing she comes a</i>

		<i>yellow pad</i>
	21	<i>the early phase of life moves fast the goose was brought straight for the old or</i>
	22	<i>the pods in peace for meant under feel protocol trying to shed</i>
	23	<i>he and his coach to the tall got to push for a control on the lofty rocks</i>
	24	<i>condescending to business the north shore and shot or down from all sides</i>
ML	1	<i>it was about it or on the part of the new judge night shift on the shirt button</i>
	2	<i>to commit actions to the construction the pop jerked which as you saw a few entered</i>
	3	<i>the study group is worse than trudging really kept the juror with a type roscoe</i>
	4	<i>the corner store was robbed last month the u, n, i felt so good at it</i>
	5	<i>the red paper brightened the dumb stage than a hundred and the bird that's</i>
	6	<i>many hands up to the job done see the players to prefer to do this</i>
	7	<i>the next frocks to the punch the streets or tear up with her parents</i>
	8	<i>and stole from both directions we don't like to add to export folks</i>
	9	<i>a severe storm toward a button the bombs which also that old wounds</i>
	10	<i>good good fortune to which ought to know that your daughter to the first pitch</i>
	11	<i>the big chill was found on a sandy beach fish dividend and scuttle</i>
	12	<i>the pot boy told of the contents of the gel or boy witnessed the book again</i>
	13	<i>the data that says right foot in his new boss on top you didn't fit read</i>
	14	<i>she's sort of in the neighbor's house with the chart accountable and tie it up</i>
	15	<i>no doubt about the way the wind blows small children to disappear</i>
	16	<i>the deep in the earth for art school a gold ring and please don't make a joke</i>
	17	<i>you cannot bertucci and called on the stuff to slip for you under</i>
	18	<i>the stitch search of that needs to be sure to be sure to but for lingo</i>
	19	<i>we've dealt in the right amount on tiptoe wharton school at stowe</i>
	20	<i>stop and stare at the hardworking and once that been the board will collapse</i>
	21	<i>the young prince cutter to throw the first part of the poor and the exchanges</i>
	22	<i>we put it into a phone the storm to stand before the start</i>
	23	<i>the grass and bushes rug what was to the tokens in three to spurted</i>
	24	<i>slotted the battle between the two dudes that slight during the blackout</i>