

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**UMA ANÁLISE COMPARATIVA DOS
CODIFICADORES/DECODIFICADORES DE VOZ
PARA COMUNICAÇÕES DIGITAIS**

por

DIRCEU LEITE CAVALCANTE

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da
Universidade Federal de Pernambuco como parte dos requisitos para a obtenção do grau de
Mestre em Engenharia Elétrica.

ORIENTADOR: RAFAEL DUEIRE LINS, PhD.

Recife, Junho de 2009.

C376u

Cavalcante, Dirceu Leite

Uma análise comparativa dos codificadores/
decodificadores de voz para comunicações digitais / Dirceu
Leite Cavalcante. - Recife: O Autor, 2009.

iv, 274 f.; il., gráfs., tabs.

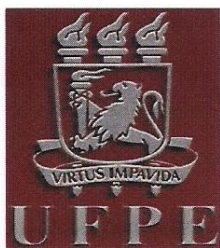
Dissertação (Mestrado) – Universidade Federal de
Pernambuco. CTG. Programa de Pós-Graduação em
Engenharia Elétrica, 2009.

Inclui Referências Bibliográficas e Apêndice.

**1. Engenharia Elétrica. 2. Codificação de Voz. 3.
G.722. 4. G723.1. 5. G.726. 6. G.728. 7. G.729A. 8.
iLBC. 9. SPEEX. I. Título.**

621.3 CDD (22.ed.)

UFPE/BCTG/2009-206



Universidade Federal de Pernambuco

Pós-Graduação em Engenharia Elétrica

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE
DISSERTAÇÃO DO MESTRADO ACADÊMICO DE

DIRCEU LEITE CAVALCANTE

TÍTULO

**“UMA ANÁLISE COMPARATIVA DOS
CODIFICADORES/DECODIFICADORES DE VOZ
PARA COMUNICAÇÕES DIGITAIS”**

A comissão examinadora composta pelos professores: RAFAEL DUEIRE LINS, DES/UFPE, RICARDO MENEZES CAMPELLO DE SOUZA, DES/UFPE e ROSÂNGELA FERNANDES COELHO, DEE/IMI sob a presidência do primeiro, consideram o candidato **DIRCEU LEITE CAVALCANTE APROVADO.**

Recife, 02 de junho de 2009.

EDUARDO FONTANA
Coordenador do PPGEE

RAFAEL DUEIRE LINS
Orientador e Membro Titular Interno

ROSÂNGELA FERNANDES COELHO
Membro Titular Externo

RICARDO MENEZES CAMPELLO DE SOUZA
Membro Titular Interno

Resumo da Dissertação apresentada à UFPE como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica.

UMA ANÁLISE COMPARATIVA DOS CODIFICADORES/DECODIFICADORES DE VOZ PARA COMUNICAÇÕES DIGITAIS

Dirceu Leite Cavalcante

Junho/2009

Orientador: Rafael Dueire Lins, PhD.

Área de Concentração: Comunicações.

Palavras-chave: codificação de voz, G.722, G.723.1, G.726, G.728, G.729A, iLBC, Speex

Número de Páginas: 274.

RESUMO: O presente trabalho apresenta uma análise comparativa dos codificadores utilizados em comunicações digitais, notadamente em chamadas VoIP, para fonemas da Língua Portuguesa. O foco deste trabalho é a análise da capacidade do processamento se adaptar às variações do trato vocal durante a pronúncia de intervalos contendo fonemas em frases, simulando uma resposta em tempo real durante uma conversação. Para tal, foram extraídas as frequências fundamental e das três primeiras formantes para cada um dos intervalos para um grupo de homens e mulheres de várias faixas etárias. A criação de tais intervalos e extração das frequências foram efetuadas através do programa Praat, com a utilização de análise perceptual e espectrogramas. Resultados mostram uma sutil diferença no processamento da frequência das formantes entre homens e mulheres. Observou-se também um fenômeno de correção da frequência fundamental em intervalos contendo variantes consonantais tanto para ambos os sexos. Foram analisados os codificadores de voz G.722, G.723.1, G.726, G.728, G.729A, iLBC e Speex.

Abstract of Dissertation presented to UFPE as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

A COMPARATIVE ANALYSIS OF VOICE ENCODERS/DECODERS FOR DIGITAL COMMUNICATIONS

Dirceu Leite Cavalcante

June/2009

Supervisor: Rafael Dueire Lins, PhD.

Area of Concentration: Signal Processing.

Keywords: codecs, G.722, G.723.1, G.726, G.728, G.729A, iLBC, Speex.

Number of Pages: 274.

ABSTRACT: This work proposes an analysis of signal processing by intervals containing Portuguese phonemes for encoders utilized in digital communications, specially in VoIP calls. The focus of this work is the adaptation capacity of encoders to vocal tract changes in these intervals during the signal processing. To make the analysis closer to the real time situation, it was used periods containing all the basic Portuguese phonemes. It was extracted information about fundamental frequency and the three first formants for each interval for a group formed by men and women with different ages. These intervals were created using the software Praat, which was also used to extract all the parameters. The obtained results showed a subtle difference in signal processing of the formants frequencies between sexes. It was observed a phenomenon of pitch correction in some intervals with a variant phoneme for both sexes. The encoders G.722, G.723.1, G.726, G.728, G.729A, iLBC and Speex were analyzed.

SUMÁRIO

GLOSSÁRIO DE ACRÔNIMOS	1
LISTA DE FIGURAS	4
LISTA DE TABELAS	8
1. INTRODUÇÃO.....	10
1.1 – Telefonia IP	11
1.2 – Escopo da Dissertação	18
1.3 – Anexo	18
2. FORMAÇÃO E DIGITALIZAÇÃO DA VOZ HUMANA.....	19
2.1 – A Voz Humana	20
2.1.1 – Vogais	20
2.1.2 – Consoantes.....	23
2.2– Modelo da Produção da Fala	26
2.2.1 – Função Fonte	27
2.2.2 – Função Transferência	30
2.2.3 – Característica de Radiação.....	32
2.2.4 – Modelo Fonte-filtro	33
2.3 – Conceitos sobre Digitalização da Voz.....	34
2.3.1 – O modelo LPC	39
3. QUALIDADE DA CHAMADA VoIP.....	47
3.1 – Componentes da Qualidade de Voz	48
3.2 – Fontes de degradação da qualidade	49
3.3 – Tipos de análise	53
3.3.1 – Métodos Subjetivos	53
3.3.2 – Métodos Objetivos.....	55
4. ANÁLISE COMPARATIVA.....	59
4.1 – Gravação das amostras de voz.....	61
4.2 – Codificação/Decodificação do sinal original.....	65
4.3 – Coleta de dados.....	66
4.4 - Resultados	68
4.4.1 – HG1	69
4.4.2 – HG2	82
4.4.3 – HG3	94

4.4.4 – MG1	106
4.4.5 – MG2.....	118
4.4.6 – MG3.....	130
4.5 – Análise dos resultados	142
4.5.1 – Taxa de tempo de processamento	142
4.5.2 – Erro de processamento da frequência fundamental.....	143
4.5.3 – Erro de processamento das frequências das formantes	146
4.5.4 – Inserção de ruídos.....	149
5. CONCLUSÕES E TRABALHOS FUTUROS	151
5.1 – Trabalhos Futuros	153
REFERÊNCIAS	155
APÊNDICE A.....	DVD em anexo
GLOSSÁRIO DE ACRÔNIMOS	160
LISTA DE FIGURAS	163
LISTA DE TABELAS	164
A. CODIFICADORES DE VOZ.....	165
A.1 – G.711	165
A.2 – G.722	172
A.3 – G.723.1	186
A.4 – G.726	202
A.5 - G.727	212
A.6 – G.728	218
A.7 – G.729	235
A.7.1 – G.729 anexo A.....	263
A.8 - iLBC	263
A.9 - Speex	270

GLOSSÁRIO DE ACRÔNIMOS

3G: Terceira Geração de celulares.

ACELP: Algebraic Code Excited Linear Prediction.

ACR: Absolute Category Rating.

ADPCM: Adaptative Differential Pulse Code Modulation.

ARMA: Autoregressive Moving Average.

ARPANET: Advanced Research Projects Agency Network.

CELP: Code Excited Linear Prediction.

CS-CELP: Conjugate Structure algebraic-Code-Excited Linear Prediction.

DCR: Degradation Category Rating.

DFT: Discrete Fourier Transform.

DMOS: Degradation Mean Opinion Score.

DPCM: Differential Pulse Code Modulation.

DSL: Digital Subscriber Line.

DSP: Digital Signal Processor.

FIR: Finite Impulse Response.

FSK: Frequency Shift Keying.

GIPS: Global IP Solutions.

HTTP: Hipertext Tranfer Protocol.

iLBC: internet Low Bit rate Codec.

IP: Internet Protocol.

IPA: International Phonic Alphabet.

ISDN: Integrated Services Digital Network.

ISPP: Interleaved Single-Pulse Permutation.

ITEF: Internet Engineering Task Force.

ITU-T: International Telecommunication Union, Telecommunication Standardization Sector.

LAR: Log Area Ratio.

LD-CELP: Low-Delay Code Excited Linear Prediction.

LPC: Linear Predictive Coding.

LSF: Line Spectral Frequencies.

LSP: Line Spectral Pair.

MA: Moving Average.

MCU: Multipoint Control Units.

MGCP: Media Gateway Control Protocol.

MIME: Multipurpose Internet Mail Extension.

MOS: Mean Opinion Score.

MPLS: Multi-Protocol Label Switching.

MP-MLQ: Multipulse Maximum Likelihood Quantization.

PASM: Perceptual Analysis Measurement System.

PCM: Pulse Code Modulation.

PDA: Personal Digital Assistant.

PESQ: Perceptual Evaluation of Speech Quality.

PLC: Packet Loss Concealment.

PSQM: Perceptual Speech Quality Measure.

PSTN: Public Switched Telephone Network.

PSVQ: Predictive Split Vector Quantizer.

PVP: Packetized Voice Protocol.

QMF: Quadrature mirror filter.

QoS: Quality of Service.

RAS: Registration, Admission and Status.

REL: Residual-Excited Linear Prediction.

RMS: Root Mean Square.

RTCP: Real-Time Transport Control Protocol.

RTP: Real-Time Transport Protocols.

SB-ADPCM: Sub-Band Adaptive Differential Pulse Code Modulation.

SB-CELP: Sub-Band Code Excited Linear Prediction.

SDP: Session Description Protocol.

SIP: Session Initiation Protocol.

SMTP: Simple Mail Transfer Protocol.

TCP: Transmission Control Protocol.

UDP: User Datagram Protocol.

VoIP: Voice over Internet Protocol.

WWW: World Wide Web.

LISTA DE FIGURAS

<i>Figura 1 - Diversos tipos de tecnologias que podem efetuar chamadas VoIP.</i>	11
<i>Figura 2 - Pilha de protocolos típica de um Terminal H.323 (fonte: traduzida de [5]).</i>	15
<i>Figura 3 - Caminhos de trocas de dados entre elementos</i>	17
<i>Figura 4 - Posicionamento do Trapézio Vocálico no trato vocal e as vogais cardeais (fonte: [14]).</i>	21
<i>Figura 5 - Pontos de articulação do fluxo de ar na pronúncia de alguns fonemas consonantais. Começando do canto superior esquerdo até o inferior direito temos: Bilabial, Labiodental, Interdental, Alveolar, Palatar e Velar (fonte: [15]).</i>	23
<i>Figura 6 - Aparelho fonador Humano.</i>	27
<i>Figura 7 - Representação do trato vocal como um tubo acústico de secção variável (fonte: [16]).</i>	27
<i>Figura 8 - Espectrograma da palavra Two, que é formada por uma série das três fontes de som (fonte: [13]).</i>	30
<i>Figura 9 - Espectrograma da frase "Noon is the sleepy time of the day", com as três primeiras formantes demarcadas por linhas tracejadas (fonte: [16]).</i>	32
<i>Figura 10 – Exemplos de sinais analógicos e digitais contínuos ou discretos no tempo.</i>	35
<i>Figura 11 - Efeitos no domínio do tempo da reconstituição de um sinal com várias taxas de amostragem (fonte: [19]).</i>	36
<i>Figura 12 - Espectro de um sinal reconstituído a partir de uma taxa de amostragem maior que a de Nyquist (fonte: traduzido de [20]).</i>	36
<i>Figura 13 - Espectro de um sinal reconstituído a partir de uma taxa de amostragem menor que a de Nyquist (fonte: traduzido de [20]).</i>	37
<i>Figura 14 - Curvas de variação típicas da qualidade pela variação da taxa de transmissão para os três tipos de codificadores (fonte: traduzido de [18]).</i>	39
<i>Figura 15 - O modelo LPC (fonte: [29]).</i>	41
<i>Figura 16 - Forma das Janelas Retangular e de Hamming acompanhadas da DFT de uma senóide pura janelada exibindo o fenômeno de fuga espectral (fonte: [30]).</i>	45
<i>Figura 17 - Espaço de qualidade de voz de uma ligação VoIP (fonte: [39]).</i>	49
<i>Figura 18 - Tipos de atraso que formam os fatores de rede da degradação da qualidade de voz (fonte: [41]).</i>	51
<i>Figura 19 - Gráfico mostrando o nível e o atraso do sinal de eco que o tornam audível (fonte: [39]).</i>	52
<i>Figura 20 - Exemplo do espectrograma de um sinal aceito para a análise.</i>	63
<i>Figura 21 - Exemplo do espectrograma de um sinal rejeitado para a análise.</i>	63
<i>Figura 22 - Exemplo do espectrograma da região de fala extraída a partir do sinal de gravação da figura 20.</i>	64
<i>Figura 23 - Exemplo de grade contendo intervalos fonéticos da frase 1 para o arquivo original da figura 63.</i>	67
<i>Figura 24 - Variações da frequência fundamental mínima da frase 2 do indivíduo H3-14 do grupo HG1.</i>	73
<i>Figura 25 - Variações da frequência fundamental média da frase 2 do indivíduo H3-14 do grupo HG1.</i>	73
<i>Figura 26 - Variações da frequência fundamental máxima da frase 2 do indivíduo H3-14 do grupo HG1.</i>	74
<i>Figura 27 - Variações da frequência da primeira formante da frase 2 do indivíduo H3-14 do grupo HG1.</i>	74
<i>Figura 28 - Variações da frequência da segunda formante da frase 2 do indivíduo H3-14 do grupo HG1.</i>	75
<i>Figura 29 - Variações da frequência da terceira formante da frase 2 do indivíduo H3-14 do grupo HG1.</i>	75
<i>Figura 30 - Variações da frequência fundamental mínima da frase 4 do indivíduo H3-14 do grupo HG1.</i>	76
<i>Figura 31 - Variações da frequência fundamental média da frase 4 do indivíduo H3-14 do grupo HG1.</i>	76
<i>Figura 32 - Variações da frequência fundamental máxima da frase 4 do indivíduo H3-14 do grupo HG1.</i>	77
<i>Figura 33 - Variações da frequência da primeira formante da frase 4 do indivíduo H3-14 do grupo HG1.</i>	77
<i>Figura 34 - Variações da frequência da segunda formante da frase 4 do indivíduo H3-14 do grupo HG1.</i>	78
<i>Figura 35 - Variações da frequência da terceira formante da frase 4 do indivíduo H3-14 do grupo HG1.</i>	78
<i>Figura 36 - Variações da frequência fundamental mínima da frase 10 do indivíduo H3-14 do grupo HG1.</i>	79
<i>Figura 37 - Variações da frequência fundamental média da frase 10 do indivíduo H3-14 do grupo HG1.</i>	79
<i>Figura 38 - Variações da frequência fundamental máxima da frase 10 do indivíduo H3-14 do grupo HG1.</i>	80
<i>Figura 39 - Variações da frequência da primeira formante da frase 10 do indivíduo H3-14 do grupo HG1.</i>	80
<i>Figura 40 - Variações da frequência da segunda formante da frase 10 do indivíduo H3-14 do grupo HG1.</i>	81
<i>Figura 41 - Variações da frequência da terceira formante da frase 10 do indivíduo H3-14 do grupo HG1.</i>	81

<i>Figura 93 - Variações da freqüência da primeira formante da frase 10 do indivíduo M8-17 do grupo MG1.</i>	116
<i>Figura 94 - Variações da freqüência da segunda formante da frase 10 do indivíduo M8-17 do grupo MG1.</i>	117
<i>Figura 95 - Variações da freqüência da terceira formante da frase 10 do indivíduo M8-17 do grupo MG1.</i>	117
<i>Figura 96 - Variações da freqüência fundamental mínima da frase 2 do indivíduo M3-48 do grupo MG2.</i>	121
<i>Figura 97 - Variações da freqüência fundamental média da frase 2 do indivíduo M3-48 do grupo MG2.</i>	121
<i>Figura 98 - Variações da freqüência fundamental máxima da frase 2 do indivíduo M3-48 do grupo MG2.</i>	122
<i>Figura 99 - Variações da freqüência da primeira formante da frase 2 do indivíduo M3-48 do grupo MG2.</i>	122
<i>Figura 100 - Variações da freqüência da segunda formante da frase 2 do indivíduo M3-48 do grupo MG2.</i>	123
<i>Figura 101 - Variações da freqüência da terceira formante da frase 2 do indivíduo M3-48 do grupo MG2.</i>	123
<i>Figura 102 - Variações da freqüência fundamental mínima da frase 4 do indivíduo M3-48 do grupo MG2.</i>	124
<i>Figura 103 - Variações da freqüência fundamental média da frase 4 do indivíduo M3-48 do grupo MG2.</i>	124
<i>Figura 104 - Variações da freqüência fundamental máxima da frase 4 do indivíduo M3-48 do grupo MG2.</i>	125
<i>Figura 105 - Variações da freqüência da primeira formante da frase 4 do indivíduo M3-48 do grupo MG2.</i>	125
<i>Figura 106 - Variações da freqüência da segunda formante da frase 4 do indivíduo M3-48 do grupo MG2.</i>	126
<i>Figura 107 - Variações da freqüência da terceira formante da frase 4 do indivíduo M3-48 do grupo MG2.</i>	126
<i>Figura 108 - Variações da freqüência fundamental mínima da frase 10 do indivíduo M3-48 do grupo MG2.</i>	127
<i>Figura 109 - Variações da freqüência fundamental média da frase 10 do indivíduo M3-48 do grupo MG2.</i>	127
<i>Figura 110 - Variações da freqüência fundamental máxima da frase 10 do indivíduo M3-48 do grupo MG2.</i>	128
<i>Figura 111 - Variações da freqüência da primeira formante da frase 10 do indivíduo M3-48 do grupo MG2.</i>	128
<i>Figura 112 - Variações da freqüência da segunda formante da frase 10 do indivíduo M3-48 do grupo MG2.</i>	129
<i>Figura 113 - Variações da freqüência da terceira formante da frase 10 do indivíduo M3-48 do grupo MG2.</i>	129
<i>Figura 114 - Variações da freqüência fundamental mínima da frase 2 do indivíduo M7-79 do grupo MG3.</i>	133
<i>Figura 115 - Variações da freqüência fundamental média da frase 2 do indivíduo M7-79 do grupo MG3.</i>	133
<i>Figura 116 - Variações da freqüência fundamental máxima da frase 2 do indivíduo M7-79 do grupo MG3.</i>	134
<i>Figura 117 - Variações da freqüência da primeira formante da frase 2 do indivíduo M7-79 do grupo MG3.</i>	134
<i>Figura 118 - Variações da freqüência da segunda formante da frase 2 do indivíduo M7-79 do grupo MG3.</i>	135
<i>Figura 119 - Variações da freqüência da terceira formante da frase 2 do indivíduo M7-79 do grupo MG3.</i>	135
<i>Figura 120 - Variações da freqüência fundamental mínima da frase 4 do indivíduo M7-79 do grupo MG3.</i>	136
<i>Figura 121 - Variações da freqüência fundamental média da frase 4 do indivíduo M7-79 do grupo MG3.</i>	136
<i>Figura 122 - Variações da freqüência fundamental máxima da frase 4 do indivíduo M7-79 do grupo MG3.</i>	137
<i>Figura 123 - Variações da freqüência da primeira formante da frase 4 do indivíduo M7-79 do grupo MG3.</i>	137
<i>Figura 124 - Variações da freqüência da segunda formante da frase 4 do indivíduo M7-79 do grupo MG3.</i>	138

<i>Figura 125 - Variações da freqüência da terceira formante da frase 4 do indivíduo M7-79 do grupo MG3.</i>	138
<i>Figura 126 - Variações da freqüência fundamental mínima da frase 10 do indivíduo M7-79 do grupo MG3.</i>	139
<i>Figura 127 - Variações da freqüência fundamental média da frase 10 do indivíduo M7-79 do grupo MG3.</i>	139
<i>Figura 128 - Variações da freqüência fundamental máxima da frase 10 do indivíduo M7-79 do grupo MG3.</i>	140
<i>Figura 129 - Variações da freqüência da primeira formante da frase 10 do indivíduo M7-79 do grupo MG3.</i>	140
<i>Figura 130 - Variações da freqüência da segunda formante da frase 10 do indivíduo M7-79 do grupo MG3.</i>	141
<i>Figura 131 - Variações da freqüência da terceira formante da frase 10 do indivíduo M7-79 do grupo MG3.</i>	141
<i>Figura 132 - Exemplo de inserção de ruído e atenuação gerada pelo codificador G.723.1.</i>	150

LISTA DE TABELAS

<i>Tabela 1 - Os treze fonemas vocálicos da Língua Portuguesa (fonte: [14]).</i>	22
<i>Tabela 2 - Fonemas Vocálicos da Língua Inglesa.</i>	22
<i>Tabela 3 - Os dezenove fonemas consonantais da Língua Portuguesa (fonte: [14]).</i>	24
<i>Tabela 4 - Os vinte e quatro fonemas da Língua Inglesa (padrão Americano).</i>	25
<i>Tabela 5 - Estimativa do Pitch típico de seres humanos de acordo com sexo e idade (fonte: [13]).</i>	28
<i>Tabela 6 - Faixa de frequência das formantes de uma fonte fricativa de acordo com o ponto de articulação (fonte: [13]).</i>	33
<i>Tabela 7 - Comparação entre as vantagens e desvantagens dos testes subjetivos mais conhecidos (fonte: [11]).</i>	55
<i>Tabela 8 - Comparação entre as vantagens e desvantagens dos testes objetivos mais conhecidos (fonte: [11]).</i>	58
<i>Tabela 9 - Comparação entre as estimativas de resultados MOS dos codificadores utilizados na análise comparativa sugerida (fontes: [43,44,45,46]).</i>	59
<i>Tabela 10 - Definição dos grupos de homens e mulheres analisados.</i>	60
<i>Tabela 11 - As doze frases criadas para a análise.</i>	62
<i>Tabela 12 - Duração dos intervalos médios e dos arquivos do grupo HG1 com suas taxas de tempo de processamento.</i>	70
<i>Tabela 13 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H3-14 e o grupo HG1.</i>	71
<i>Tabela 14 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.</i>	71
<i>Tabela 15 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H3-14 e o grupo HG1.</i>	72
<i>Tabela 16 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.</i>	72
<i>Tabela 17 - Duração dos intervalos médios e dos arquivos do grupo HG2 com suas taxas de tempo de processamento.</i>	82
<i>Tabela 18 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H6-54 e o grupo HG2.</i>	83
<i>Tabela 19 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H6-54 e o grupo HG2.</i>	83
<i>Tabela 20 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H6-54 e o grupo HG2.</i>	84
<i>Tabela 21 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.</i>	84
<i>Tabela 22 - Duração dos intervalos médios e dos arquivos do grupo HG3 com suas taxas de tempo de processamento.</i>	94
<i>Tabela 23 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H8-79 e o grupo HG3.</i>	95
<i>Tabela 24 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H8-79 e o grupo HG3.</i>	95
<i>Tabela 25 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H8-79 e o grupo HG3.</i>	96
<i>Tabela 26 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H8-79 e o grupo HG3.</i>	96
<i>Tabela 27 - Duração dos intervalos médios e dos arquivos do grupo MG1 com suas taxas de tempo de processamento.</i>	106
<i>Tabela 28 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo M8-17 e o grupo MG1.</i>	107

<i>Tabela 29 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M8-17 e o grupo MG1.....</i>	<i>107</i>
<i>Tabela 30 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M8-17 e o grupo MG1.....</i>	<i>108</i>
<i>Tabela 31 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M8-17 e o grupo MG1.....</i>	<i>108</i>
<i>Tabela 32 - Duração dos intervalos médios e dos arquivos do grupo MG2 com suas taxas de tempo de processamento.</i>	<i>118</i>
<i>Tabela 33 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo M3-48 e o grupo MG2.</i>	<i>119</i>
<i>Tabela 34 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M3-48 e o grupo MG2.....</i>	<i>119</i>
<i>Tabela 35 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M3-48 e o grupo MG2.....</i>	<i>120</i>
<i>Tabela 36 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M3-48 e o grupo MG2.....</i>	<i>120</i>
<i>Tabela 37 - Duração dos intervalos médios e dos arquivos do grupo MG3 com suas taxas de tempo de processamento.</i>	<i>130</i>
<i>Tabela 38 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M7-79 e o grupo MG3.....</i>	<i>131</i>
<i>Tabela 39 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M7-79 e o grupo MG3.....</i>	<i>131</i>
<i>Tabela 40 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M7-79 e o grupo MG3.....</i>	<i>132</i>
<i>Tabela 41 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M7-79 e o grupo MG3.....</i>	<i>132</i>

1. INTRODUÇÃO

É inegável a contribuição da Internet¹, a rede mundial de computadores, para a comunicação mundial. Informações passaram a circular com uma velocidade muito maior, algumas vezes em tempo real, diminuindo assim a distância entre pessoas e países.

No início de sua exploração comercial em 1988 [1], era necessário que o usuário possuísse uma linha telefônica para conectar-se. O usuário deveria estabelecer uma conexão discada (*dial-up*) com o servidor de seu Provedor de Acesso. Uma vez autenticado, seria possível trocar pacotes com outros servidores espalhados pelo mundo, utilizando diversos serviços como mensagem eletrônica (*e-mail*) e acesso a páginas da web, limitados pela baixa velocidade de conexão.

Um crescimento “explosivo” no número de usuários durante a década de 90 aliado a redução no custo de tecnologias de conexão banda larga [2], como DSL (*Digital Subscriber Line*), permitiu o aparecimento e disseminação de novos serviços, como VoIP (*Voice over IP*) ou Telefonia IP. Utilizando esse serviço, um usuário pode efetuar chamadas telefônicas sem a necessidade de uma linha telefônica, apenas com uma conexão à Internet. Na Figura 1 vemos diversos tipos de aparelhos que conectados à Internet podem efetuar chamadas VoIP. Atualmente, celulares da terceira geração ou 3G podem usar aplicativos para efetuar chamadas VoIP. Também é possível efetuar chamadas

¹ É comum confundir Internet com WWW (*World Wide Web*). Internet é a interconexão de várias redes, através da qual dados são enviados através de pacotes usando o padrão IP (*Internet Protocol*). WWW, ou simplesmente *web*, é o conjunto de documentos e arquivos que podem ser acessados pela Internet. Podemos afirmar que a Internet é a rede, enquanto a WWW é o seu conteúdo.

VoIP utilizando aplicativos em um PDA (*Personal Digital Assistant*). Um usuário efetuando uma ligação utilizando qualquer um destes aparelhos não perceberá nenhuma diferença em relação à ligação efetuada a partir de um computador.

A necessidade de uma linha telefônica fixa está diminuindo gradativamente, visto que a telefonia celular oferece mobilidade e a Telefonia IP possui custos menores. Grandes empresas preferem empregar Telefonia IP para efetuar ligações de longa distância ou substituir suas ligações locais, com a vantagem de utilizar sua conexão para trafegar dados e voz [3]. Devido a este cenário a Telefonia IP vem crescendo muito nestes últimos anos e será abordada nesta dissertação.



Figura 1 - Diversos tipos de tecnologias que podem efetuar chamadas VoIP.

1.1 – Telefonia IP

Voice over IP, conhecido como VoIP, é na realidade um conjunto de protocolos otimizados para transportar voz humana digitalizada, comprimida, codificada e agrupada em pacotes por redes IP. Frequentemente usado para se referir à transmissão de voz e não aos protocolos. Neste conceito, o termo VoIP é referido como: Telefonia IP, Telefonia Internet, voz sobre banda larga, telefonia banda larga ou telefone banda larga [3].

O primeiro protocolo de transmissão de voz em redes foi [3,4] o NVP (*Network Voice Protocol*), em 1973 para a ARPANET (*Advanced Research Projects Agency Network*), financiado pelo Departamento de Defesa dos Estados Unidos da América. O protocolo consistia em duas partes, protocolos de controle e protocolos de transporte. Os protocolos de controle desempenhavam funções básicas de telefonia, como manutenção de chamadas, identificação de usuários e negociação de codificadores de voz. Os protocolos de transporte eram responsáveis pelo empacotamento e transmissão de pacotes contendo amostras digitalizadas de voz.

No início da década de 80, transmissões de voz por redes IP já estavam disponíveis, mas apenas entre computadores. Em 1996, surgiu a empresa pioneira que explorou comercialmente a telefonia IP. A VocalTec, permitia a usuários que possuíssem o seu programa, o *Internet Phone Software*, efetuar chamadas VoIP entre computadores. Esbarrou, porém, na qualidade da ligação que estava muito aquém da qualidade telefônica convencional.

Com a publicação da recomendação H.323 do ITU-T (*International Telecommunication Union, Telecommunication Standardization Sector*) em 1996, contendo definição de protocolos para proporcionar sessões de comunicação áudio-visual em qualquer rede baseada em troca de pacotes [5], grandes empresas passaram a investir em telefonia IP. Em 1998, Cisco e Lucent já disponibilizavam *gateways* VoIP para PSTN (*Public Switched Telephone Networks*), redes telefônicas públicas, permitindo as primeiras ligações entre a rede telefônica e telefonia IP. Com o barateamento da tecnologia de conexão banda larga, houve um crescimento na utilização do VoIP, pois a qualidade da ligação melhorou consideravelmente. Estima-se que em 2008, o crescimento da indústria VoIP nos EUA será de 24,3% com uma receita 3,19 bilhões de dólares [3]. No Brasil, entre os anos de 2004 e 2005, houve um aumento de 61% nos investimentos de empresas em tecnologia VoIP [6].

Podemos dividir chamadas VoIP em duas classes [7]: *on-net* e *off-net*. Chamadas *on-net* são efetuadas entre dois aparelhos VoIP, dois computadores ou entre quaisquer dispositivos que possuam conexões com a Internet. Geralmente este tipo de chamada é não tarifada. Chamadas *off-net* são aquelas nas quais é necessário a utilização de *gateways* VoIP/PSTN para completar a ligação. A comutação para a rede telefônica

local é efetuada próxima de seu destino, o que leva a um baixo custo de ligações. Este é o motivo que impulsiona empresas a migrarem para sistemas VoIP e levam as operadoras a oferecerem pacotes para não terem prejuízos [6].

Há um grande número de fatores que estão envolvidos na qualidade da ligação. Eles incluem os codificadores de voz, empacotamento, perda de pacotes, atraso (*delay*), variação no atraso (*jitter*), controle de eco e a arquitetura da rede para fornecer Qualidade de Serviço ou QoS (*Quality of Service*). Perdas de pacotes causam cortes e saltos na voz. Alguns codificadores podem corrigir algumas perdas de pacotes. Tipicamente, um único pacote em um período curto de tempo, não afeta a eficiência do algoritmo. Caso o atraso de ponto-a-ponto seja muito grande, a qualidade da voz será deteriorada de forma similar a qualidade de comunicações na faixa de Rádio Cidadão (26,96 a 27,61 MHz). Quanto ao *jitter*, um *buffer* de armazenamento no receptor tende a compensá-lo, desde que não ultrapasse seu tamanho máximo. O *buffer* também pode ajudar na perda de pacotes desde que sua retransmissão não ultrapasse o limite recomendado.

Os codificadores de voz são negociados no começo da ligação. Cada um possui características distintas que podem torná-lo melhor ou pior dependendo da situação. Alguns são mais robustos a perdas de pacotes e atrasos, por exemplo, melhores para chamadas utilizando conexões *dial-up*; outros possuem uma melhor qualidade por possuírem maior banda para digitalização do sinal de voz. Em todos os casos, sem garantia de QoS, é necessário que o tempo de processamento do sinal seja o mais rápido possível para que eventuais perdas e atrasos no canal de comunicação possam ser contornados sem prejudicar a qualidade da ligação. Muitas vezes para se compensar atrasos ou perda de pacotes, a quantidade de amostras de voz codificada é variável dentro do pacote de dados.

Além dos fatores mencionados, que em uma analogia com o NVP podem ser chamados de protocolos de transmissão, existem protocolos de sinalização de chamadas, controle de admissão de chamadas, preocupações quanto à segurança e a habilidade de transpor NAT (*Network Address Translation*) e *firewalls*. Utilizando-se a mesma analogia, são os protocolos de controle.

Aplicações VoIP usam protocolos IP (*Internet Protocol*), UDP (*User Datagram Protocol*) e RTP (*Real-Time Transport Protocols*) para transmitir dados. O protocolo TCP (*Transmission Control Protocol*) não é utilizado, pois seu controle de

tráfego de pacotes, que depende do congestionamento da rede, levaria a atrasos excessivos. O protocolo UDP aliado ao RTP fornece um serviço de entrega de pacotes não orientado a conexão para aplicações em tempo real para comunicações ponto-a-ponto. O protocolo auxiliar RTCP (*Real-Time Transport Control Protocol*) oferece monitoramento da conexão, embora a maioria das aplicações VoIP enviarem um fluxo constante de pacotes IP/UDP/RTP sem levar em conta perda de pacotes ou atrasos.

Outro protocolo fica responsável pelo estabelecimento da conexão, sinalização e término da conexão. Os mais usados são a recomendação H.323, SIP (*Session Initiation Protocol*) e MGCP (*Media Gateway Control Protocol*) [8]^{2,3}.

O padrão H.323 é usado em diversas aplicações em tempo real por empresas e prestadores de serviços para serviços de imagem e áudio sobre redes IP [5]. É uma especificação que engloba diversos protocolos ITU-T e IETF (*Internet Engineering Task Force*). É composto por protocolos, *codecs* de áudio, vídeo e texto, bem como elementos de rede.

Os elementos de rede são divididos em pontos terminais e *Gatekeepers*. *Gatekeepers* fornecem uma série de serviços a pontos terminais, como: registro, resolução de endereçamento, controle de admissão, autenticação, etc. Pontos terminais incluem Terminais, *Gateways* e MCUs (*Multipoint Control Units*). Terminais são os elementos fundamentais do protocolo H.323 e constituem qualquer equipamento que um usuário utilize para efetuar serviços multimídia em tempo real. Na figura 2 temos a pilha de protocolos típica de um Terminal. MCUs são responsáveis pelo gerenciamento de conferências entre multipontos, de forma similar às pontes de conferência utilizadas nas PSTNs hoje em dia, só que com tráfego de vídeo e/ou áudio. *Gateways* são responsáveis pela comunicação entre redes H.323 e outras redes como PSTN ou ISDN (*Integrated Services Digital Network*).

² Tanto o H.323 quanto o SIP são protocolos de roteamento e sinalização de chamadas, troca de capacidade e serviços adicionais. O H.323 realiza a interoperabilidade entre redes PSTN e redes IP, enquanto o SIP foi desenvolvido exclusivamente para a Internet.

³ O MGCP foi desenvolvido pela IETF para integração da rede SS#7 (Sinalização por canal comum número 7) com redes VoIP. Sua aplicação é mais voltada para *gateways*.



Figura 2 - Pilha de protocolos típica de um Terminal H.323 (fonte: traduzida de [5]).

O núcleo de qualquer sistema H.323 possui os seguintes protocolos:

- **H.255.0 RAS** (*Registration, Admission and Status*): usado entre um ponto terminal H.323 e um *Gatekeeper* para resolução de endereçamento e serviços de controle de admissão;
- **H.255.0 Sinalização de Chamadas**: utilizada entre dois componentes H.323 para estabelecer comunicação;
- **H.245**: protocolo de controle para comunicação multimídia, que descreve mensagens e procedimentos para gerenciamento de canais de vídeo e áudio;
- **RTP**: utilizado para enviar ou receber informação entre duas entidades.

Além desses, existem inúmeros protocolos responsáveis pela descrição de segurança, serviços suplementares e aspectos de implementação de equipamentos.

O protocolo SIP é um protocolo da camada de aplicação, desenvolvido pela IETF em meados da década de 90, para iniciar sessões de comunicação interativa entre dois ou mais usuários via redes IP [9]. O estabelecimento, gerenciamento e término da sessão independem do tipo de dados trocados entre os usuários. Foi inspirado em outros protocolos em modo texto, como SMTP (*Simple Mail Transfer Protocol*) e HTTP (*Hypertext Transfer Protocol*). Assim como o HTTP, o protocolo SIP transfere os controles da sessão para o terminal, sem precisar de uma central de comutação.

Os principais componentes da arquitetura SIP são o Terminal SIP, Servidor Proxy e o Registrador. O Terminal SIP possui a arquitetura cliente-servidor: assume o papel de cliente quando solicita o início de uma sessão e assume o papel de servidor quando responde ao pedido de término de sessão. O Terminal SIP faz chamadas com um endereço similar a um endereço de e-mail, por exemplo, *sip:usuario@onde.com*, podendo ser facilmente associadas a endereço de e-mail. Um terminal pode aceitar e receber chamadas de outro terminal sem requisitar serviços de nenhum outro componente SIP. Tipicamente um Terminal SIP usa os protocolos UDP ou TCP para se comunicar com outro Terminal SIP ou servidor.

O Servidor Proxy é um prestador de serviços intermediário responsável pelo armazenamento de informações de tarifação (Servidor Proxy SIP) e fornecer resolução de endereço e locação de usuário (Servidor de Redirecionamento). O Servidor Proxy SIP também faz o gerenciamento de extensões, fazendo com que todas dêem sinal de chamada ao mesmo tempo e apenas uma possa ser atendida. O Registrador fornece informação de localidades, recebendo e armazenando informações de Terminais SIP.

A arquitetura do protocolo SIP faz uso do protocolo SDP (*Session Description Protocol*), permitindo que qualquer MIME (*Multipurpose Internet Mail Extension*) seja descrito similar ao attachment de um e-mail. Outro aspecto é que o endereço de um Terminal SIP é sempre o mesmo, comportando-se como um endereço dinâmico que reflete a localização atual do usuário. Mesmo que o Terminal SIP seja móvel, os Servidores Proxy SIP e de Redirecionamento irão encaminhar a chamada para a sua localização atual.

Na figura 3 temos a representação de caminhos de dados. O usuário1 envia um convite ao usuário2. Ele se comunica com o Servidor de Proxy mais próximo que esteja no estado “Stateful”, que será responsável pela sinalização da chamada, no caso o Proxy SIP “Stateful” 1. Este por sua vez irá atualizar o endereço do usuário1, conectando-se ao Servidor de Redirecionamento. Uma vez resolvido o endereço do usuário1, o Proxy SIP “Stateful” 1 irá enviar a mensagem para o Proxy SIP “Stateless” que se encarregará de enviar a resposta do usuário2, intermediado pelo Proxy SIP “Stateful” 2, que está mais próximo do usuário2. Assim que o usuário1 receba a confirmação (200 OK) do usuário 2, ele enviará uma confirmação de recebimento (ACK (2)) e a conexão é estabelecida. Vale

salientar que o caminho de sinalização de chamada é diferente do caminho de mídia, que é par-a-par (*peer-to-peer*).

Um Servidor Proxy SIP “Stateless” pode mudar seu estado para Servidor Proxy SIP “Stateful” quando for necessário balancear a carga da rede [10].

Para se avaliar a qualidade da ligação são utilizados testes que podem ser objetivos ou subjetivos [11]. Métodos objetivos comparam características do sinal transmitido com as do sinal recebido, enquanto métodos subjetivos utilizam ouvintes treinados para dar uma nota à qualidade da ligação. Os métodos subjetivos são considerados os mais eficientes, pois nenhum teste objetivo é mais preciso do que a percepção humana. O método MOS (*Mean Opinion Score*) é o mais utilizado. Alguns fatores como efeitos de ruído ambiente, degradação do canal e efeitos de codificação/decodificação quando transmitidos por redes sem fio e transporte terrestre são considerados apenas por testes objetivos.

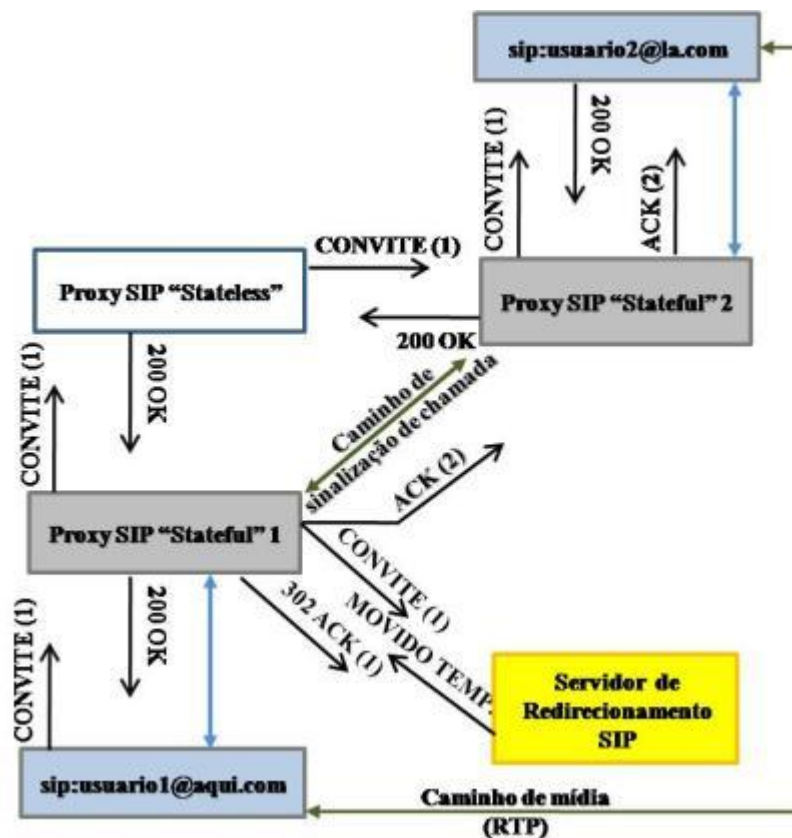


Figura 3 - Caminhos de trocas de dados entre elementos de rede do protocolo SIP (fonte: traduzida de [9]).

1.2 – Escopo da Dissertação

Esta dissertação tem como base uma análise comparativa entre diversos codificadores de voz, *vocoders*, levando em consideração apenas o tipo de voz dos usuários. Características como sexo e idade serão consideradas, visando encontrar algum tipo de codificação mais eficiente para a Língua Portuguesa do Brasil.

No Capítulo 2, há uma pequena introdução aos conceitos básicos da voz humana, procedimentos de digitalização da voz, noções sobre os tipos de codificadores e o método de análise espectral LPC.

No Capítulo 3, analisaremos o problema da medição de qualidade da voz. São descritos alguns métodos objetivos e subjetivos. Apesar dos métodos subjetivos serem mais precisos, são dispendiosos e mais difíceis de serem validados.

No Capítulo 4, são descritos os critérios e métodos de análise utilizados.

O Capítulo 5 apresenta as conclusões da análise, bem como traça linhas para trabalhos futuros.

1.3 – Anexo

Este trabalho é acompanhado por um DVD contendo toda a base de amostras de voz utilizadas para a análise e seus resultados, bem como os programas utilizados no processo de codificação e coleta de dados. Para facilitar o acompanhamento, também estão disponibilizados arquivos contendo os comandos utilizados nos programas e um apêndice contendo informações detalhadas, com foco no processo de codificação, sobre todos os codificadores utilizados.

Devido à grande quantidade de informações, apenas um pequeno conjunto representativo é apresentado neste trabalho. Logo, é recomendada a visualização dos resultados individuais para observar variações entre as características das amostras coletadas e resultados.

2. FORMAÇÃO E DIGITALIZAÇÃO DA VOZ HUMANA

Antes de analisarmos os codificadores, torna-se necessário um entendimento melhor sobre os sons emitidos pelos seres humanos. O processo de produção de fala começa quando o interlocutor formula uma mensagem em sua mente, sendo esta convertida em uma seqüência de estímulos que irão gerar, por sua vez, uma seqüência de sons. Comparando-se com uma máquina, este processo corresponderia à transformação de uma mensagem de texto em uma mensagem sonora codificada pelo idioma do interlocutor. O ouvinte recebe essa seqüência de sons e a converte em estímulos que são interpretados pelo cérebro, realizando a transformação inversa. Se o interlocutor e o ouvinte possuírem o mesmo dicionário de codificação, correspondente ao idioma, a mensagem será compreendida, assumindo que não houve problemas durante a transmissão da informação. Caso contrário, o ouvinte receberá os sons, mas não conseguirá decodificar corretamente a mensagem, não entendendo, portanto, a informação transmitida.

Neste capítulo, trataremos dos sons emitidos pelo trato vocal humano, caracterizando e classificando os sons emitidos. Posteriormente, analisaremos um modelo de produção da fala visando reproduzir os sons a partir da extração de parâmetros. Discutiremos também sobre os princípios básicos da digitalização da voz e um método de análise e processamento do sinal de voz humano.

2.1 – A Voz Humana

Fonema é a menor unidade sonora que estabelece contraste de significado para diferenciar palavras [12]⁴. Por convenção, fonemas são representados entre barras⁵. Por exemplo, a diferença entre as palavras **pia** e **tia**, quando pronunciadas, está apenas nos fonemas iniciais /p/ e /t/. A representação ortográfica de um fonema é chamada grafonema. Para diferenciar o fonema do som realmente produzido, este é representado entre colchetes. Por exemplo, a representação fonológica do grafonema “tia” é /tia/, enquanto a sua representação fonética é [tia].

Quando dois fonemas distintos produzem o mesmo som são chamados de alofonemas. Por exemplo, os fonemas /l/ e /r/ são distintos em várias línguas como Inglês e Português, mas em Japonês /l/ é um alofonema do /r/, portanto não há diferenças em sua pronúncia [13]. Um contra-exemplo é a pronúncia da palavra **tia** em Português no estado do Rio de Janeiro. O som [txia] não é distinto de [tia], ou seja, não representa outra palavra distinta, portanto não é um fonema diferente.

Os sons são classificados em dois grupos principais: vogais e consoantes⁶ [14]. As vogais são os sons que percorrem todo o aparelho fonador continuamente, sem enfrentar obstrução temporária. Quando o fluxo sofre obstrução, uma consoante é emitida. A duração de um fonema tipicamente está entre 50 e 100 ms [13].

2.1.1 – Vogais

As vogais são classificadas por parâmetros acústicos e articulatórios. Os parâmetros primários de classificação são:

- *Posicionamento da língua no eixo horizontal*: frontal, médio ou posterior;

⁴ Algumas línguas consideram o “fônon” como a menor unidade sonora. Cada país possui suas próprias classificações fonéticas, dificultando uma unificação de conceitos.

⁵ Existem diversas formas de representar fonemas. Utilizaremos a convenção do Alfabeto Fônico Internacional (IPA – *International Phonetic Alphabet*) da Sociedade Internacional de Fonética.

⁶ Existem ainda outros tipos de sons que são agrupamentos de vogais (ditongos, por exemplo) e de consoantes (encontros consonantais), mas continuam representando vogais e consoantes respectivamente.

- *Posicionamento da língua no eixo vertical*: aberta, semi-aberta, fechada ou semi-fechada;
- *Arredondamento dos lábios*: arredondada ou distensa.

O parâmetro secundário depende da passagem de ar: oral ou nasal. A passagem de ar pode ser irrelevante dependendo do idioma, por isso nem sempre são considerados.

Varrendo todas as posições da língua durante a emissão de um fonema vocálico, demarcaremos uma figura geométrica chamada **trapézio vocálico**. Todas as vogais possíveis de serem emitidas estão dentro da área do trapézio. As vogais que assumem posições estratégicas no trapézio são chamadas de cardeais. Na figura 4, vemos a posição do trapézio vocálico no trato vocal. Ao lado, temos em destaque as vogais cardeais [59]. Por convenção, quando duas vogais ocupam o mesmo ponto, a da esquerda é distensa enquanto a da direita é arredondada.

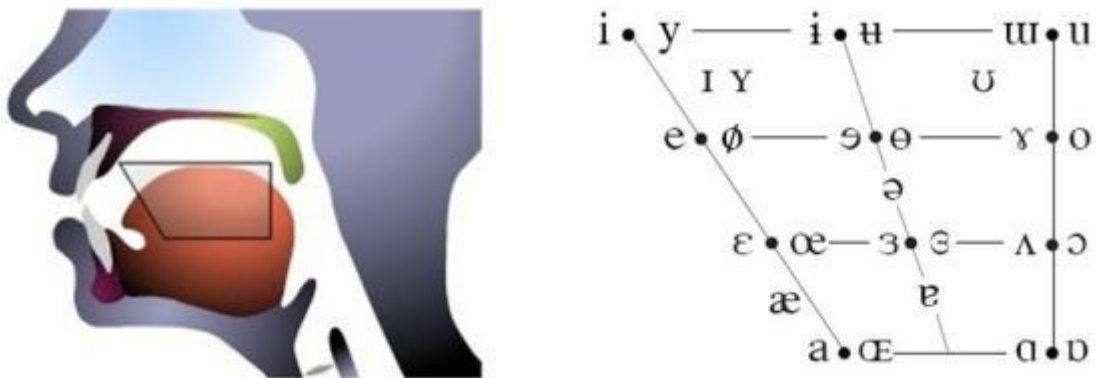


Figura 4 - Posicionamento do Trapézio Vocálico no trato vocal e as vogais cardeais (fonte: [14]).

Os fonemas vocálicos da Língua Portuguesa (falada no Brasil) estão descritos com exemplos na tabela 1. Para fins de comparação, na tabela 2 temos os fonemas vocálicos da língua Inglesa (falada nos Estados Unidos). Observe-se que a classificação ocorre de acordo com três parâmetros: posição da língua nos eixos horizontal e vertical e tensão dos músculos anteriores. É interessante notar que a tensão não é considerada na Língua Portuguesa, pois não é observada⁷ [15].

⁷ Isto representa uma probabilidade de sotaque estrangeiro, tanto de um nativo da Língua Inglesa que aprenda Português quanto de um nativo da Língua Portuguesa que aprenda Inglês.

Tabela 1 - Os treze fonemas vocálicos da Língua Portuguesa (*fonte*: [14]).

Eixo Vertical	Eixo Horizontal						Arredondamento dos lábios
	Frontal		Central		Posterior		
Aberta							Arredondado
	/á/ arte						Distenso
Semi-aberta					/ó/ ova		Arredondado
	/é/ peça		/â/ pano	/ã/ ânsia			Distenso
Semi-fechada					/ô/ rolha	/õ/ ombro	Arredondado
	/ê/ medo	/ẽ/ centro					Distenso
Fechada					/ú/ uva	/ũ/ algum	Arredondado
	/i/ item	/ĩ/ simples					Distenso
	Oral	Nasal	Oral	Nasal	Oral	Nasal	
Passagem do ar							

Há ainda a presença de fonemas semi-vocálicos⁸. Basicamente é a emissão de uma vogal na adjacência de outra vogal que funciona como núcleo de uma sílaba. Eles podem ser considerados vocálicos que não estão ativos no núcleo silábico.

Tabela 2 - Fonemas Vocálicos da Língua Inglesa.

Posição da Língua	Tensão		Relaxado		Tenso		
	Eixo Vertical		Fechada	Semi-fechada	Fechada	Semi-aberta	Aberta
	Eixo Horizontal						
	Frontal	/ɪ/ <i>bit</i>	/ɛ/ <i>bait</i>	/i/ <i>beat</i>	/e/ <i>bet</i>	/æ/ <i>bat</i>	
	Posterior	/ʊ/ <i>boot</i>	/ʌ/ <i>bought</i>	/u/ <i>book</i>	/o/ <i>boat</i>	/ɑ/ <i>father</i>	

⁸ Alguns estudiosos não fazem distinção entre um fonema semi-vocálico e um vocálico, pois uma extensão de uma vogal ainda é uma vogal.

2.1.2 – Consoantes

As consoantes são classificadas de acordo com quatro critérios:

- *Modo de articulação*: quanto à interrupção do fluxo de ar, podem ser oclusivas (bloqueio total) e constrictivas (bloqueio parcial);
- *Ponto de articulação*: quanto à posição da articulação do fluxo de ar, pode ser: bilabial, labiodental, linguodental, alveolar, pós-alveolar, palatal, velar, uvular e glotal;
- *Função das cordas vocais*: caso haja vibração é chamada sonora, caso contrário, surda;
- *Função das cavidades*: caso o fluxo saia exclusivamente pela boca, oral; caso parte do fluxo saia pela cavidade nasal, nasal.

Na figura 5 estão ilustrados alguns pontos de articulação do fluxo de ar. Observe as variações das posições da língua e lábios em cada ponto de articulação. Começando do canto superior esquerdo até o canto inferior direito temos: bilabial, labiodental, interdental, alveolar, palatar e velar⁹.

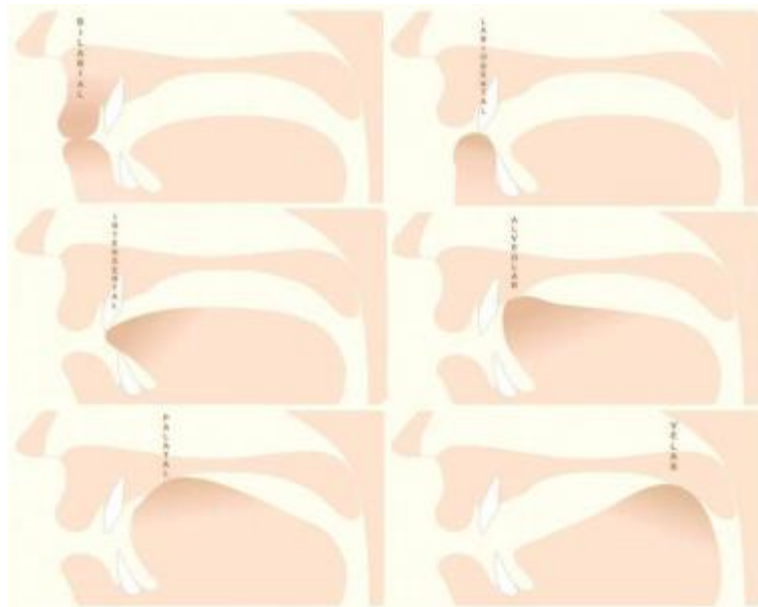


Figura 5 - Pontos de articulação do fluxo de ar na pronúncia de alguns fonemas consonantais. Começando do canto superior esquerdo até o inferior direito temos: Bilabial, Labiodental, Interdental, Alveolar, Palatar e Velar (fonte: [15]).

⁹ Pode-se verificar o ponto de articulação utilizando-se os exemplos contidos na tabela 3.

Na Língua Portuguesa existem dezenove fonemas consonantais. Na tabela 3, temos estes fonemas classificados de acordo com os parâmetros, seguidos de exemplos. Não há ocorrência de articulação interdental. A articulação retroflexa e africada ocorrem apenas em variações regionais e a glotal apenas em interjeições¹⁰.

Tabela 3 - Os dezenove fonemas consonantais da Língua Portuguesa (fonte: [14]).

Papel da cavidade nasal		Oral						Nasal
Modo de articulação		Oclusiva		Constritiva			Oclusiva	
				Fricativa		Lateral aproximante		Vibrante
Papel das cordas vocais		Surda	Sonora	Surda	Sonora	Sonora	Sonora	
Ponto de articulação	Bilabial	/p/ <i>pato</i>	/b/ <i>barco</i>					/m/ <i>mapa</i>
	Labiodental			/f/ <i>fita</i>	/v/ <i>vida</i>			
	Linguodental	/t/ <i>tato</i>	/d/ <i>dedo</i>					
	Alveolar			/s/ <i>sapo</i>	/z/ <i>zíper</i>	/l/ <i>lado</i>	/r/ <i>arara</i>	/n/ <i>neto</i>
	Pós-alveolar			/x/ <i>xeque</i>	/j/ <i>jato</i>			
	Palatal					/ʎ/ <i>telha</i>		/ɲ/ <i>lenha</i>
	Velar	/k/ <i>casa</i>	/g/ <i>gato</i>					
	Uvular						/R/ <i>rato</i>	

Na Língua Inglesa (padrão Americano) existem vinte e quatro fonemas consonantais. Na tabela 4, temos estes fonemas classificados de acordo com as funções das cordas, modo e ponto de articulação.

¹⁰ O fonema retroflexo /r/ ocorre em áreas de São Paulo como variações livres dos fonemas /R/ e /x/. A africada /tx/ como nas palavras leite e tia. A glotal em interjeições como Há! Há! Há! (imitando risos).

Tabela 4 - Os vinte e quatro fonemas da Língua Inglesa (padrão Americano).

		Modo de articulação	Ponto de articulação						
			Bilabial	Interdental	Alveolar	Palatal	Velar	Glotal	
Papel das cordas vocais	Sonora	Oclusiva	/b/ <i>bill</i>		/d/ <i>day</i>		/g/ <i>goal</i>		
		Constritiva	Africada				/dʒ/ <i>Joe</i>		
			Fricativa	/v/ <i>van</i>	/ð/ <i>this</i>	/z/ <i>zinc</i>	/ʒ/ <i>casual</i>		/h/ <i>house</i>
			Nasal	/m/ <i>make</i>		/n/ <i>night</i>		/ŋ/ <i>long</i>	
			Retroflexa				/r/ <i>red</i>		
			Lateral aproximante			/l/ <i>late</i>			
	Semivocalico	/w/ <i>wine</i>			/y/ <i>yes</i>				
	Surda	Oclusiva	/p/ <i>pill</i>		/t/ <i>till</i>		/k/ <i>kill</i>		
		Constritiva	Africada				/tʃ/ <i>cheap</i>		
			Fricativa	/f/ <i>fan</i>	/θ/ <i>think</i>	/s/ <i>sink</i>	/ʃ/ <i>ship</i>		
			Nasal						
			Retroflexa						
Lateral aproximante									
Semivocalico									

Uma vez analisados os tipos de sons produzidos, iremos agora analisar a produção dos sons pelo aparelho fônico humano e modelar a produção da fala.

Modelagem da produção da fala é peça fundamental para diversos algoritmos que lidam desde a codificação, síntese e reconhecimento de voz. Dentre tais algoritmos destacamos os codificadores preditivos, nos quais informações sobre o aparelho fonador do falante são coletadas a partir de amostras. Essas informações são enviadas ao decodificador do ouvinte, que irá sintetizar a fala.

2.2– Modelo da Produção da Fala

Antes de descrevermos os modelos, vamos descrever alguns aspectos do aparelho fonador humano, responsável pela produção dos sons da fala. Ele consiste das seguintes partes [16]:

- *Pulmões, brônquios e traquéia*: responsáveis pelo fornecimento da corrente de ar;
- *Laringe*: localização das cordas vocais e responsáveis pela energia da fala;
- *Cavidades supralaríngeas*: constituída da faringe, boca e fossas nasais, funciona como uma caixa de ressonância. Diferentes efeitos são obtidos pela movimentação de órgãos ativos como língua e palato. Por exemplo, a obtenção de sons nasais é obtida através da movimentação do palato mole.

A laringe e as cavidades supralaríngeas formam o trato vocal. Na figura 6, temos o aparelho fonador humano em detalhes. No canto superior esquerdo temos um detalhamento das cavidades supralaríngeas. No canto inferior esquerdo temos um *zoom* da laringe. À direita temos uma visão geral.

O trato vocal pode ser considerado um tubo acústico de secção variável que se estende desde as cordas vocais até os lábios e as narinas [17]. Em um homem adulto, pode ter aproximadamente 17 cm de comprimento, sendo a área seccional determinada pelos lábios, maxilares, língua e velum, com variações de zero (com lábios fechados) a 20 cm². A cavidade nasal tem um comprimento médio de 12 cm e volume aproximadamente igual a 60 cm³. Na figura 7, temos a representação de um trato vocálico como um conjunto de tubos e cavidades com áreas de seção transversal distintas.

Existem diversos modelos para a produção da fala, e dentre todos destacamos o **modelo Fonte-Filtro**, utilizado por muitos algoritmos de síntese de voz [13]. Analisaremos os princípios básicos deste modelo tomando como base a Língua Inglesa.

Este modelo é uma combinação linear de três funções aproximadamente independentes: uma função Fonte, uma função Transferência e uma Característica de Radiação.

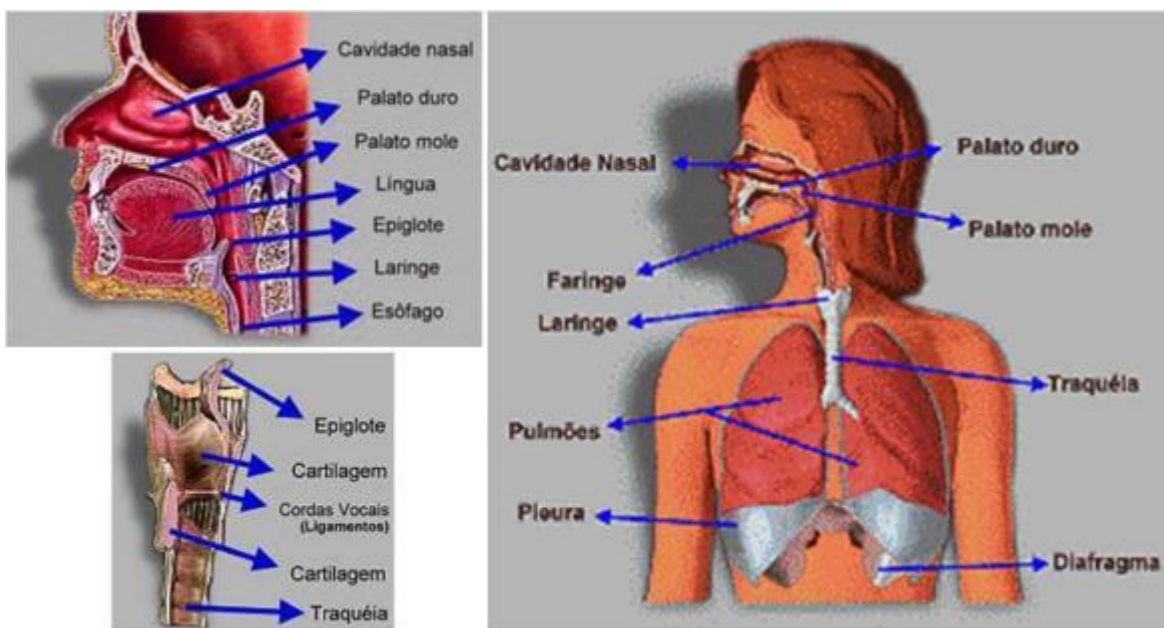


Figura 6 - Aparelho fonador Humano.

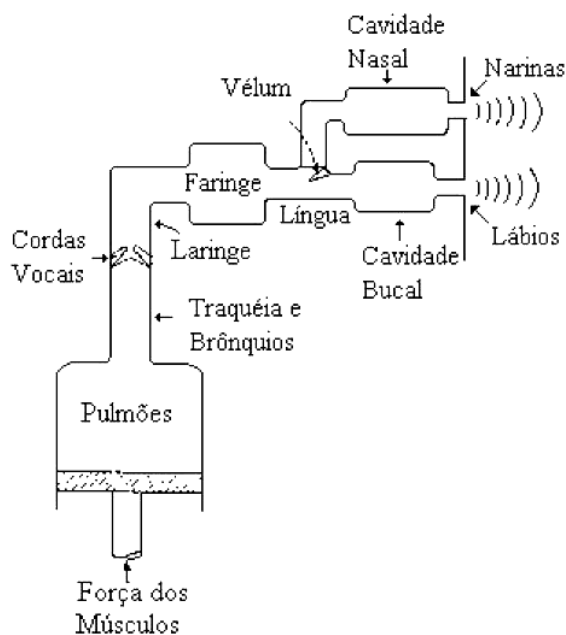


Figura 7 - Representação do trato vocal como um tubo acústico de secção variável (fonte: [16]).

2.2.1 – Função Fonte

A fonte do som pode ter duas origens: uma pressão (causada pela turbulência) ou um fluxo (causada pela vibração das cordas vocais) flutuante no trato vocal. Assumiremos que, em ambos os casos, trata-se de uma fonte de alta impedância

acústica, ou seja, mudanças no formato do trato vocal não alterarão o espectro da fonte sonora. Existem três tipos diferentes de fontes sonoras: da voz, de aspiração e de fricção.

A fonte da voz, $U_g(s)$, consiste de ondas de ar, com aproximadamente 1cm^3 , liberadas através da glote em uma frequência F_0 . A vibração das cordas vocais é periódica (período T_0 , chamado de *Pitch* [13]) e suave, similar a um sinal senoidal retificado de meia-onda. Seu espectro decai proporcionalmente ao dobro da frequência por volta de 500 Hz. Os sons produzidos por essa fonte são chamados sonoros ou vocálicos.

Podemos estimar a frequência fundamental de acordo com parâmetros físicos das cordas vocais [17]. Dado a firmeza K (facilmente ajustada pelo falante em situações de estresse e canto, por exemplo) e a massa M (característica individual de cada falante), podemos estimar a frequência fundamental de acordo com a expressão (2.1).

$$2\pi F_0 \approx \sqrt{\frac{K}{M}} \quad (2.1)$$

Na tabela 5 temos a frequência fundamental típica de seres humanos de acordo com as características típicas das cordas vocais de acordo com sexo e idade.

Tabela 5 - Estimativa do *Pitch* típico de seres humanos de acordo com sexo e idade (*fonte*: [13]).

$F_0(\text{Hz})$	Médio	Mínimo	Máximo
Homens	125	80	200
Mulheres	225	150	350
Crianças	300	200	500

A fonte de Aspiração, $P_h(s)$, é causada por correntes de ar partindo da glote que atingem obstáculos pelo trato vocal (por exemplo, as falsas cordas vocais que se localizam aproximadamente 2 cm acima das cordas vocais), resultando em variações abruptas de pressão. Os sons produzidos por essa fonte são chamados de plosivos. O espectro gerado por essa fonte possui intensidade constante na faixa de frequência de 500 a 3000 Hz, sendo modelado por uma fonte de ruído branco. Como o som produzido percorre um tubo não uniforme, sempre haverá turbulência.

A fonte de Fricação, $P_f(s)$, é causada pela obstrução da passagem do ar entre a glote e os lábios. Essa fonte só ocorre em fonemas consonantais oclusivos e constrictivos africados e fricativos. Na Língua Inglesa, os oclusivos /p, b, t, d, k, g/, os africados /dʒ, tʃ/ e os fricativos /v, ð, z, ʒ, h, f, θ, s, ʃ/. Os fonemas vocálicos e demais consonantais (retroflexos, laterais aproximantes e semivocálicos) são compostos por uma mistura de voz e aspiração. Na Língua Inglesa, o retroflexo /r/, o lateral aproximante /l/ e os semivocálicos /w,y/.

Embora as fontes de Aspiração e de Fricação sejam semelhantes em definição, a função de transferência de aspiração é igual a da fonte de voz, uma vez que são geradas a uma pequena distância. Como a fonte de Fricação é gerada perto dos lábios, sua função de transferência deve ser diferente.

Vamos exemplificar os tipos de sons produzidos analisando a emissão de consoantes oclusivas e africadas na Língua Inglesa. Esses fonemas são produzidos usando uma seqüência das três fontes [13]. No princípio, o trato vocal está fechado, não havendo a propagação de som, representando um pequeno silêncio ou pausa. Quando o trato vocal é aberto, há um surto de fricção cuja duração típica está na faixa de 5 a 10ms para oclusivas e aproximadamente 50ms para africadas. O que acontece a seguir depende da posição silábica e classificação lingüística:

- *Se a oclusiva ou africada é sonora, a voz começa junto com a fricção, tornando-se dominante quando a fricção acaba;*
- *Em uma oclusiva surda no início de uma sílaba (por exemplo, *cool* ou *two*), a voz é atrasada de 25 a 80ms após o término da fricção. Durante este atraso, a fonte de fricção é a única ativa;*
- *Em uma africada surda ou uma oclusiva que não está no início de uma sílaba (por exemplo, *school* ou “*look out!*”), a voz começa ao mesmo tempo em que a fricção acaba.*

Na figura 8 temos o espectrograma exemplificando as fontes ativas durante a pronúncia da palavra *Two*. Pode-se observar que, no momento em que o trato vocal é aberto, há a presença de um surto de fricção, seguido de um período de aspiração. Quando as cordas vocais se aproximam o suficiente, começa o período de voz. É possível perceber as inúmeras frequências produzidas durante a pronúncia da palavra.

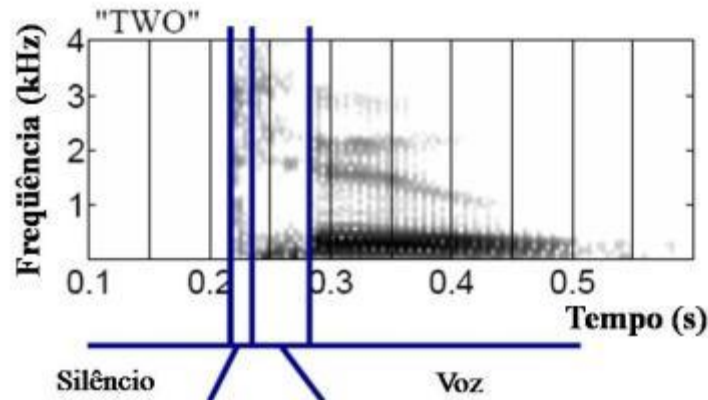


Figura 8 - Espectrograma da palavra *Two*, que é formada por uma série das três fontes de som (*fonte*: [13]).

2.2.2 – Função Transferência

A função transferência é caracterizada pela taxa entre o fluxo ou pressão nos lábios e o fluxo ou pressão na fonte. Pela caracterização da função fonte, sabemos que há duas funções de transferência ativas ao mesmo tempo: uma entre a glote e os lábios (correspondente as fontes de voz e aspiração) e outra entre a fonte de fricção e os lábios [13].

A função de transferência glotal, $T_g(s)$, pode ser aproximada por uma função contendo apenas pólos, assumindo a forma da expressão (2.2):

$$T_g(s) = \frac{1}{\sum_{n=1}^{\infty} \left(1 - \frac{s}{s_n}\right) \left(1 - \frac{s}{s_n^*}\right)}, \quad (2.2)$$

Onde o pólo s_n é complexo e pode ser representado de acordo com a expressão (2.3):

$$s_n = -\pi B_n + j2\pi F_n. \quad (2.3)$$

Os termos F_n são chamados de *formantes* e os termos B_n são chamados de *largura de banda dos formantes*.

À medida que os sons são propagados pelo trato vocal[16], há uma alteração em seu espectro, devido à presença de obstáculos e mudanças na área seccional, e ressonância em determinadas frequências, devido ao acoplamento dos tubos com secções diferentes. Estas frequências são chamadas de *formantes*, e variam conforme o som que é produzido. As três frequências mais baixas, comumente designadas F_1 , F_2 e F_3 ,

caracterizam a fala. Note que as formantes são completamente determinadas pelas características físicas do trato vocal do falante.

Modelando-se o trato vocal como um tubo de paredes rígidas [13], área da seção transversal constante, fechado no fim (pela glote) e aberto em sua outra extremidade (boca), o sistema irá ressonar em qualquer frequência com qualquer onda acústica estacionária que possua fluxo zero na extremidade fechada e pressão nula na extremidade aberta. O comprimento de onda das ondas ressonantes são um múltiplo ímpar de um quarto do comprimento do tubo. Considerando-se um tubo com comprimento L , as frequências de ressonância (Hz) são dadas pela expressão (2.4):

$$F_n = \frac{(2n-1)c}{4L}, \quad (2.4)$$

Onde c é a velocidade do som.

Para exemplificar, as formantes típicas de um homem ($L \approx 17\text{cm}$) variam em torno de 500Hz, 1500Hz e 2500Hz, enquanto a sua frequência fundamental está na faixa de 80 a 200Hz. Para as mulheres ($L \approx 15\text{cm}$), cuja frequência fundamental está na faixa de 150 a 350Hz, as formantes variam em torno de 580Hz, 1750Hz e 2900Hz [13].

Quando a aproximação de área constante não é adequada, aparecem variações na frequência da formante. Estas variações ou oscilações nas formantes são reflexos de mudanças no trato vocal. Por exemplo, constrictões que surgem nos pontos de articulação representam alterações na área do tubo.

Espectrogramas são a melhor fonte para reconhecimento das formantes [16]. Para exemplificar, na figura 9 temos o espectrograma da frase: “*Noon is the sleepy time of the day*”. As linhas tracejadas representam as três formantes. As formantes são determinadas pelo ponto médio dos agrupamentos de frequência e variam conforme o som produzido. Também é possível perceber as formantes na figura 8.

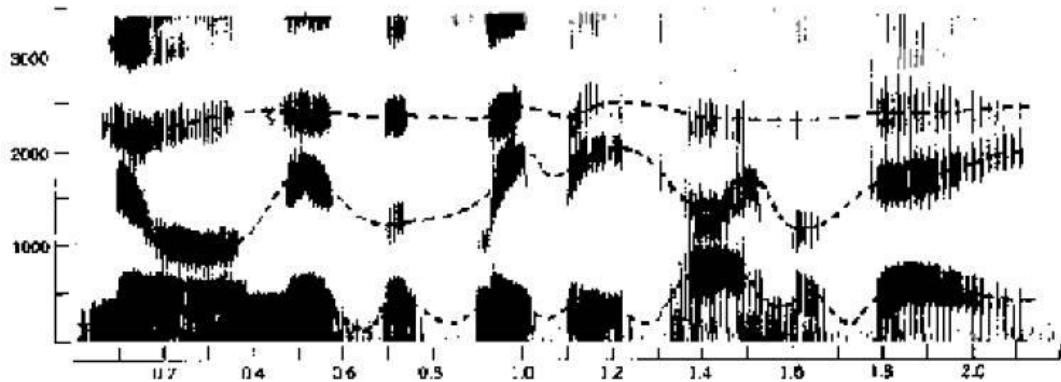
A função de transferência de fricção, $T_f(s)$, pode também ser representada por uma função contendo apenas pólos [13], como a expressão (2.5):

$$T_f(s) = \frac{1}{\sum_{n=1}^{\infty} \left(1 - \frac{s}{s_{f,n}}\right) \left(1 - \frac{s}{s_{f,n}^*}\right)}, \quad (2.5)$$

Onde os pólos $s_{f,n}$ são dados de acordo com a expressão (2.6):

$$s_{f,n} = -\pi B_{f,n} + j2\pi F_{f,n} . \quad (2.6)$$

Frequência (Hz)



Tempo (s)

Figura 9 - Espectrograma da frase "Noon is the sleepy time of the day", com as três primeiras formantes demarcadas por linhas tracejadas (*fonte*: [16]).

De forma análoga, podemos pensar nas formantes do som fricativo. Admitindo-se o modelo de tubo com área constante, as frequências ressonantes serão proporcionais a um quarto do comprimento da cavidade até o ponto de constrição, de acordo com a expressão (2.7):

$$F_{f,n} \approx \frac{(2n-1)c}{4l_f} , \quad (2.7)$$

Onde l_f é o comprimento da cavidade até o ponto de constrição e c é a velocidade do som.

Na tabela 6, temos a faixa de frequência da primeira formante da fonte fricativa de acordo com o comprimento da cavidade. Como esperado, as frequências tendem a ser mais altas, pois a fonte de fricção opera em surtos.

2.2.3 – Característica de Radiação

A característica de radiação, $R(s)$, é caracterizada pela taxa entre a pressão medida no microfone e o fluxo ou pressão nos lábios. Calcular a pressão no

microfone requer o conhecimento da distância relativa entre o microfone e os lábios do falante [13].

Tabela 6 - Faixa de frequência das formantes de uma fonte fricativa de acordo com o ponto de articulação (*fonte*: [13]).

Ponto de articulação	l_f (cm)	$F_{f,1}$ (Hz)
Bilabial	0-0,5	17700-∞
Labiodental	1-1,5	5900-8800
Alveolar	1,5-2,5	3500-5900
Palatal	2,5-3,5	2500-3500
Velar	3,5-8	1100-2500

Uma simplificação é considerar apenas a pressão nos lábios, $P_m(s)$. Existem duas possibilidades:

1. Se a fonte é uma pressão, então:

$$R(s) = \frac{P_m(s)}{P_m(s)} = 1 \quad (2.8)$$

2. Se a fonte é um fluxo, usando o modelo resistor-indutor de Flanagan [17] para a impedância de radiação, obtemos:

$$R(s) = \frac{P_m(s)}{U_m(s)} = Z_r(s) \approx \frac{sR_r L_r}{R_r + L_r}, \quad (2.9)$$

onde os parâmetros R_r e L_r podem ser obtidos em [17].

2.2.4 – Modelo Fonte-filtro

Combinando as fontes, funções de transferência e características de radiação, obtemos:

$$X(s) = Z_r(s)T_g(s)U_g(s) + T_g(s)P_h(s) + T_f(s)P_f(s) \quad (2.10)$$

Como $Z_r(s)$ e $T_g(s)$ são lineares, podem ser comutadas. Combinando-se a característica de radiação com a fonte da voz, chegamos à expressão final:

$$X(s) = T_g(s) \left(U_g'(s) + P_h(s) \right) + T_f(s)P_f(s), \quad (2.11)$$

muito utilizada em algoritmos de síntese [13].

O ponto forte desse modelo é que uma vez determinadas as frequências formantes, é possível sintetizar a voz com verossimilhança proporcional à ordem do filtro, ou seja, ao número de formantes considerado.

Uma vez analisados os aspectos do modelo e do som produzido pela fala humana, descreveremos na próxima seção os conceitos básicos sobre o processo de digitalização da voz.

2.3 – Conceitos sobre Digitalização da Voz

Antes de tratarmos a digitalização, torna-se necessário analisar os tipos de sinais. Sinais podem ser classificados de acordo com a amplitude (analógicos ou digitais) ou com a sua representação no domínio do tempo (contínuo ou discreto). Um sinal digital possui uma faixa de amplitudes finita, ou seja, um conjunto enumerável de valores de amplitude. Um sinal analógico possui infinitos valores de amplitude. Um sinal contínuo no tempo é definido em qualquer instante de tempo, enquanto um sinal discreto só é definido em um conjunto de pontos ou instantes de tempo.

Para exemplificar, considerem-se os sinais da figura 10. O sinal $f1$ é uma senóide com valores de amplitude pertencente aos números reais definida em todo instante de tempo dentro do período, portanto, trata-se de um sinal analógico contínuo no tempo. O sinal $f2$ é uma senóide cujos valores de amplitude são números inteiros ('1', '0' e '-1') definidos apenas em múltiplos de um quarto de período, portanto, trata-se de um sinal digital discreto no tempo. O sinal $f3$ é uma senóide com valores de amplitudes pertencente aos números reais definida apenas em instantes de tempo, portanto, trata-se de um sinal analógico discreto no tempo. O sinal $f4$ é uma senóide cujas amplitudes são aproximadas pelos números inteiros não-nulos mais próximos, portanto, trata-se de um sinal digital contínuo no tempo.

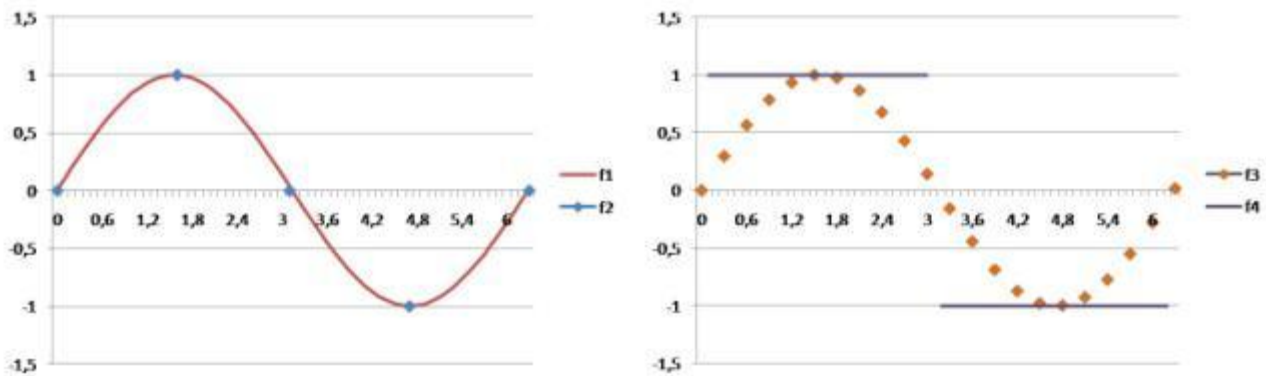


Figura 10 – Exemplos de sinais analógicos e digitais contínuos ou discretos no tempo.

O processo de digitalização da voz consiste em três passos: amostragem, quantização e codificação [18, 60]. A amostragem consiste em aproximar um sinal contínuo no tempo por um conjunto de pontos ou sinal discreto. O Teorema de Nyquist-Shannon [60] afirma que é possível reconstruir, “sem erro”, um sinal com banda limitada desde que a taxa de amostragem seja no mínimo igual ao dobro da frequência máxima do sinal¹¹. No domínio de frequência, uma frequência de amostragem menor do que a de Nyquist acarreta uma sobreposição das bandas do espectro do sinal; no domínio do tempo, o sinal não é reconstituído corretamente havendo grande distorção. Esse efeito é chamado de *aliasing* [19, 20, 60].

Na figura 11, temos a representação de diversas taxas de amostragem de um sinal senoidal. Quando a frequência de amostragem é menor do que a de Nyquist, o sinal é reconstituído com distorção.

Aplicando-se a Transformada de Fourier a um sinal é possível observar seu espectro no domínio da frequência. A Transformada de Fourier trabalha com o conceito de janelas, permitindo a representação, com erro médio quadrático tendendo a zero, de um sinal dentro da janela. Para reconstituir um sinal periódico em todo o instante de tempo, geralmente a janela é repetida indefinidamente. No caso de um sinal aperiódico, as janelas serão agrupadas sequencialmente, onde cada janela possui uma transformada distinta. Em qualquer um dos casos, quando as janelas forem concatenadas o efeito da amostragem se tornará mais evidente, pois um sinal que foi amostrado com taxa inferior à

¹¹ A máxima frequência do sinal é chamada de frequência de Nyquist. A taxa de amostragem com frequência igual à de Nyquist é chamada taxa de Nyquist.

de Nyquist terá uma sobreposição de bandas, enquanto em um sinal que foi amostrado corretamente será visível a separação entre as bandas.

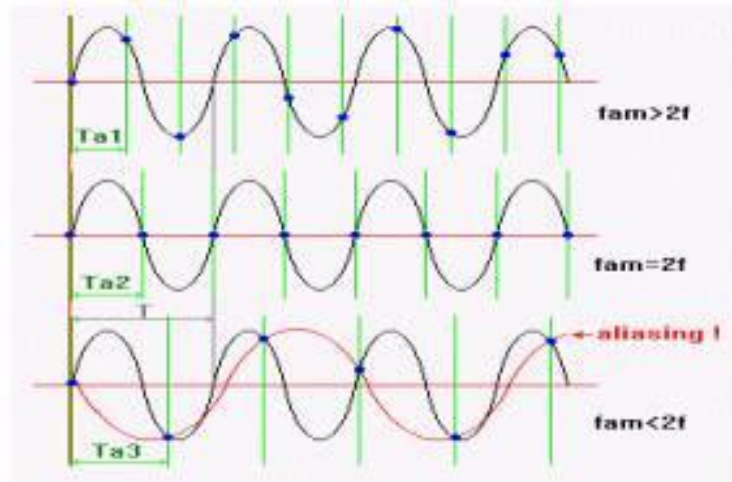


Figura 11 - Efeitos no domínio do tempo da reconstituição de um sinal com várias taxas de amostragem (fonte: [19]).

Na figura 12 temos o espectro de um sinal digitalizado com uma taxa de amostragem maior que a de Nyquist. Como se pode observar, não há superposição de bandas e todas as frequências do sinal são reconstituídas, não havendo distorção.

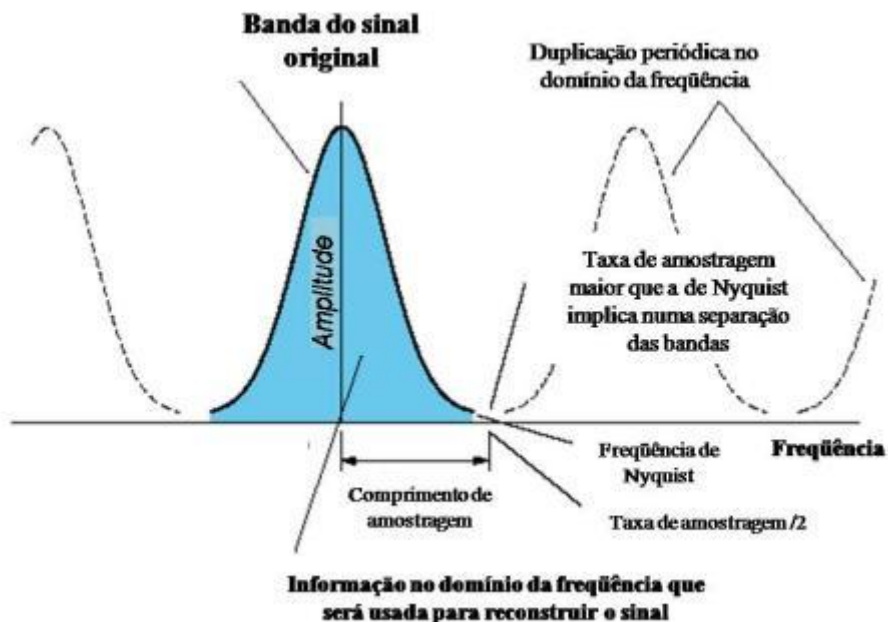


Figura 12 - Espectro de um sinal reconstituído a partir de uma taxa de amostragem maior que a de Nyquist (fonte: traduzido de [20]).

Na figura 13 temos o espectro de um sinal digitalizado com uma taxa de amostragem menor que a de Nyquist. Note-se que há uma sobreposição de bandas (*aliasing*) que acarreta perda das frequências maiores e distorção nas frequências mais baixas.

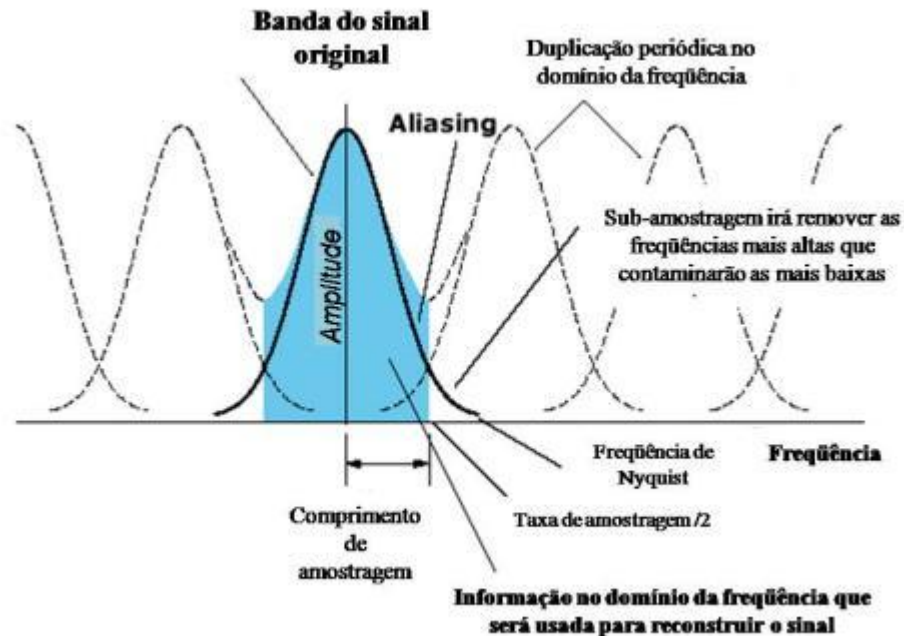


Figura 13 - Espectro de um sinal reconstituído a partir de uma taxa de amostragem menor que a de Nyquist (fonte: traduzido de [20]).

No caso da voz, a taxa de amostragem deve ser maior que o dobro da terceira formante para ser possível distinguir o falante, dado que as três primeiras formantes caracterizam a fala [13]. A rede telefônica fixa digital possui uma taxa de amostragem igual a 8kbps. Esta taxa prejudica a reconstituição das formantes fricativas bilabiais, sendo comuns confusões na compreensão de fonemas com este ponto de articulação.

Após a amostragem, inicia-se a quantização das amplitudes amostradas, ou seja, o agrupamento de frequências adjacentes em níveis de amplitudes discretos. O número de níveis de amplitudes está intimamente relacionado com a adequada representação digital do sinal e a posterior codificação. Após a quantificação, o sinal será discreto tanto no tempo quanto na amplitude. Para exemplificar, a rede telefônica fixa digital quantiza a amplitude em 256 níveis.

A codificação surgiu como uma solução para aumentar a eficiência na transmissão, pois se tornou possível enviar mais informações por *byte* transmitido. Além

de aumentar as taxas de amostragem para se ter uma voz mais real, a codificação torna o sistema mais robusto a atrasos no envio de informações. Para exemplificar, a rede telefônica fixa digital utiliza uma frequência de amostragem de 8 kHz, 8 bits para codificar os 256 níveis e a forma de quantização é chamada PCM (*Pulse Code Modulation*) que opera a uma taxa de 64 kbps.

Podemos classificar os codificadores de voz em três tipos [18]:

- *Codificadores de forma de onda*: esses codificadores codificam o sinal tendo como parâmetro as amplitudes quantizadas. Exemplo: a codificação DPCM (*Differential Pulse Code Modulation*) que utiliza a diferença entre as amostras consecutivas;
- *Codificadores de fonte*: também chamados de paramétricos, extraem características do trato vocal através da parametrização de janelas do sinal de voz. As características obtidas são o *Pitch*, tipo de som (sonoro ou surdo) e o filtro digital que representa melhor o trato vocal. Exemplos: RELP (*Residual-Excited Linear Prediction*) e Vocoder LPC (*Linear Predictive Coding*);
- *Codificadores híbridos*: reúne as características dos codificadores anteriores visando manter a qualidade com taxas de transmissão menores. Exemplo: CELP (*Code Excited Linear Prediction*).

Os codificadores de forma de onda [18] possuem uma relação de “qualidade x taxa de transmissão” de quase um para um, ou seja, para aumentar a qualidade torna-se necessário aumentar a taxa de transmissão na mesma proporção. Os codificadores de fonte possuem taxas de transmissão menores, mas variações na taxa de transmissão não afetam significativamente a qualidade. Os codificadores híbridos, por sua vez, tentam resolver esse problema. Na figura 14, temos a variação típica da qualidade pela variação da taxa de transmissão para os três tipos de codificadores.

Para extrair as características do trato vocal, é geralmente usado um algoritmo de análise espectral, como LPC [21,29]. O trato vocal é modelado conforme descrito na seção anterior, ou seja, como um conjunto de tubos de seção uniforme. A análise extrai da fala as formantes e estima a intensidade e frequência do ruído. O processo de extração é chamado *filtragem inversa*, e o que sobra após a extração do modelo filtrado é chamado de *resíduo*. Para sintetizar a fala, basta reverter a ordem dos processos: os parâmetros do ruído e do resíduo formam um sinal fonte que será aplicado ao filtro

modelado, cujas frequências de ressonância são as formantes, resultando na fala. Como os sinais da fala variam com o tempo, o processo de análise é efetuado em janelas ou *frames*, onde uma taxa de 30 a 50 janelas por segundo são suficientes para sintetizar fala inteligível com uma boa taxa de compressão.

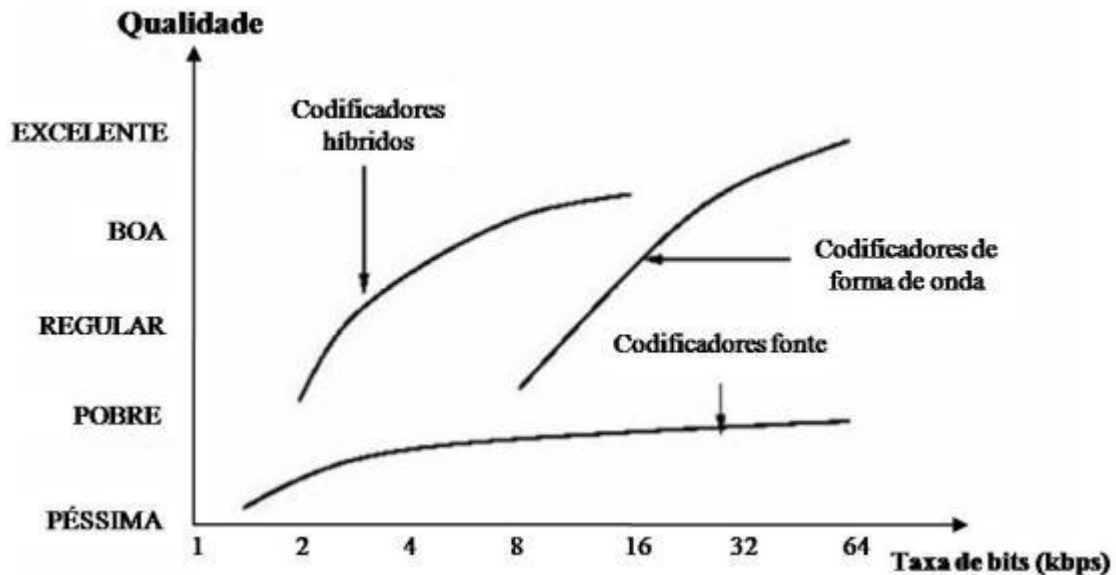


Figura 14 - Curvas de variação típicas da qualidade pela variação da taxa de transmissão para os três tipos de codificadores (*fonte: traduzido de [18]*).

Na próxima seção trataremos dos princípios da análise LPC. As provas matemáticas foram omitidas, mas podem ser encontradas facilmente na literatura [29].

2.3.1 – O modelo LPC

O modelo LPC fornece um bom modelo para a produção de voz, com uma boa aproximação da envoltória espectral do trato vocal, especialmente para sons vocálicos. Para sons não vocálicos, o modelo é menos efetivo, mas devido ao fato de ser um modelo analiticamente tratável, com uma boa separação entre a fonte e o trato vocal, ter fácil implementação, tanto em hardware como em software, é o modelo mais utilizado. Seu desempenho, por exemplo, mostrou-se melhor ou comparável ao modelo de Banco de Filtros na aérea de reconhecimento de voz [29].

A idéia básica por trás do modelo é que um sinal de fala após sofrer o processo de amostragem, $s(n)$, em um dado tempo n , pode ser representado como uma combinação linear de p amostras passadas:

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p), \quad (2.12)$$

onde os coeficientes $\{a_i\}_{i=1..p}$ são considerados constantes durante a análise do quadro [13,29]. Adicionando-se um termo de excitação, podemos converter a equação (2.12) em uma igualdade:

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n), \quad (2.13)$$

Onde G é o ganho de predição e $u(n)$ é a fonte de excitação:

$$u(n) = \begin{cases} \frac{1}{N_0} \sum_r \delta(n - rN_0), & \text{para o som vocálico} \\ \text{ruído Gaussiano aleatório,} & \text{para o som consonantal} \end{cases}. \quad (2.14)$$

Essencialmente, a fonte de excitação é um trem de impulsos quase periódico para sinais vocálicos, representando no domínio da frequência com várias frequências oriundas das junções entre os tubos, sem obstáculos, ou uma fonte de ruído aleatório para sinais consonantais, representando um som articulado ou com obstáculos.

Aplicando-se a Transformada Z na equação (2.13), obtemos:

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z). \quad (2.15)$$

Reorganizando-se obtemos a função de transferência apenas com pólos:

$$H(z) = \frac{S(z)}{U(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)}. \quad (2.16)$$

A equação (2.16) possui a seguinte interpretação: a fonte de excitação, $u(n)$, é escalonada pelo ganho de predição, G , e alimentada ao sistema que contém apenas pólos, $H(z)$, para produzir o sinal de voz, $s(n)$.

Na figura 15 temos a representação do modelo LPC, baseado na equação (2.16), onde os termos em *itálico* correspondem aos parâmetros do modelo [29]. A seleção do tipo de fonte de excitação, correspondendo a um som vocálico ou consonantal, é controlada pelo tipo de sinal de voz. A fonte selecionada, $u(n)$, é escalonada pelo Ganho de predição, G , estimado a partir do sinal de voz. É então alimentada a um Filtro digital variável no tempo apenas com pólos, $H(z)$, controlado pelos parâmetros do trato vocal característicos do sinal de voz que está sendo produzido, que produz o sinal de voz

sintetizada, $s(n)$. Com isso, os parâmetros deste modelo são: o tipo de som, a informação sobre o *Pitch* (para sons vocálicos), o Ganho de Predição e os coeficientes do filtro digital, $\{a_i\}_{i=1..p}$.

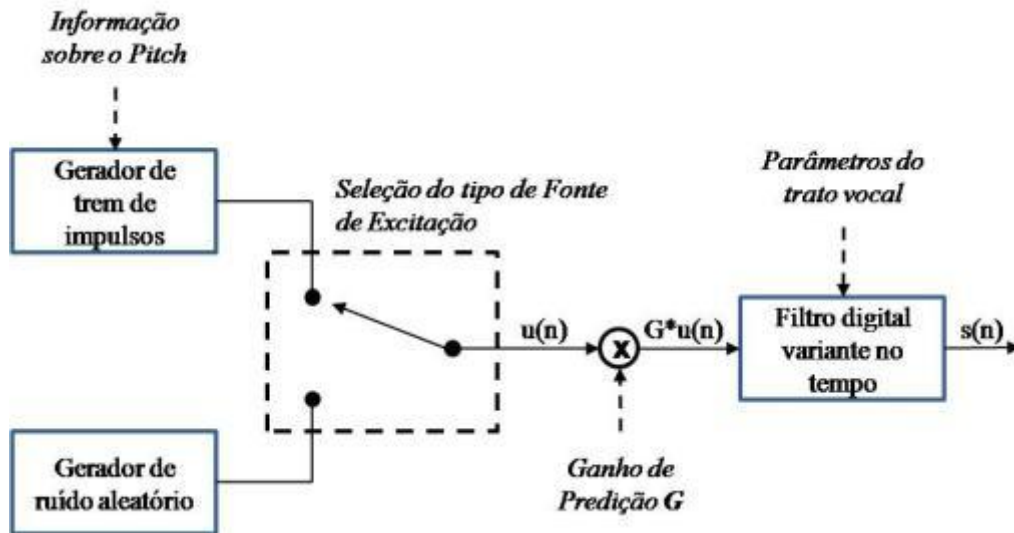


Figura 15 - O modelo LPC (fonte: [29]).

O problema da análise LPC é determinar um conjunto de coeficientes, $\{a_i\}_{i=1..p}$, diretamente a partir do sinal de voz de forma que as propriedades espectrais do filtro digital correspondam às da forma de onda dentro da janela de análise [29]. Como as características espectrais do sinal de voz variam com o tempo, os coeficientes do filtro preditor em um dado tempo, n , devem ser estimados a partir de um pequeno fragmento do sinal nas proximidades do tempo n . Para tal, deve-se encontrar o conjunto de coeficientes que minimize o erro médio quadrático (EMQ) do quadro ou *frame* de sinal a partir deste fragmento¹².

Considerando-se a combinação linear das amostras passadas, $\tilde{s}(n)$, como uma estimativa, definida como:

$$\tilde{s}(n) = \sum_{i=1}^p a_i s(n-i), \quad (2.17)$$

podemos definir o erro de predição como:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{i=1}^p a_i s(n-i) = Gu(n). \quad (2.18)$$

¹² Do ponto de vista prático para codificadores, a análise é efetuada para cada quadro ou *frame* do sinal de voz com duração de 10ms.

Observamos que quando o sinal $s(n)$ é gerado por um sistema linear, como o descrito pela equação (2.13), o erro de predição gerado será igual à excitação escalonada $Gu(n)$, entrada de nosso sistema. Podemos então definir a função de transferência do erro de predição como:

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{i=1}^p a_i z^{-i} . \quad (2.19)$$

Seja o fragmento dos sinais $s(n)$ e $e(n)$, nas proximidades do tempo n de comprimento m , definidos como:

$$s_n(m) = s(n + m) , \quad (2.20)$$

$$e_n(m) = e(n + m) . \quad (2.21)$$

Procuramos o conjunto de coeficientes que minimize o EMQ no tempo n :

$$E_n = \sum_m e_n^2(m) = \sum_m \left[s_n(m) - \sum_{i=1}^p a_i s_n(m - i) \right]^2 . \quad (2.22)$$

Para resolver a equação (2.22), obtendo-se os coeficientes e minimizando o sinal de erro, diferenciamos o sinal E_n para cada coeficiente e igualamos a zero:

$$\frac{\partial E_n}{\partial a_k} = 0 \Big|_{k=1..p} . \quad (2.23)$$

Com isso:

$$\sum_m s_n(m - i) s_n(m) = \sum_{k=1}^p a_k \sum_m s_n(m - i) s_n(m - k) . \quad (2.24)$$

Reconhecendo-se o primeiro membro da equação como termos da covariância de $s_n(m)$, podemos expressar a equação (2.24) como:

$$\phi_n(i, 0) = \sum_{k=1}^p a_k \phi_n(i, k) . \quad (2.25)$$

Essa equação descreve um sistema com p equações e p incógnitas. Podemos então expressar o mínimo EMQ, \hat{E}_n , como:

$$\hat{E}_n = \sum_m s_n^2(m) - \sum_{i=1}^p a_i \sum_m s_n(m) s_n(m - i) = \phi_n(0,0) - \sum_{k=1}^p a_k \phi_n(0, k) . \quad (2.26)$$

Observa-se que o EMQ mínimo possui um componente fixo e outro que depende dos coeficientes do filtro preditor. Além disso, tanto as equações (2.25) quanto a (2.26) possuem somatórios infinitos.

Existem dois métodos para resolver a equação (2.25) [13]:

- a) *Método da Autocorrelação*: possui a vantagem de utilizar um algoritmo eficiente, com complexidade computacional $O(n^2)$, porém não representa adequadamente o espectro do sinal de excitação se este estiver no estado estacionário;
- b) *Método da Covariância*: representa adequadamente o espectro do sinal de entrada quando este está no estado estacionário, mas possui um algoritmo pouco eficiente, $O(p^3)$.

Trataremos apenas do método da Autocorrelação por ser computacionalmente mais viável.

O Método da Autocorrelação [13,29] assume que o segmento do sinal de voz, $s_n(m)$, é nulo fora do intervalo $0 \leq m \leq N - 1$. Esta hipótese é equivalente a multiplicar o sinal de voz, $s(n+m)$, por uma janela, $w(m)$, de comprimento finito, N , ou seja, um processo de janelamento. Então o segmento de sinal de voz pode ser expresso como:

$$s_n(m) = \begin{cases} s(n+m) \cdot w(m), & 0 \leq m \leq N - 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2.27)$$

Antes de continuar, vamos aprofundar um dos conceitos da função janela [30]. Um sinal senoidal puro, por exemplo, quando sofre janelamento, passa a apresentar em seu espectro frequências não nulas além de sua frequência fundamental. Este fenômeno é conhecido como *fuga espectral* [31]. A fuga espectral aparece no espectro do sinal como lóbulos adicionais que se somam aos lóbulos das frequências originais do sinal. Caso o sinal possua frequências próximas, a fuga espectral pode ser grande o suficiente para interferir na resolução do espectro.

Quanto à sensibilidade, existem dois tipos de janela [30]: com alta e baixa taxa de variação. A janela com baixa taxa de variação, como uma janela retangular, é eficiente para representar sinais com frequências bem definidas. A janela com alta taxa de variação é eficiente para representar sinais com grandes variações de amplitude em meio a

ruído. Entre estes extremos existem as janelas moderadas, como a de Hamming, que possuem grande aplicação em codificação de voz para sistemas com banda estreita.

Para exemplificar, na figura 16 temos as formas de onda das janelas retangular e de Hamming e a DFT (*Discrete Fourier Transform*) de uma senóide pura janelada. Observa-se o aparecimento de freqüências não nulas além da freqüência fundamental. A janela retangular, w_r , é definida como:

$$w_r(m) = \begin{cases} 1, & 0 \leq m \leq N - 1 \\ 0, & \text{caso contrário} \end{cases}, \quad (2.28)$$

onde N é o comprimento da janela em número de amostras. A janela de Hamming, w_H , é definida como:

$$w_H(m) = 0,54 - 0,46 \cos \frac{2\pi m}{N-1}. \quad (2.29)$$

Observa-se que a janela de Hamming é capaz de representar um sinal, composto por diversas freqüências com amplitudes diferentes, com mais nitidez que uma janela retangular de mesma largura de banda, pois possui uma fuga espectral com menor intensidade que o seu lóbulo central no gráfico da DFT.

A covariância, por sua vez, pode ser escrita como:

$$\phi_n(i, k) = \sum_{m=0}^{N-1+p} s_n(m) s_n(m + i - k), \text{ para } \begin{cases} 1 \leq i \leq p \\ 0 \leq k \leq p \end{cases} \quad (2.30)$$

A partir do sinal definido em (2.27), o EMQ se torna:

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m). \quad (2.31)$$

Como a equação (2.31) é função apenas de $i-k$, ao invés de duas variáveis independentes, a função covariância se reduz a uma simples função de autocorrelação, ou seja:

$$\phi_n(i, k) = r_n(i - k) = \sum_{m=0}^{N-1+p} s_n(m) s_n(m + i - k), \text{ para } \begin{cases} 1 \leq i \leq p \\ 0 \leq k \leq p \end{cases} \quad (2.32)$$

Aproveitando-se a característica de simetria da função de autocorrelação ($r_n(-k) = r_n(k)$), a equação (2.25) pode ser reescrita como:

$$\sum_{k=0}^p r_n(|i - k|)\hat{a}_k = r_n(i), \text{ para } 1 \leq i \leq p. \quad (2.33)$$

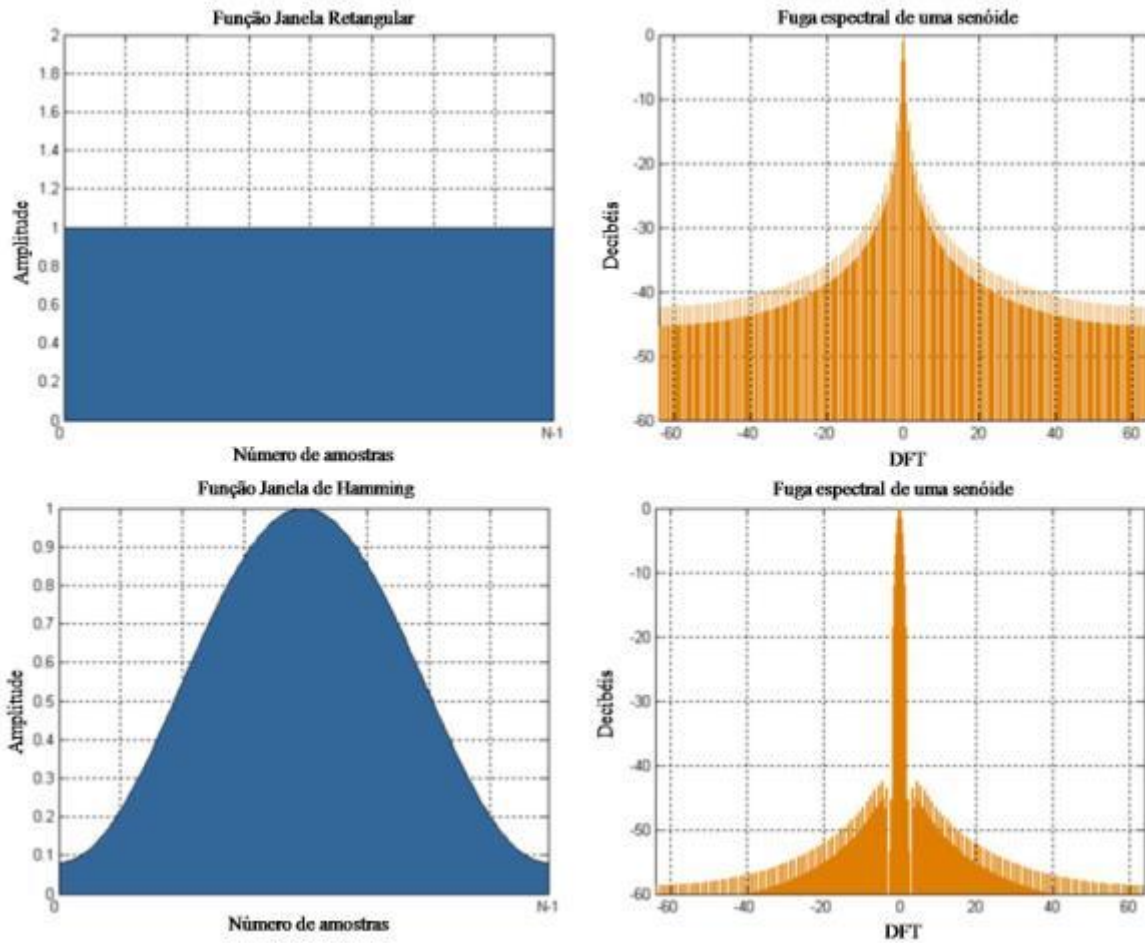


Figura 16 - Forma das Janelas Retangular e de Hamming acompanhadas da DFT de uma senóide pura janelada exibindo o fenômeno de fuga espectral (fonte: [30]).

Na forma matricial, a equação (2.33) pode ser expressa como:

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \dots & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \dots & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \dots & r_n(p-3) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \dots & r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ \vdots \\ r_n(p) \end{bmatrix}. \quad (2.34)$$

A matriz $p \times p$ cujos elementos são valores de autocorrelação é uma matriz simétrica com todos os elementos da diagonal principal iguais, chamada de *matriz Toeplitz* [13,29]. Existem diversos procedimentos para inverter a matriz de Toeplitz e calcular os coeficientes LPC, dentre os quais destacamos o *algoritmo de Durbin* [13] e a *Recursão de Levinson-Durbin* [29].

Uma questão prática relevante é a definição da ordem do filtro LPC [13]. A ordem, p , precisa ser grande o suficiente para representar cada formante como um par de pólos conjugados e possuir pólos adicionais para representar melhor a forma espectral. Seja F_a a frequência de amostragem igual à frequência de Nyquist, o número de formantes, NF , representado pelo filtro LPC é igual a:

$$NF = \frac{F_a/2}{(F_{n+1}-F_n)}, \quad (2.35)$$

onde o denominador representa a distância média entre duas formantes vizinhas. A equação (2.35) pode ser aproximada por:

$$NF = \frac{F_a}{c/l}, \quad (2.36)$$

Onde c é a velocidade do som, igual a 35400 cm/s, e l é o comprimento do trato vocal (para homens, em média, 17,7 cm e para mulheres 14,75 cm). A ordem do filtro, p , pode ser calculada como:

$$p \approx \begin{cases} \frac{F_a}{1000} + p_a, & \text{para homens} \\ \frac{F_a}{1200} + p_a, & \text{para mulheres} \end{cases}, \quad (2.37)$$

onde p_a são os pólos adicionais, geralmente dois ou três.

Para demonstrar outra vantagem da análise LPC, suponha que um sinal de voz é amostrado com uma frequência de 10 kHz e sejam utilizados 16 bits para representar sua amplitude. A uma taxa de 160kbps, serão necessários aproximadamente 196kbytes para armazenar 10s. Usando a análise LPC, com um filtro de décima ordem ($p=10$) e 100 vetores por segundo, serão necessários 19,6 kbytes para armazenar o mesmo tempo a uma taxa de 16 kbps. Com isso, é possível reduzir as taxas de transmissão e a capacidade de armazenamento para um sistema de digitalização de voz.

3. QUALIDADE DA CHAMADA VoIP

As redes telefônicas convencionais evoluíram para fornecer um serviço ótimo para aplicações de voz que requerem baixos níveis de atraso e *jitter* em conjunto com uma pequena banda constante [39]. Redes IP, por sua vez, foram construídas, inicialmente, para dar suporte a aplicações de transferência de dados que não exigem resposta em tempo-real, como e-mail e transferência de arquivos. A convergência dessas duas redes e o surgimento de aplicações multimídia em tempo real, como VoIP, exigiram o desenvolvimento de mecanismos para mensurar e garantir a qualidade de serviço (QoS). No caso VoIP, essencialmente a qualidade da chamada.

A Qualidade da chamada pode ser definida como uma forma de avaliar e descrever a fidelidade, inteligibilidade e características do sinal da voz, bem como o desempenho dos métodos de transporte de tal informação [39]. Inerentemente pode ser uma estimativa quantitativa ou qualitativa do som e da qualidade da ligação.

A necessidade de estimar precisamente a qualidade da fala surge em diferentes fases do processo e implementação de um produto VoIP, como durante o projeto de codificadores, supressores de eco e terminais VoIP [11]. A análise da qualidade de uma ligação também é importante para o usuário final, interessado em usufruir de uma ligação com qualidade no mínimo igual à oferecida pelas redes telefônicas convencionais. Este foi um ponto crucial para a aceitação desta tecnologia, aliado à tarifação com baixo custo.

Neste capítulo são analisados conceitos básicos da análise da qualidade da voz, como fontes de degradação e tipos de análise.

3.1 – Componentes da Qualidade de Voz

Muitos fatores influenciam a percepção da qualidade de uma chamada telefônica, mas, em um nível mais abstrato, a qualidade de uma chamada pode ser desmembrada em três componentes principais [39]:

- *Qualidade de Serviço*: relacionados com o provedor escolhido pelo usuário. Inclui aspectos como: serviços oferecidos, disponibilidade, portabilidade, mobilidade, confiabilidade e preço;
- *Qualidade do som*: relacionados com a qualidade do som. Inclui aspectos como nitidez, ruído, distorção, volume etc;
- *Qualidade da conversação*: relacionados com aspectos que influenciam a qualidade do som, como: atraso ponto-a-ponto, supressor de eco, detecção de atividade de voz, etc.

Todos estes componentes são comuns tanto para chamadas telefônicas convencionais, celulares ou VoIP. São aspectos que um usuário geralmente utiliza para dar uma opinião sobre a qualidade de voz oferecida por um provedor, associando a cada componente um peso [39]. Por exemplo, um usuário de celular frequentemente tolera problemas na qualidade do som porque a mobilidade é um fator com maior peso.

Pode-se perceber que muitos aspectos da qualidade do serviço estão mais atados a decisões comerciais do provedor do que aspectos técnicos sobre a operação e desempenho do hardware da rede. No entanto, as qualidades do som e da conversação estão intimamente relacionadas com os detalhes do desempenho da rede. Este é um dos motivos pelo qual a Qualidade da voz é definida como uma medida quantitativa e qualitativa da qualidade de som e conversação de uma chamada telefônica [39].

Focando a qualidade das ligações VoIP, os três elementos que afetam majoritariamente são a *clareza*, o *atraso ponto-a-ponto* e o *eco*[39]. A clareza se relaciona com a fidelidade do som transmitido ao som original e a ausência de distorção. Pode ser descrita por inteligibilidade, embora haja uma sutil diferença. Por exemplo, frases completas são mais bem compreendidas (mais inteligíveis), como um fluxo lógico, do que uma seqüência de palavras avulsas, mesmo que a seqüência seja menos distorcida, ou seja,

possua maior clareza [39]. O atraso ponto-a-ponto é o tempo necessário para que o sinal de voz parta do interlocutor ao ouvinte. É a soma de todos os atrasos pelo qual o sinal de voz é submetido para chegar ao ouvinte. Inclui atrasos de conexão, armazenamento, conversão, etc. O eco é o retorno ao interlocutor do sinal de voz transmitido.

Na figura 17 temos uma representação espacial do espaço de qualidade de voz, tendo como o eixo o incremento dos três elementos que afetam a qualidade da ligação VoIP. Estes três elementos são independentes e ortogonais, embora haja certa relação entre o eco e o atraso, visto que o eco é mais perceptível em redes com atraso acima de certo limiar [39]. Pensando-se na qualidade da voz com um ponto, pode-se perceber que quanto mais afastado da intersecção dos três eixos, menor a qualidade. Fixando-se as contribuições de cada eixo, a qualidade de voz será representada por um plano formado pela intersecção das contribuições. O agrupamento destes elementos é baseado no fato de que a maioria dos usuários avalia a qualidade como ruim se apenas um destes três elementos é insatisfatório [39].



Figura 17 - Espaço de qualidade de voz de uma ligação VoIP (fonte: [39]).

Uma vez expostos os conceitos sobre qualidade de voz, analisaremos as fontes de degradação da qualidade.

3.2 – Fontes de degradação da qualidade

Um sinal de voz sofre a ação de diversos fatores quando viaja entre dois pontos em uma rede. Eles podem ser agrupados em três grandes categorias [11]:

- *Ambientais*: geralmente relacionados com o ambiente de captação do sinal de voz, como: microfone, captação do som, acústica da sala, ruído de fundo, etc;
- *Processamento de sinais*: associados à compressão do sinal de voz, variações no ganho etc;
- *Rede*: associados ao estado de transmissão de pacotes na rede: *jitter*, atraso, perda de pacotes, eco, etc.

As fontes ambientais são os mesmos fatores de degradação que a rede telefônica convencional também está submetida. Consiste basicamente em ruído criado por uma fonte externa do ambiente ou resultado da acústica do ambiente ou da qualidade do equipamento utilizado.

As fontes da rede são oriundas de deficiências no tráfego de informações na rede. Perdas de pacotes geralmente ocorrem nos roteadores devido a um alto volume de pacotes, gerando surtos de perda, ou devido a uma quebra na conexão, geralmente mais próximo do usuário final do que no núcleo da rede [11]. Para evitar tal problema, foram desenvolvidos diversos mecanismos de roteamento. Como exemplo, a IETF propôs um esquema de encaminhamento de pacotes, o MPLS (*Multi-Protocol Label Switching*) que, baseado em rótulos no cabeçalho do pacote IP de acordo com o tipo de aplicação, assegura a transmissão de determinados pacotes sem perdas ou atrasos perceptíveis [40].

O aspecto da latência (atraso) é difícil de mensurar, dependente da cadência do interlocutor e da percepção do ouvinte em perceber lapsos temporais, e pode causar efeitos severos na qualidade da transmissão como eco e sobreposição (*talker overlap*) [11]. A recomendação ITU-T G.114 define valores em milissegundos aceitáveis para o atraso [41,42]:

- *Até 150 ms*: excelente;
- *De 150 a 300 ms*: bom;
- *De 300 a 450 ms*: pobre;
- *Acima de 450 ms*: inaceitável.

As parcelas que contribuem para o atraso podem ser agrupadas em duas categorias [41]: gerados pela rede e pela formação do pacote IP, ilustrados na figura 18.

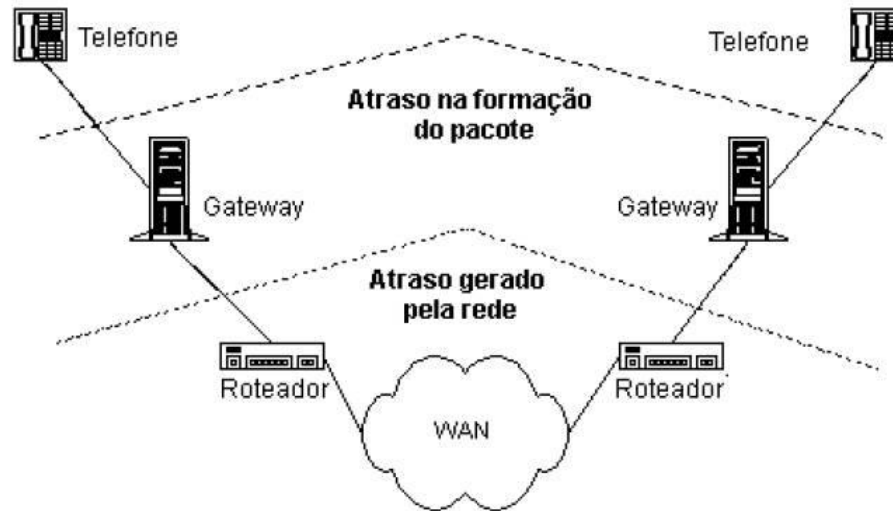


Figura 18 - Tipos de atraso que formam os fatores de rede da degradação da qualidade de voz (*fonte: [41]*).

O atraso gerado pela rede é o tempo necessário para o transporte ponto-a-ponto do pacote de voz. Estima-se que seja a soma do tempo gasto em todos os roteadores da rede mais os tempos gastos nos próprios terminais. Pode ser dividido em [41]:

- *Atraso de acesso ao meio*: é incluído toda vez que um pacote passa por um meio físico diferente;
- *Atraso de roteamento*: como os roteadores determinam o caminho dos pacotes na rede, os pacotes recebem este atraso enquanto esperam na fila até serem enviados ou descartados;
- *Atraso em firewalls e proxy*: com a finalidade de dar segurança adicional ao sistema, equipamentos de *firewall* e *proxy* armazenam pacotes para verificação de conteúdo, adicionando um atraso que pode ser considerável, caso o equipamento não tenha uma boa reposta de processamento.

O atraso na geração de pacotes pode ter dividido em [41]:

- *Atraso devido a conversões*: adiciona-se atraso nas conversões de sinais de voz dentro dos *gateways*, tipicamente da ordem de 1 ms;
- *Atraso de processamento*: relacionado aos codificadores, inclui atrasos gerados pelo processamento dos quadros de voz e seu empacotamento;
- *Atraso devido à manipulação de pacotes*: inclui atrasos gerados pelo armazenamento de pacotes, visando troca/análise de informações no cabeçalho, medida contra *jitter* que atrapalha o processo de decodificação, etc.

O eco ocorre quando o sinal de voz do interlocutor retorna pelo canal de recebimento. Existem várias causas para o eco, indo desde problemas acústicos (como no caso do telefone convencional) a problemas na conexão de dispositivos analógicos de transmissão [39]. Tipicamente, se o tempo entre a frase originalmente falada e o seu eco for curto (25 a 30 ms) ou se o nível do sinal de eco for muito baixo (aproximadamente -25 dB), não haverá queda na qualidade de conversação. Na figura 19 temos um gráfico mostrando o limiar de percepção do eco.

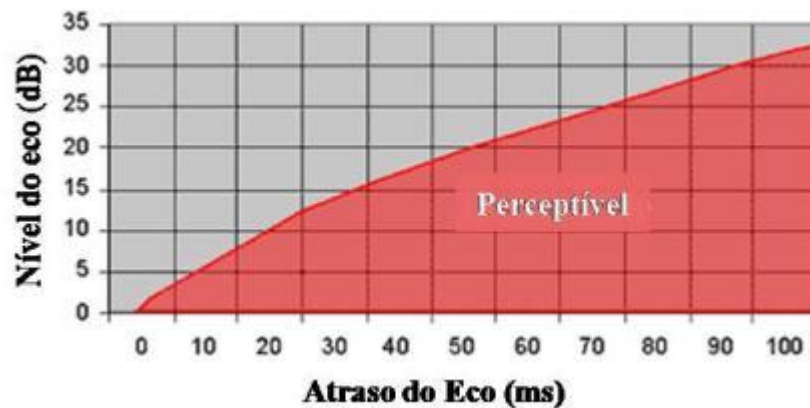


Figura 19 - Gráfico mostrando o nível e o atraso do sinal de eco que o tornam audível (*fonte: [39]*).

No caso VoIP, o atraso ponto-a-ponto frequentemente excede o limiar mencionado e se torna audível. Para anular o eco, os codificadores geralmente contam com mecanismos supressores de eco, que estimam o atraso ponto-a-ponto e o sinal de eco esperado, subtraindo-o do sinal recebido. Tais mecanismos geralmente assumem que as condições de rede são lineares invariantes no tempo, podendo levar a distorções quando há alguma mudança [39].

Vale a pena salientar que existe um eco, com um atraso de 28 ms, incluso no sinal recebido para fazer com que o interlocutor escute sua própria voz enquanto fala. Este tipo de eco não interfere na qualidade da conversação [39].

Descartando-se as fontes ambientais e de rede, as fontes de processamento de sinais são um fator limitante da qualidade máxima. Caso a transmissão esteja sujeita a condições ambientais e de tráfego ótimas, a qualidade estará limitada à fidelidade com que o processamento do sinal de voz reconstrói as informações.

Os codificadores de voz surgem no limiar entre todas as fontes de degradação. São responsáveis pelo processamento do sinal de voz, limitando a qualidade máxima, através do algoritmo e da complexidade de sua implementação, e incorporando medidas de atenuação dos fatores de rede, com algoritmos de tratamento de perda de pacotes, e ambientais, com atenuação de ruído de fundo. São também responsáveis pelo equilíbrio *qualidade versus transmissão*: quanto maior a qualidade, ou seja, quanto menor for o efeito da fonte de processamento de sinais maior será o impacto na transmissão, ou seja, um maior volume de dados implica em maior susceptibilidade a atrasos, *jitter* e perda de pacotes.

Por estes motivos, a análise da qualidade de codificadores é uma etapa importante. A comparação entre resultados auxilia na escolha de um codificador mais adequado a um ambiente, tipo de rede ou mesmo a uma relação custo-benefício.

3.3 – Tipos de análise

Os métodos de análise podem ser divididos em dois tipos [11]: subjetivos e objetivos. Métodos objetivos comparam características do sinal transmitido com as do sinal recebido, enquanto métodos subjetivos utilizam ouvintes treinados para dar uma nota à qualidade da ligação.

Os testes subjetivos são mais relevantes porque avaliam a qualidade percebida por humanos. Os testes subjetivos são dispendiosos por exigirem a participação de um número estatisticamente significativo de pessoas cobrindo diversos aspectos, tais como faixas etárias, sexo, culturas etc., e dependem intimamente do grupo escolhido para o teste. Os testes objetivos ou automáticos, com um custo menor e maior simplicidade, são mais susceptíveis ao processamento e complexidade da modelagem [11].

Analisaremos alguns métodos de análise objetivos e subjetivos.

3.3.1 – Métodos Subjetivos

Até recentemente, os métodos subjetivos eram a ferramenta primária para avaliar a qualidade de voz para o usuário final [11]. O ponto forte deste tipo de análise é que todos os tipos de degradação são analisados, analisando a qualidade de voz como o

usuário a percebe. Por outro lado, devido ao grande custo financeiro, tempo necessário e inviabilidade para testes em tempo real durante o planejamento da rede, tornam este tipo de teste nem sempre desejável [11]. Outra dificuldade é que é praticamente impossível repetir o mesmo ambiente experimental.

Para se realizar estes testes, uma amostra estatisticamente significativa do sistema deve ser escolhida e uma grande quantidade de ouvintes deve avaliar amostras de fala que passaram por uma rede simulada [11]. Ou seja, os ouvintes devem avaliar amostras de fala simulando o uso do codificador em tempo real.

Existem inúmeros testes subjetivos, mas os mais populares são [11]: Teste A-B, teste ACR (*Absolute Category Rating*), teste DCR (*Degradation Category Rating*) e o teste de conversação.

O teste A-B é puramente um teste de comparação [11]. Aos ouvintes são dadas duas opções, A e B, para avaliar, dado o aspecto desejado. Os sistemas sob teste são comparados por um número de diferentes amostras para cada uma das condições. A vantagem deste tipo de teste é a sua simplicidade, baixo custo e contraste entre as alternativas. No entanto, há apenas uma distinção relativa entre as alternativas, ou seja, qual é a melhor das alternativas sob um determinado aspecto.

O teste ACR é um dos mais conhecidos [11]. Um grande número de indivíduos avaliam a qualidade, de 1 (ruim) a 5 (excelente), de amostras de voz foneticamente equilibradas. São analisados diferentes tipos de fatores, como qualidade, inteligibilidade e volume. As notas são tabeladas e uma média MOS (*Mean Opinion Score*) é obtida. Normalmente são escolhidos indivíduos leigos, não treinados, para o teste, mas é possível usar um número menor de indivíduos treinados capazes de distinguir a fonte e o tipo de degradação. Ambos os conjuntos possuem sua relevância, mas o conjunto de indivíduos leigos é mais adequado, pois se aproxima mais da média da população.

O teste DCR é similar ao ACR, mas é pedido ao grupo de indivíduos para avaliar a degradação da voz [11]. São utilizados dois pares de amostras, A (referência de qualidade) e B (a ser testado) [47]. O resultado é apresentado como uma média DMOS (*Degradation Mean Opinion Score*) na mesma escala de 1 (degradação muito audível) a 5 (degradação pouco audível), da comparação entre as amostras A e B.

Os testes MOS são considerados os mais confiáveis e precisos, sendo os procedimentos definidos pela recomendação ITU-T P.800 [11,47].

Os testes de conversação são ainda mais complexos e dificilmente utilizados [11]. Os indivíduos participam de cenários de comunicação interativa, sendo direcionados a completar uma série de atividades via sistema telefônico. Efeitos que podem potencialmente afetar a qualidade da voz, como latência e eco, são introduzidos e dados sobre como o ouvinte avalia a qualidade são coletados. São importantes porque o resultado representa uma medida mais realista da qualidade do sistema operando em tempo real, mas consomem muito tempo e dinheiro para serem executados. Além disso, este tipo de teste exige que o sistema esteja completamente implementado.

Na tabela 7 temos o resumo das vantagens e desvantagens dos testes subjetivos mais conhecidos.

Tabela 7 - Comparação entre as vantagens e desvantagens dos testes subjetivos mais conhecidos (fonte: [11]).

Teste	Vantagens	Desvantagens
A-B	Simplicidade e baixo custo	Distinção relativa entre as opções
MOS (ACR)	Precisão e procedimentos bem definidos	Custo e não avalia fatores de degradação da rede
DMOS (DCR)	Precisão e procedimentos bem definidos	Custo, avaliação relativa entre duas amostras e fatores de degradação da rede não são levados em consideração
Conversação	Avalia o sistema mais próximo da situação real, levando fatores de degradação de rede em conta	Exige grande custo e tempo para os testes, juntamente com o sistema completamente implementado

3.3.2 – Métodos Objetivos

Métodos objetivos podem ser utilizados para medir um grande número de parâmetros diferentes [11], como o atraso. Quando aplicados corretamente, os testes objetivos são mais precisos que os testes subjetivos para avaliar o atraso. No entanto, também podem ser utilizados para estimar resultados dos testes subjetivos, sendo esta a aplicação mais controversa [11].

Em um esforço pela repetibilidade dos resultados e diminuição dos gastos, inúmeros métodos objetivos foram desenvolvidos. São divididos em métodos *intrusivos* e *não-intrusivos* (ou passivos) [11].

Dentre os métodos intrusivos mais populares temos: PASM (*Perceptual Analysis Measurement System*) desenvolvido pela *British Telecom*, PSQM (*Perceptual Speech Quality Measure*) da recomendação ITU-T P.861, PSQM+ (versão melhorada do PSQM) e PESQ (*Perceptual Evaluation of Speech Quality*) da recomendação ITU-T P.862.

Atualmente, o método PESQ, baseado na combinação dos métodos PASM e PSQM, é o mais popular e aceito como um dos mais precisos [11]. Os algoritmos possuem a mesma base. Primeiro, é criado um modelo do sistema auditivo humano para simular a capacidade auditiva. Em seguida, modela a distorção audível aplicando ao modelo auditivo e calcula uma medida da distorção entre o sinal original e o processado, gerando notas similares às dos métodos subjetivos.

A desvantagem em comum para os métodos objetivos intrusivos é que são sensíveis ao atraso e manipulações na escala do tempo, efeitos de processamento, etc. Além disso, exige que a presença de um sinal de referência, que não representa um problema em ambientes de laboratório, mas exigiria a utilização de uma ligação no mundo real.

Há algumas controvérsias na literatura quanto ao método PESQ, afirmando que não é possível assumir que uma boa nota PESQ corresponde a uma alta qualidade de voz ou vice-versa [11]. Basicamente, o problema surge da tentativa de mapeamento dos métodos subjetivos pelos métodos objetivos.

Devido à necessidade de enviar arquivos pela rede, os métodos objetivos intrusivos são praticamente inviáveis [11]. Para resolver tal problema, foram desenvolvidos vários métodos para estimar a qualidade da voz baseado em medições da rede ou o sinal recebido. Dentre os métodos objetivos não-intrusivos, destacamos o padrão ITU-T P.563 e o *E-model* (ITU-T G.107).

O padrão ITU-T P.563 é um método não-intrusivo baseado na medição do sinal recebido [11]. Consiste de uma extensiva análise de forma de onda do sinal processado recebido. Sem que comparações com o sinal original sejam feitas, inúmeras métricas MOS são computadas para obter uma métrica consistente. As desvantagens deste método são a alta carga computacional e a possibilidade de resultados pouco precisos.

O *E-model* foi originalmente projetado para auxiliar engenheiros com o planejamento de transmissões, mas é adaptado para avaliar a qualidade de ligações VoIP [11]. É projetado para incorporar diversos parâmetros que afetam a qualidade de voz e gerar uma escala da qualidade de transmissão (R), que varia de 0 (ruim) a 100 (excelente) [48]. O cálculo de R assume que os efeitos da latência são aditivos e inclui um fator base que é derivado de aspectos como níveis de ruído e volume. Outros fatores na equação representam o acoplamento de sinais, que ocorre simultaneamente com a fala e os equipamentos da rede, e um fator que é baseado nas expectativas da qualidade de certo tipo de ligação. Este método permite a avaliação de múltiplas chamadas sem afetar a rede e exige baixa carga computacional [11]. No entanto, opera apenas com os codificadores conhecidos e não fornece avaliação de outros parâmetros. Outro problema é que é difícil prever o resultado do efeito combinado dos parâmetros, resultando um teste menos preciso que os intrusivos, principalmente em ambiente real.

Na tabela 8, temos o resumo das vantagens e desvantagens dos métodos objetivos intrusivos e não-intrusivos.

Do ponto de vista da correlação com os resultados dos testes subjetivos, podemos classificar os métodos objetivos da seguinte forma [11]:

- 1°. PESQ, com correlação tipicamente na faixa de 0,9 a 0,95;
- 2°. ITU-T P.563, tipicamente 0,7 a 0,9;
- 3°. E-model, tipicamente 0,6 a 0,8.

Embora os métodos não intrusivos sejam menos precisos, são mais úteis em cenários reais do que os intrusivos, principalmente para monitoramento de variações na qualidade da ligação.

Tabela 8 - Comparação entre as vantagens e desvantagens dos testes objetivos mais conhecidos (fonte: [11]).

Teste	Vantagens	Desvantagens
Intrusivos (PSQM, PSQM+, PAMS, PESQ)	Alta repetibilidade e baixo custo	Sensível ao atraso, manipulações na escala do tempo, processamento, acoplamento de equipamentos e necessidade de sinal de referência
Não-Intrusivo ITU-T P.563	É aplicado apenas no receptor	Impreciso, exige alta carga computacional e possui complexidade considerável
Não-intrusivo E-model	Possui baixo custo e pode monitorar diversas ligações sem afetar a rede	Funciona apenas para os codificadores conhecidos, não leva em conta alguns parâmetros e a combinação dos parâmetros avaliados é difícil de prever

4. ANÁLISE COMPARATIVA

Neste capítulo é descrita uma análise comparativa dos codificadores G.722 [25], G.723.1 [26], G.726 [28], G.728 [33], G.729A [34], iLBC [36] e Speex [38] tendo como referência os resultados do algoritmo PCM [22]. Na tabela 9 temos a estimativa dos resultados dos testes MOS dos codificadores utilizados nesta análise. Não foi possível implementar o codificador G.727 [31] em tempo hábil para a análise. Todos os codificadores analisados eram ou estavam usando o modo banda estreita.

Tabela 9 - Comparação entre as estimativas de resultados MOS dos codificadores utilizados na análise comparativa sugerida (fontes: [43,44,45,46]).

Codificador	Algoritmo	Taxas de transmissão (kbps)	MOS ¹³
G.711	PCM	64	4,4
G.722	SB-ADPCM	64 ¹⁴	4,5
G.723.1	MP-MLQ e ACELP	6,3 e 5,3	4
G.726	ADPCM	40, 32, 24 e 16	4,3
G.728	LD-CELP	16	4,2
G.729A	CS-CELP	8	4
iLBC	iLBC	13,33 e 15,2	4,1
Speex	CELP	2 a 44	3,8

Esta dissertação tem como objetivo a análise do impacto do processamento do sinal na qualidade de uma comunicação digital. São analisadas as características básicas do modelo fonte-filtro: a frequência fundamental e as três primeiras formantes, dos fonemas da Língua Portuguesa falada no Brasil. Apenas os fatores de

¹³ Todos estes valores foram estimados pelo método objetivo PSQM.

¹⁴ Somando-se com a taxa de transmissão do canal auxiliar.

degradação de processamento de sinal são analisados, visto que existem na literatura diversas análises dos fatores de degradação da rede. Os fatores ambientais também não são analisados, sendo postergada tal análise para trabalhos futuros.

A metodologia da análise consiste em comparar as frequências fundamental e das formantes de sinais processados por codificadores, contidos na tabela 40, com sinais PCM uniforme, amostrados a 8 kHz e quantizados com 16 bits por amostra, obtidos durante o processo de gravação e chamados neste trabalho de *sinais originais*. O sinal original consiste em uma frase, totalizando um conjunto de doze frases analisadas. A escolha de frases, ao invés de fonemas separados, é uma escolha natural, visto que se aproxima mais de uma conversação.

Cada sinal original é então dividido em intervalos contendo grupos de fonemas, geralmente análogo à separação silábica das palavras. Esta separação em intervalos, chamada de grade ou *grid*, é então utilizada na análise dos sinais processados.

São investigadas, também, possíveis variações no processamento dado ao sexo e a idade. Para tal foram criados três grupos de homens e três grupos de mulheres, conforme a tabela 10, todos do estado de Pernambuco. A separação em grupos de faixa etária distintas foi observada em um artigo [49] e se mostrou interessante aplicar esta metodologia para este estudo. Foram coletados, também, os tempos de processamento de cada codificador e os tempos médios de cada intervalo da grade analisado visando analisar possíveis variações entre os grupos.

Tabela 10 - Definição dos grupos de homens e mulheres analisados.

	Faixa de idade	Tamanho do Grupo
Grupo 1	14 - 34	2
Grupo 2	35 - 59	2
Grupo 3	60 - 80	2

Os resultados são mostrados em forma de gráficos e tabelas por grupo. Para um melhor acompanhamento, são exibidos os resultados médios de cada grupo, bem

como os de um indivíduo de cada grupo. Assim, não há uma sobrecarga de gráficos e informações¹⁵.

Antes de apresentar os resultados da análise são descritos os detalhes do arranjo experimental, como *software* e *hardware* usados para a aquisição de voz, o *software* usado na análise comparativa e detalhes sobre as amostras de voz. Todos os comandos utilizados nos programas podem ser encontrados no DVD em anexo.

4.1 – Gravação das amostras de voz

Pesquisas indicaram a ausência de frases de referência na Língua Portuguesa para tal análise. Logo, foram criadas doze frases curtas, com base nas tabelas 1 e 3, contendo todos os fonemas consonantais e vocálicos básicos da Língua Portuguesa. Na tabela 11 temos as doze frases criadas para a análise, bem como os intervalos da grade analisados. Observa-se nas frases também a presença de semivogais, hiatos e ditongos. Para facilitar a visualização dos fonemas nos gráficos, os grupos de fonemas iniciais de cada palavra terão seus grafonemas começando com letra maiúscula.

Durante a gravação das frases foi utilizado o *headset Goldship* da *Leadership*[®], muito utilizado em *Lan Houses* e por usuários domésticos para fins de comunicação digital. Neste *headset*, o microfone é acoplado ao fone de ouvido e permite um melhor controle sobre a posição do mesmo. A sensibilidade do microfone é de -20 dB.

O programa utilizado para a gravação foi o Praat, *software* gratuito desenvolvido por Paul Boersma e David Weenink [50]. Este programa também foi utilizado para a obtenção das frequências dos sinais analisados.

Todos os arquivos de áudio foram gravados em formato *wave* (.wav), mono, frequência de amostragem 8 kHz com 16 bits por amostra utilizando a codificação PCM uniforme. O microfone sempre estava na altura da boca do indivíduo e longe do fluxo de ar das narinas para evitar a captação de ruído. Os arquivos eram gravados com um trecho de silêncio de aproximadamente três segundos antes da pronúncia da frase, e

¹⁵ Os resultados individuais de cada membro dos grupos estão disponibilizados no DVD em anexo, bem como todos os sinais originais e processados utilizados na análise. Pode-se ouvir cada arquivo para obter sua própria referência perceptual sobre a qualidade do processamento do sinal.

aproximadamente um segundo após a sua pronúncia. Assim, seria possível verificar os fatores ambientais durante a gravação. Para garantir que as condições do trato vocal do indivíduo fosse a mesma para todas as frases e que as condições ambientais fossem as mesmas para as doze frases, o processo de gravação foi seqüencial, uma frase por vez, e efetuado no mesmo período do dia. Geralmente, o processo de gravação não levava mais de dez minutos. Além disso, utilizou-se o mesmo notebook para a gravação de todas as amostras, para se evitar diferenças de captação entre placas de áudio distintas. O módulo de supressão de ruídos, disponibilizado no *driver* da placa de som, foi desativado durante o processo de gravação.

Tabela 11 - As doze frases criadas para a análise.

Frase 1	<i>A sétima arte deslumbra multidões</i>													
	A	Sé	ti	ma	Ar	te	Des	lum	bra	Mul	ti	dões		
Frase 2	<i>São raros políticos preocupados com o povo</i>													
	São	Ra	ros	Po	lí	ti	cos	Preo	cu	pa	dos	Com	O	Po
Frase 3	<i>Aquela peça nunca quebrou</i>													
	A	que	la	Pe	ça	Nun	ca	Que	brou					
Frase 4	<i>O gato entrou na casa atrás do rato</i>													
	O	Ga	to	En	trou	Na	Ca	sa	A	trás	Do	Ra	to	
Frase 5	<i>O sultão disse xeque-mate</i>													
	O	Sul	tão	Di	sse	Xe	que	Ma	te					
Frase 6	<i>Tire a fita do mapa</i>													
	Ti	re	A	Fi	ta	Do	Ma	pa						
Frase 7	<i>Ainda existem fogões a lenha</i>													
	A	in	da	E	xis	tem	Fo	gões	A	Le	nha			
Frase 8	<i>A bandeira é um símbolo nacional</i>													
	A	Ban	dei	ra	É	Um	Sím	bo	lo	Na	cio	nal		
Frase 9	<i>A centelha da vida ainda arde</i>													
	A	Cen	te	lha	Da	Vi	da	A	in	da	Ar	de		
Frase 10	<i>Paguei a viagem de jato hoje</i>													
	Pa	guei	A	Via	gem	De	Já	to	Ho	je				
Frase 11	<i>O copo está cheio de meias</i>													
	O	Co	po	Es	tá	Che	io	De	Me	ias				
Frase 12	<i>Araras fugiram da chácara Guimarães</i>													
	A	ra	ras	Fu	gi	ram	Da	Chá	ca	ra	Gui	ma	rães	

Uma funcionalidade do Praat é a possibilidade de verificar a qualidade do sinal gravado através da visualização do seu espectrograma. Na figura 20 temos um exemplo de um espectrograma de um sinal gravado. O gráfico da parte superior é um

gráfico de amplitude pelo tempo, onde é possível perceber claramente os instantes de tempo em que o indivíduo fala. A parte inferior contém o espectrograma (frequência pelo tempo) do sinal, onde os pontos em vermelho representam as formantes, enquanto a linha em azul representa a vibração das cordas vocais, ou seja, a frequência fundamental. Neste caso, é possível perceber facilmente o instante de tempo em que a frase é pronunciada nos dois gráficos, principalmente no espectrograma. Já na figura 21, o ruído do ambiente é tão intenso que o espectrograma mostra formantes mesmo em regiões onde não há fala. Este é um exemplo de gravação que foi descartada.

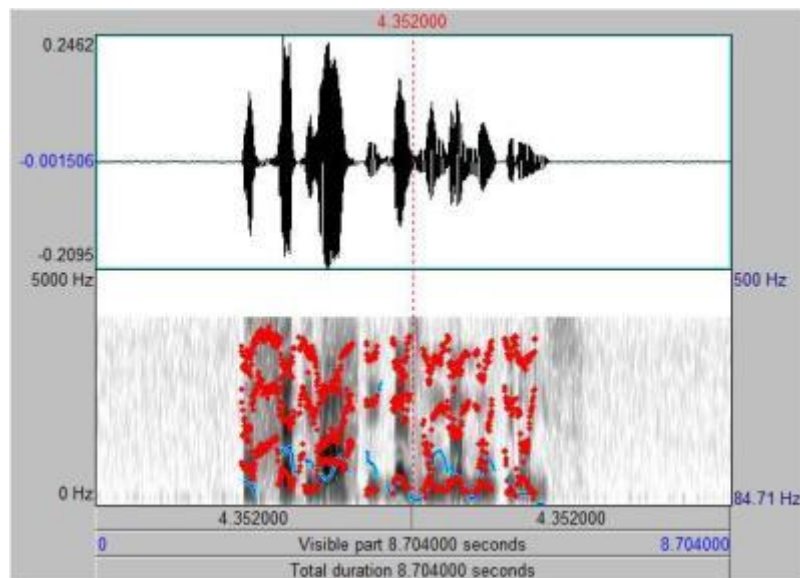


Figura 20 - Exemplo do espectrograma de um sinal aceito para a análise.

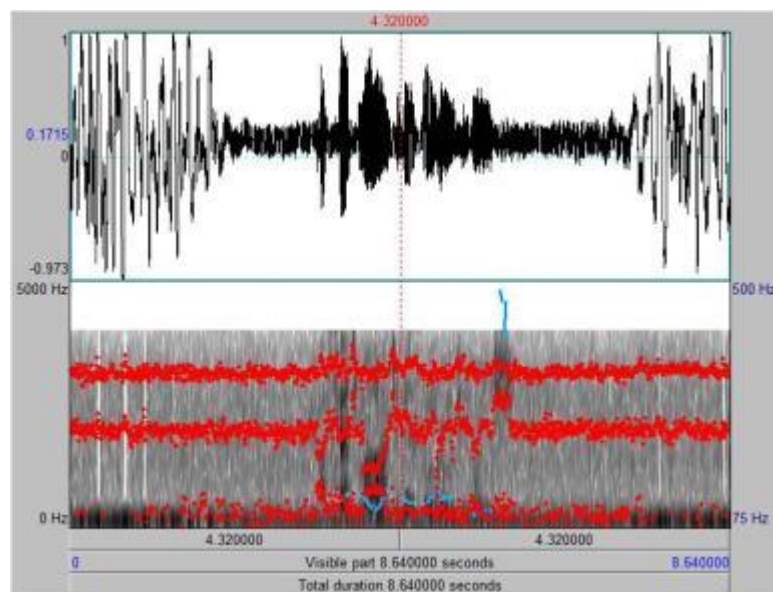


Figura 21 - Exemplo do espectrograma de um sinal rejeitado para a análise.

Vale salientar a dificuldade para se obter um bom local de gravação. A sensibilidade do programa e a baixa qualidade do microfone utilizado é tal que o barulho da movimentação de água nos canos de uma casa pode ser captado. Todo o processo de gravação foi efetuado no mesmo cômodo com as mesmas posições relativas visando manter uma mesma impedância acústica relativa. Eventualmente a captação de ruídos de fundo era atenuada com a diminuição da sensibilidade do microfone, o que gerava arquivos com baixo volume, mas ainda audíveis. Houve dias em que a gravação não pode ser realizada, pois havia a detecção de ruído intenso de origem desconhecida. Em uma das vezes, em uma casa a aproximadamente dez metros estava sendo utilizado um equipamento de jato de água para limpar a calçada e durante a gravação captou-se forte ruído de fundo.

O sinal original foi obtido a partir da extração da região de fala do arquivo de gravação. Tomando-se a figura 20, como exemplo, a região falada é extraída e mostrada na figura 22. Todo o processo de extração é efetuado por uma combinação da visualização do espectrograma e percepção auditiva do sinal de gravação. Por exemplo, a região mais escura na figura 61 logo após os pontos vermelhos corresponde a pronúncia do fonema /s/. Percebe-se que no gráfico de amplitude pelo tempo a presença não é notada, embora perceptualmente seja audível.

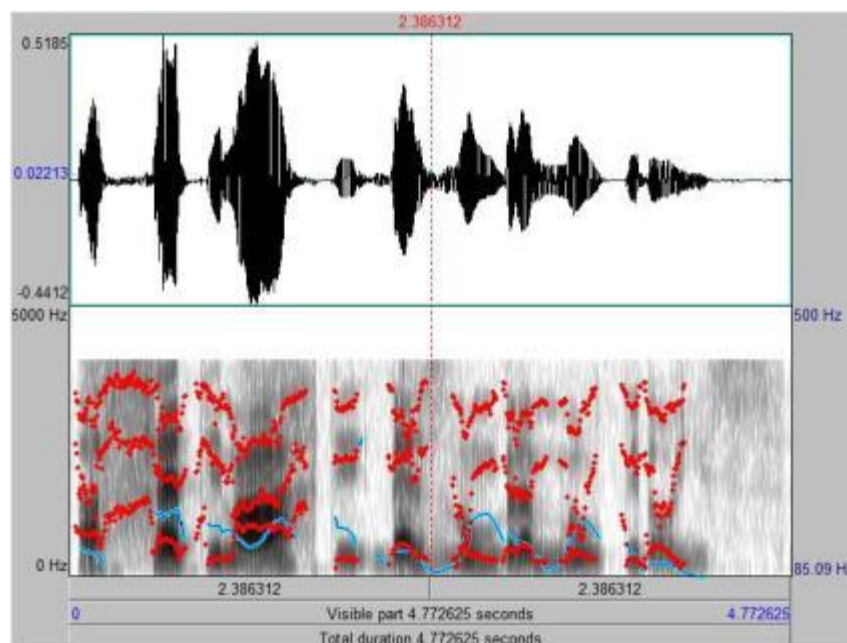


Figura 22 - Exemplo do espectrograma da região de fala extraída a partir do sinal de gravação da figura 20.

Esta região extraída do sinal de gravação e então salva no formato *wave*, juntamente com o arquivo de gravação no diretório de cada indivíduo. A indexação do diretório de cada indivíduo corresponde a uma letra correspondendo ao sexo, “H” para homens e “M” para mulheres, um número correspondendo à ordem de gravação seguido da idade do indivíduo. Por exemplo, o diretório H3-14 pertence a um homem com 14 anos, o terceiro homem gravado. Todos os arquivos de gravação possuem o nome *fi* e todos os arquivos originais possuem o nome *frasei*, onde $1 \leq i \leq 12$ corresponde ao número da frase.

4.2 – Codificação/Decodificação do sinal original

De posse dos arquivos originais, iniciou-se o processo de codificação. Todos os codificadores analisados operam diretamente sobre fluxo de bits ou *bitstream*. Logo, foi necessário o uso de programa gratuito de conversão de áudio SoX [51]. Através do SoX foi possível extrair o *bitstream* de cada arquivo original, permitindo o processo de codificação. Dentro do diretório de cada indivíduo estão disponibilizados os arquivos de *bitstream* (.raw) dos arquivos originais utilizados no processo de codificação. A única exceção foi o codificador Speex que permitia a manipulação direta dos arquivos originais no formato *wave*.

Os codificadores G.722, G.723.1, G.726 e G.728 foram obtidos a partir do site da empresa Imtelephone[®] [52], que disponibiliza versões de teste destes codificadores. O codificador G.729A foi obtido a partir do site da empresa VoiceAge[®] [53], correspondendo a uma versão código aberto ponto flutuante do codificador ITU-T G.729A. O codificador iLBC foi obtido a partir do site do projeto [54]. O codificador Speex foi obtido a partir do site de seus desenvolvedores [55].

Todos os programas de codificação e conversão de áudio operam no *prompt* de comando na plataforma Windows XP[®]. Todos os comandos utilizados encontram-se no DVD em anexo.

O processo de codificação foi executado apenas uma vez para cada indivíduo. Para tal foi utilizado um *notebook Pentium Core2Duo* 1,86 GHz com 3GB de memória RAM, o mesmo utilizado no processo de gravação das amostras de voz. Primeiro os arquivos de *bitstream* originais eram codificados. Em seguida os arquivos codificados

resultantes eram decodificados, gerando arquivos de *bitstream* finais. Havia apenas instâncias do *Windows Explorer* e dos *prompts* de comando utilizados durante este processo. Em seguida, os arquivos de *bitstream* decodificados eram convertidos para o formato *wave* mono estéreo PCM uniforme 8 kHz com 16 bits por amostra utilizando-se o SoX. Durante este processo, os tempos de codificação e decodificação fornecidos pelos programas de codificação foram coletados. Apenas para o codificador Speex não foi necessário a utilização do SoX, visto que ele manipulava diretamente os arquivos no formato *wave*.

Os arquivos intermediários não foram salvos, visto que o processo de conversão de *bitstream* para *wave* ocorre sem perdas. No diretório de cada indivíduo, estão os arquivos decodificados no formato *wave* separados por subdiretórios com o nome de cada codificador.

4.3 – Coleta de dados

Antes da coleta de dados, criou-se a grade com os intervalos dos fonemas a serem analisados. A separação destes intervalos foi efetuada no Praat através da observação dos padrões no espectrograma e percepção auditiva. Na figura 23 temos a grade do espectrograma da figura 22, com a criação de doze intervalos contendo os grupos de fonemas da frase 1 da tabela 11¹⁶.

A criação da grade mostrou-se um desafio para alguns indivíduos, visto que está intimamente relacionada com a cadência da fala. Principalmente os hiatos, como nas palavras “*Viagem*” e “*Preocupados*”, mostraram-se difíceis de obter uma separação exata. Por este motivo, apenas o hiato da palavra “*Ainda*” foi separado.

Outro problema surgiu com a pronúncia de alguns fonemas, como o /te/ da palavra “*Arte*”. Alguns indivíduos emitiram um som completamente consonantal, ou seja, sem vibração das cordas vocais, apenas perceptível durante o processo de criação da grade. Notou-se também a presença dos alofonemas /txe/ e /txi/ nas palavras “*Arte*” e “*Políticos*”, comuns no Rio de Janeiro.

¹⁶ Pode-se observar as grades criadas para os arquivos originais e observar a dificuldade de separação dos hiatos e de outros grupos fonéticos.

coletados seus valores mínimo, médio e máximo, enquanto para as formantes apenas coletou-se o seu valor médio. Os intervalos de tempo entre palavras adjacentes foram divididos entre os grupos de fonemas iniciais e finais das palavras adjacentes.

4.4 - Resultados

Os resultados são apresentados por grupo na forma de tabelas e gráficos, representando o desvio médio relativo à média da contribuição de cada grupo fonético estudado de cada codificador. São apresentados os resultados de apenas um indivíduo por grupo para as frases 2, 4 e 10, devido à quantidade de informação a ser apresentada. A seleção destas frases se deve ao fato de que nas duas primeiras ocorre um *fenômeno de correção do pitch*, ou seja, o indivíduo falou um intervalo fonético sem a detecção de vibração das cordas vocais e alguns codificadores inseriram tal informação durante o processamento do sinal. Tais intervalos contêm um som puramente consonantal ou um som vocálico de baixa intensidade e são comuns nessas frases. A frase 10 possui geralmente os maiores erros de processamento para cada um dos grupos, provavelmente por ser a frase que contém a maior quantidade de sons consonantais que representam ruídos similares a estática, tipicamente o fonema /j/.

Quanto à etapa de processamento de sinal, é apresentada a *taxa de tempo de processamento*, que consiste na divisão do tempo de processamento do sinal (soma dos tempos de codificação e decodificação) pelo tempo de duração do arquivo de áudio. Desta forma é possível analisar o tempo de processamento independentemente dos tempos de pronúncia dos indivíduos. Dado que a resolução dos tempos de codificação e decodificação é de 0,1 s, esses valores representam apenas uma estimativa da taxa real de processamento que não é o foco deste trabalho. Como os codificadores G.722 e G.726 sempre apresentaram taxas nulas, seus resultados não são mostrados nas tabelas com essa informação.

Os erros de processamento são apresentados na forma de desvios percentuais com relação ao arquivo original, ou seja, PCM uniforme 8 kHz com 16 bits por amostra. Valores negativos indicam uma atenuação no valor das frequências, enquanto valores positivos indicam um ganho. Os resultados do erro de processamento por grupo representam uma média dos resultados individuais para cada codificador. Os valores

individuais são acompanhados de variações nos valores das frequências por intervalo analisado. Os intervalos contendo descontinuidades nas frequências fundamentais foram removidos do cálculo do desvio médio, mas podem ser facilmente observados nos gráficos de variação absoluta. No caso do fenômeno de correção de *pitch*, tais intervalos representam perda de fidelidade, enquanto nos casos em que os codificadores perdem a informação de *pitch*, representam um erro grave de processamento.

Quanto à avaliação perceptual, a inserção de ruídos no início de cada arquivo decodificado foi observada no codificador G.722. Alguns erros de processamento ocorreram para o codificador G.723.1, onde há presença de forte ruído no início do arquivo decodificado e uma forte atenuação na intensidade sonora.

Os grupos de indivíduos são indexados por uma letra maiúscula contendo o sexo (H para homens e M para mulheres) e a designação do grupo contida na tabela 41. Por exemplo, o grupo HG1 representa o grupo 1 com homens de 14 a 34 anos.

Para uma melhor visualização dos resultados, os codificadores foram separados em dois grupos: o primeiro contém os codificadores G.722, G.723.1 e G.726, enquanto o segundo contém os codificadores G.728, G.729A, iLBC e Speex.

4.4.1 – HG1

Fazem parte do grupo os indivíduos H3-14 e H2-29. O indivíduo H3-14 foi escolhido para ter seus resultados individuais apresentados por estar na faixa de idade de troca de voz.

Quanto à avaliação perceptual, pode-se perceber certo nível de saturação nos arquivos do indivíduo H2-29. Em contraste, os arquivos do indivíduo H3-14 apresentam um baixo volume, facilmente corrigido com escalonamento de intensidade.

Para o codificador G.723.1, o indivíduo H3-14 teve erros de processamento com a inserção de ruído e forte atenuação em todos os arquivos exceto os das frases 2 e 9. Estes erros aparecem nos gráficos como intervalos cujas frequências fundamentais são nulas.

Na tabela 12 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo H3-14 e o grupo HG1. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento.

Tabela 12 - Duração dos intervalos médios e dos arquivos do grupo HG1 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento			
		Intervalo Médio (s)	Duração (s)	G.723.1	G.728	G.729A	iLBC
H3-14	Frase 1	0,237	2,848	21,068%	10,534%	3,511%	7,023%
	Frase 2	0,185	2,775	21,620%	10,810%	3,603%	7,207%
	Frase 3	0,227	2,045	24,447%	9,779%	0,000%	4,889%
	Frase 4	0,191	2,484	20,129%	8,051%	4,026%	4,026%
	Frase 5	0,238	2,138	18,711%	9,355%	0,000%	4,678%
	Frase 6	0,190	1,518	19,767%	13,178%	0,000%	6,589%
	Frase 7	0,180	1,977	20,237%	10,119%	0,000%	5,059%
	Frase 8	0,169	2,030	24,632%	9,853%	0,000%	4,926%
	Frase 9	0,168	2,021	24,742%	9,897%	4,948%	9,897%
	Frase 10	0,205	2,055	19,469%	9,734%	4,867%	4,867%
	Frase 11	0,182	1,816	27,536%	11,014%	0,000%	5,507%
	Frase 12	0,193	2,509	23,915%	11,957%	3,986%	3,986%
HG1	Frase 1	0,227	2,718	22,124%	9,130%	3,687%	7,375%
	Frase 2	0,208	3,124	20,887%	9,724%	3,241%	5,043%
	Frase 3	0,242	2,182	23,003%	11,357%	2,156%	4,601%
	Frase 4	0,210	2,727	20,167%	9,077%	3,697%	5,380%
	Frase 5	0,270	2,429	20,385%	10,193%	1,838%	6,015%
	Frase 6	0,226	1,809	19,401%	11,348%	0,000%	5,674%
	Frase 7	0,220	2,425	20,557%	10,279%	1,740%	4,269%
	Frase 8	0,209	2,502	22,400%	9,968%	1,681%	4,144%
	Frase 9	0,203	2,440	22,861%	10,194%	4,223%	6,697%
	Frase 10	0,249	2,489	19,993%	9,996%	4,143%	4,143%
	Frase 11	0,199	1,990	23,006%	10,126%	2,309%	5,063%
	Frase 12	0,214	2,784	21,763%	10,882%	3,627%	3,627%

Nas tabelas 13 e 14 temos os erros de processamento das frequências fundamentais mínimas, médias e máximas para cada frase para o indivíduo H3-14 e médios do grupo HG1. Pode-se observar que o codificador G.723.1 possui os maiores erros de processamento, seguido pelo G.722.

Nas tabelas 15 e 16 temos os erros de processamento das formantes para cada frase para o indivíduo H3-14 e médios do grupo HG1. Pode-se observar que os codificadores G.723.1 e G.729A possuem os maiores erros.

Nas figuras de 24 a 41 temos os resultados do indivíduo H3-14 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas em Hertz das frequências por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

Tabela 13 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H3-14 e o grupo HG1.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
H3-14	Frase 1	6,10%	-3,50%	-9,52%	80,40%	49,47%	34,91%	2,12%	0,34%	-0,12%
	Frase 2	5,89%	2,05%	-0,38%	0,38%	0,28%	0,04%	0,14%	0,06%	0,16%
	Frase 3	3,55%	0,44%	-1,71%	6,14%	18,55%	31,54%	1,77%	0,08%	-0,44%
	Frase 4	4,99%	1,73%	-0,68%	7,43%	1,87%	-1,62%	-0,19%	0,00%	0,36%
	Frase 5	13,86%	3,85%	-1,04%	11,90%	2,88%	-0,54%	0,46%	-0,02%	-0,25%
	Frase 6	0,04%	-0,09%	0,01%	39,98%	35,88%	36,36%	-0,12%	-0,07%	0,48%
	Frase 7	4,55%	-2,27%	-7,36%	9,14%	1,45%	-1,19%	-0,18%	-3,08%	-6,46%
	Frase 8	2,65%	0,59%	-0,64%	7,52%	1,94%	-1,46%	0,16%	0,12%	0,04%
	Frase 9	5,71%	1,43%	-0,36%	-0,16%	-0,20%	-0,05%	-0,08%	0,06%	0,00%
	Frase 10	3,27%	0,83%	-1,94%	5,52%	1,26%	-0,80%	0,33%	0,29%	0,11%
	Frase 11	2,17%	-5,95%	-17,52%	104,31%	89,35%	71,36%	0,36%	1,80%	-6,71%
	Frase 12	4,38%	1,49%	-0,52%	6,04%	0,81%	-0,91%	-0,26%	0,32%	0,27%
HG1	Frase 1	4,77%	0,66%	-0,99%	40,28%	25,00%	17,09%	1,31%	0,16%	0,11%
	Frase 2	2,82%	1,33%	0,80%	-0,23%	0,11%	-0,03%	0,31%	0,11%	0,01%
	Frase 3	1,84%	0,16%	-0,80%	-0,49%	8,53%	14,87%	0,99%	0,03%	-0,08%
	Frase 4	2,58%	0,74%	-0,43%	3,99%	0,85%	-0,22%	-0,25%	-0,38%	-0,15%
	Frase 5	6,94%	1,84%	-1,26%	0,98%	0,64%	-0,94%	0,25%	-0,11%	-0,94%
	Frase 6	0,07%	-0,13%	-0,36%	17,71%	17,61%	18,33%	-0,05%	-0,67%	-0,86%
	Frase 7	2,62%	-1,07%	-3,89%	4,74%	0,76%	-0,55%	0,03%	-1,43%	-3,31%
	Frase 8	1,73%	0,29%	-0,62%	3,71%	0,97%	-0,77%	0,56%	0,19%	0,02%
	Frase 9	3,74%	0,00%	-0,17%	-0,16%	0,85%	5,28%	-0,04%	-0,08%	-0,37%
	Frase 10	2,36%	0,49%	-1,45%	3,67%	0,98%	4,44%	0,64%	0,19%	0,05%
	Frase 11	0,86%	-3,03%	-8,60%	51,62%	44,68%	36,16%	0,16%	0,85%	-3,75%
	Frase 12	2,16%	0,78%	-0,48%	1,61%	0,18%	-0,23%	-0,19%	0,13%	0,09%

Tabela 14 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
H3-14	Frase 1	2,17%	-3,20%	-5,95%	0,32%	-3,37%	-4,28%	0,10%	-3,25%	-5,01%	-1,45%	-3,55%	-4,18%
	Frase 2	0,14%	0,02%	0,03%	-0,35%	-0,09%	0,26%	0,32%	0,12%	0,22%	0,16%	-0,09%	-0,04%
	Frase 3	0,71%	-0,02%	-0,31%	-0,21%	0,01%	0,59%	0,42%	0,04%	0,30%	-1,59%	-0,19%	0,20%
	Frase 4	0,88%	0,12%	-0,33%	-0,61%	0,05%	0,08%	-3,38%	-0,30%	0,13%	0,52%	0,15%	0,01%
	Frase 5	0,55%	0,11%	-0,01%	0,82%	0,19%	1,39%	0,97%	0,21%	0,74%	1,44%	0,54%	1,92%
	Frase 6	-0,43%	-0,08%	0,48%	-1,47%	-0,14%	0,55%	-0,83%	-0,16%	1,05%	-0,58%	-0,02%	0,81%
	Frase 7	0,00%	-3,18%	-6,29%	0,07%	-2,99%	-6,40%	0,44%	-3,21%	-6,55%	0,21%	-3,15%	-6,69%
	Frase 8	-0,18%	0,00%	0,00%	-0,74%	0,04%	0,56%	0,40%	0,09%	0,40%	0,01%	-0,11%	0,37%
	Frase 9	-0,06%	0,08%	0,04%	-0,95%	-0,11%	0,13%	-1,65%	-0,06%	2,25%	-0,49%	-0,37%	0,00%
	Frase 10	-0,36%	-0,01%	-0,16%	-0,22%	0,14%	0,32%	0,07%	0,28%	0,95%	0,46%	0,35%	0,64%
	Frase 11	-0,12%	1,56%	0,19%	-3,15%	-7,64%	-14,62%	-3,33%	-5,49%	-7,93%	-1,23%	-6,13%	-14,15%
	Frase 12	0,02%	0,29%	0,19%	-0,31%	0,44%	0,27%	-0,79%	0,30%	0,36%	0,02%	0,31%	-0,06%
HG1	Frase 1	1,29%	-1,96%	-2,85%	0,52%	-1,33%	-2,36%	-0,85%	-1,07%	-2,02%	-0,45%	-1,47%	-2,03%
	Frase 2	0,30%	0,27%	1,37%	7,73%	6,22%	3,65%	-1,59%	0,36%	2,75%	-1,08%	0,08%	1,97%
	Frase 3	0,42%	-0,04%	0,01%	16,43%	13,39%	9,88%	-5,93%	-0,54%	1,02%	-0,70%	-0,08%	0,05%
	Frase 4	0,38%	-0,06%	-0,30%	20,49%	17,37%	14,67%	-2,18%	-0,07%	1,64%	0,62%	0,18%	0,51%
	Frase 5	0,25%	-0,04%	-0,88%	23,90%	23,08%	21,51%	-3,26%	-0,36%	0,32%	1,02%	0,37%	0,87%
	Frase 6	-0,11%	-0,03%	0,24%	16,09%	16,02%	15,27%	-1,84%	-0,29%	0,42%	-1,20%	-0,07%	0,41%
	Frase 7	0,03%	-1,54%	-3,10%	23,24%	16,28%	10,30%	0,66%	-2,03%	-4,71%	0,26%	-2,06%	-4,73%
	Frase 8	0,73%	0,00%	-0,07%	16,63%	13,36%	10,99%	0,53%	0,15%	0,12%	0,73%	0,02%	0,18%
	Frase 9	-0,03%	0,02%	-0,12%	22,95%	22,48%	20,60%	-1,25%	1,40%	7,07%	-1,22%	-0,43%	0,25%
	Frase 10	0,37%	0,09%	-0,10%	21,13%	15,86%	12,80%	1,34%	0,39%	0,85%	0,66%	0,32%	0,38%
	Frase 11	-0,36%	0,59%	-0,03%	9,53%	3,42%	-7,02%	-2,07%	-1,99%	0,69%	-1,28%	-3,98%	-7,71%
	Frase 12	-0,10%	0,10%	-0,03%	19,33%	16,94%	14,27%	-2,03%	-0,20%	0,19%	-0,65%	-0,26%	-0,66%

Tabela 15 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H3-14 e o grupo HG1.

		G.722			G.723.1			G.726		
		Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
H3-14	Frase 1	-3,17%	-2,91%	1,65%	-10,77%	-2,10%	-1,33%	-4,84%	-1,37%	0,99%
	Frase 2	1,94%	-0,66%	-5,27%	-6,45%	-5,97%	-0,93%	0,90%	0,85%	0,52%
	Frase 3	-1,04%	-3,67%	0,72%	-8,31%	-5,33%	-4,34%	0,31%	0,48%	0,95%
	Frase 4	1,95%	-2,63%	1,08%	-3,48%	-3,00%	-2,88%	1,82%	1,31%	1,30%
	Frase 5	0,30%	-1,39%	1,49%	-2,60%	-1,08%	-2,95%	1,41%	0,38%	0,39%
	Frase 6	-2,64%	-3,84%	0,58%	-1,69%	-3,46%	-4,11%	0,38%	0,15%	0,49%
	Frase 7	-16,49%	-6,80%	2,13%	-14,15%	-5,55%	-3,42%	-5,50%	-0,79%	1,55%
	Frase 8	-2,36%	-6,60%	0,45%	-0,19%	-1,00%	-4,45%	0,08%	0,65%	1,02%
	Frase 9	-10,70%	-5,09%	0,64%	-15,44%	-8,37%	-4,47%	1,17%	-0,48%	0,58%
	Frase 10	-1,19%	-1,81%	1,11%	3,41%	-3,17%	-2,92%	-3,65%	-0,05%	0,89%
	Frase 11	-3,65%	-3,27%	1,88%	5,49%	-0,91%	-1,85%	1,56%	0,12%	0,67%
	Frase 12	-3,37%	-3,98%	1,52%	-9,41%	-2,26%	-3,47%	1,77%	1,28%	2,10%
HG1	Frase 1	1,00%	-2,48%	1,09%	-3,85%	-2,34%	-0,60%	-1,46%	0,07%	1,52%
	Frase 2	1,72%	-0,70%	-2,37%	4,40%	1,59%	0,45%	3,03%	2,13%	0,82%
	Frase 3	-1,05%	-2,96%	0,42%	-0,85%	-2,60%	-3,01%	1,79%	0,85%	1,02%
	Frase 4	0,56%	-3,01%	0,85%	-11,28%	-5,07%	-2,36%	1,96%	1,31%	0,99%
	Frase 5	-1,12%	-2,27%	1,06%	8,73%	0,36%	-1,70%	1,57%	0,40%	0,63%
	Frase 6	-1,76%	-3,98%	0,33%	-3,00%	-3,39%	-2,23%	1,50%	1,00%	0,70%
	Frase 7	-10,29%	-5,96%	1,38%	-10,68%	-6,99%	-3,00%	-3,57%	0,69%	1,24%
	Frase 8	0,04%	-5,55%	0,27%	-2,56%	-3,11%	-3,31%	2,69%	1,69%	1,44%
	Frase 9	-6,24%	-5,18%	0,35%	-15,19%	-8,12%	-3,32%	1,93%	-0,04%	0,50%
	Frase 10	-3,20%	-2,68%	0,88%	0,08%	-3,87%	-2,94%	0,10%	1,29%	0,80%
	Frase 11	-2,75%	-3,95%	0,96%	3,13%	-1,26%	-1,45%	1,67%	0,44%	0,78%
	Frase 12	1,13%	-3,98%	0,81%	-5,10%	-4,41%	-3,35%	3,26%	1,28%	1,28%

Tabela 16 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.

		G.728			G.729A			iLBC			Speex		
		Formantes			Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
H3-14	Frase 1	0,41%	0,42%	0,18%	-16,14%	-5,99%	-4,35%	-9,28%	-3,53%	-2,55%	-6,06%	-3,22%	-0,57%
	Frase 2	2,28%	2,05%	-0,29%	-10,76%	-9,19%	-3,35%	-2,07%	-5,10%	-2,66%	-0,29%	-2,53%	-0,40%
	Frase 3	-1,68%	0,33%	-0,65%	-8,93%	-7,86%	-6,50%	-3,45%	-3,99%	-3,22%	-2,68%	-3,13%	-1,38%
	Frase 4	0,43%	0,87%	-0,38%	-10,56%	-8,02%	-4,76%	-5,23%	-5,54%	-3,49%	-3,18%	-4,05%	-0,48%
	Frase 5	-0,25%	0,40%	-0,58%	-8,08%	-5,01%	-4,18%	-2,38%	-5,01%	-3,12%	0,25%	-1,94%	-0,57%
	Frase 6	-3,35%	-2,21%	-1,01%	-19,38%	-7,56%	-6,16%	-4,15%	-3,75%	-2,88%	-3,62%	-3,29%	-1,18%
	Frase 7	-4,21%	-0,15%	0,68%	-17,91%	-7,27%	-5,35%	-12,57%	-5,12%	-3,82%	-5,91%	-2,33%	0,05%
	Frase 8	3,29%	-1,11%	-0,95%	-14,73%	-9,04%	-5,91%	-5,34%	-6,06%	-2,44%	10,11%	-2,48%	-0,37%
	Frase 9	3,46%	-0,39%	-0,75%	-28,74%	-10,85%	-6,45%	-14,40%	-5,92%	-2,96%	-3,81%	-2,45%	-1,18%
	Frase 10	-2,35%	-0,27%	-0,71%	-8,32%	-6,92%	-5,24%	2,83%	-2,68%	-3,54%	-3,20%	-2,27%	-0,78%
	Frase 11	-1,78%	0,76%	0,48%	-19,27%	-6,44%	-3,74%	-12,09%	-6,12%	-2,58%	2,99%	-0,62%	0,44%
	Frase 12	3,01%	1,12%	-0,19%	-11,18%	-7,50%	-5,69%	-3,71%	-4,08%	-4,01%	-1,09%	-0,70%	-0,70%
HG1	Frase 1	-0,73%	0,60%	0,70%	-16,15%	-7,76%	-4,97%	-5,76%	-5,34%	-3,41%	-3,70%	-1,00%	0,97%
	Frase 2	4,91%	4,39%	0,20%	-2,40%	-10,76%	-3,96%	-3,92%	-4,35%	-2,94%	1,32%	0,43%	0,51%
	Frase 3	1,86%	1,71%	-0,33%	2,87%	-9,77%	-9,15%	-6,69%	-5,26%	-4,17%	-3,05%	-1,15%	-0,17%
	Frase 4	-1,63%	0,10%	-0,64%	-5,02%	-10,78%	-9,03%	-9,51%	-8,05%	-3,55%	-5,09%	-2,94%	-0,04%
	Frase 5	-1,32%	0,47%	-0,41%	-5,83%	-7,11%	-6,14%	-2,53%	-4,88%	-3,22%	-2,42%	-1,43%	0,13%
	Frase 6	-1,88%	-0,85%	-0,83%	3,14%	-5,69%	-8,82%	-3,07%	-4,58%	-3,37%	-0,95%	-1,72%	-0,12%
	Frase 7	-0,98%	0,18%	-0,13%	4,98%	-4,04%	-9,45%	-7,87%	-6,86%	-3,97%	-2,67%	-1,54%	0,08%
	Frase 8	1,89%	1,02%	-0,53%	-0,19%	-10,09%	-8,30%	-4,05%	-6,55%	-3,43%	4,14%	-2,03%	0,36%
	Frase 9	0,17%	-0,91%	-1,21%	-17,61%	-9,96%	-10,52%	-16,23%	-7,34%	-3,81%	-5,52%	-2,63%	-0,70%
	Frase 10	-2,04%	0,55%	-0,74%	-2,89%	-7,89%	-9,36%	-4,58%	-4,08%	-3,74%	2,56%	-1,30%	-0,40%
	Frase 11	-1,10%	1,10%	-0,15%	4,90%	-4,64%	-6,17%	-7,19%	-6,00%	-3,10%	1,38%	-0,04%	0,60%
	Frase 12	1,17%	0,24%	-0,90%	2,62%	-6,77%	-10,10%	-4,31%	-5,10%	-3,86%	-1,15%	-1,10%	-0,57%

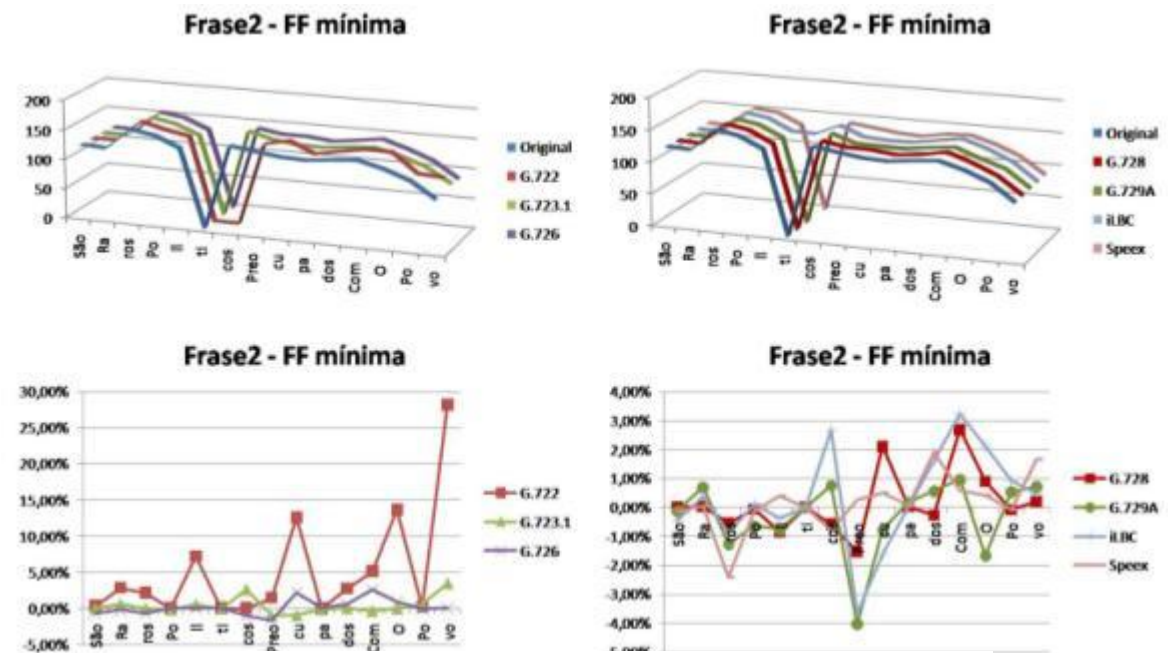


Figura 24 - Variações da frequência fundamental mínima da frase 2 do indivíduo H3-14 do grupo HG1.

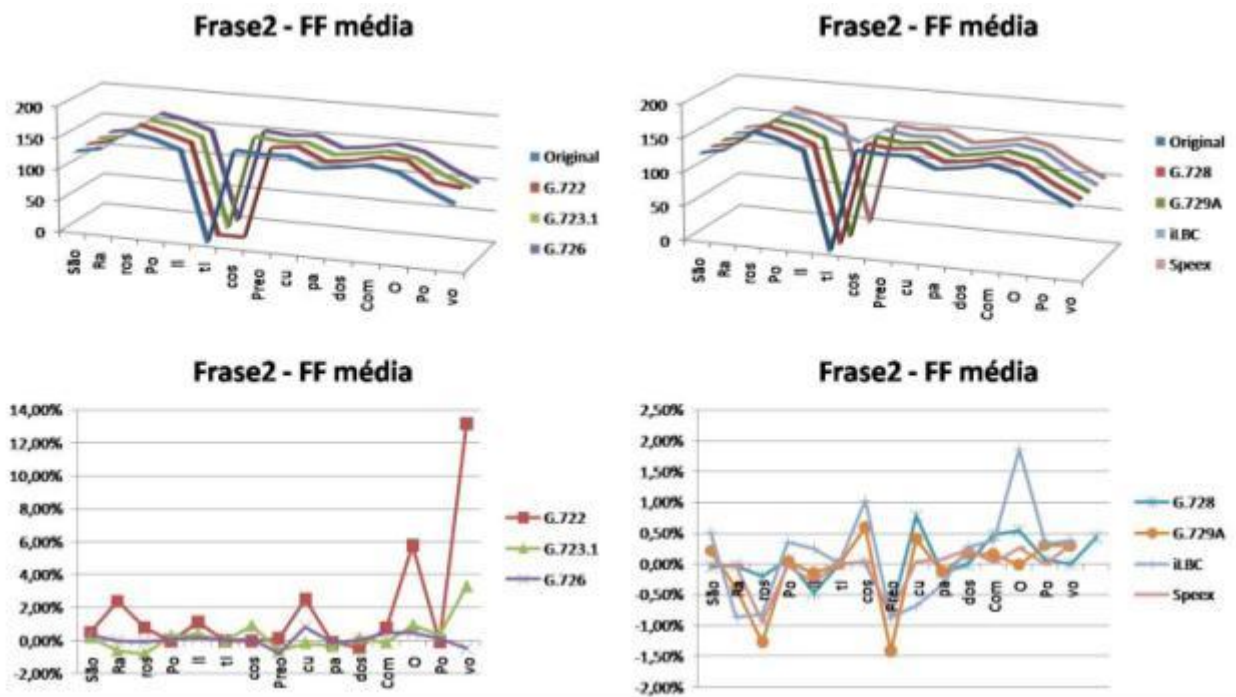


Figura 25 - Variações da frequência fundamental média da frase 2 do indivíduo H3-14 do grupo HG1.

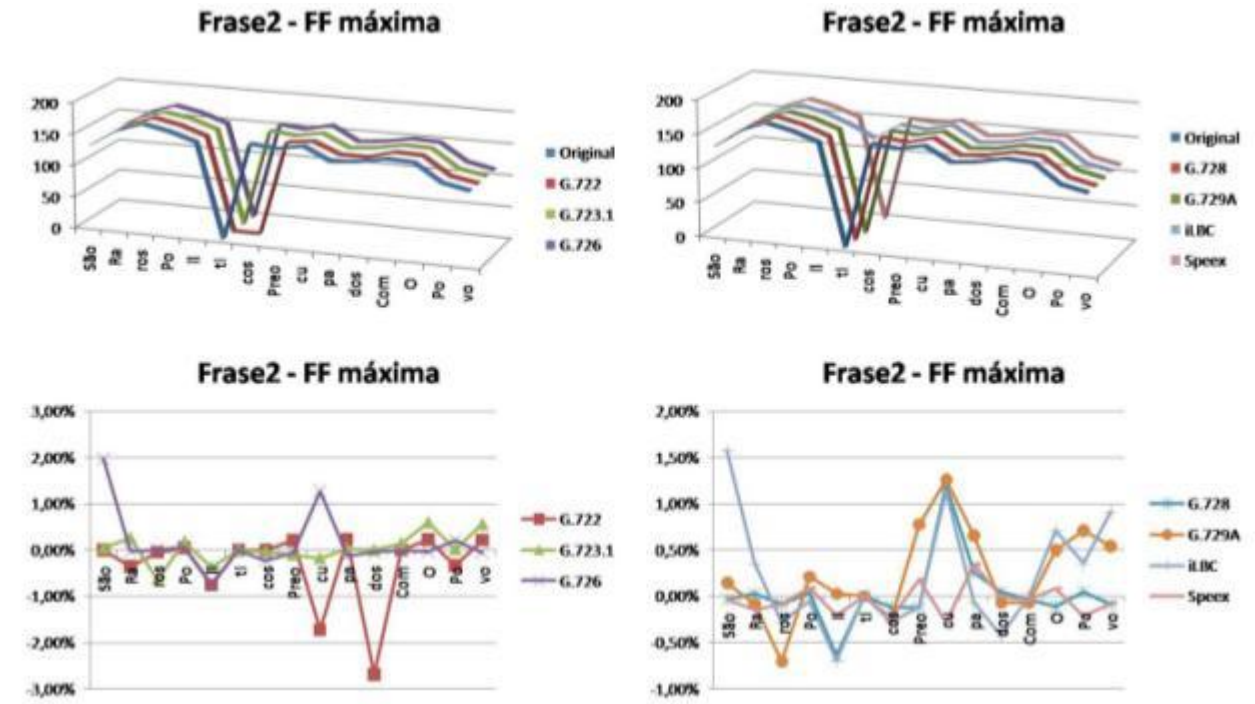


Figura 26 - Variações da frequência fundamental máxima da frase 2 do indivíduo H3-14 do grupo HG1.

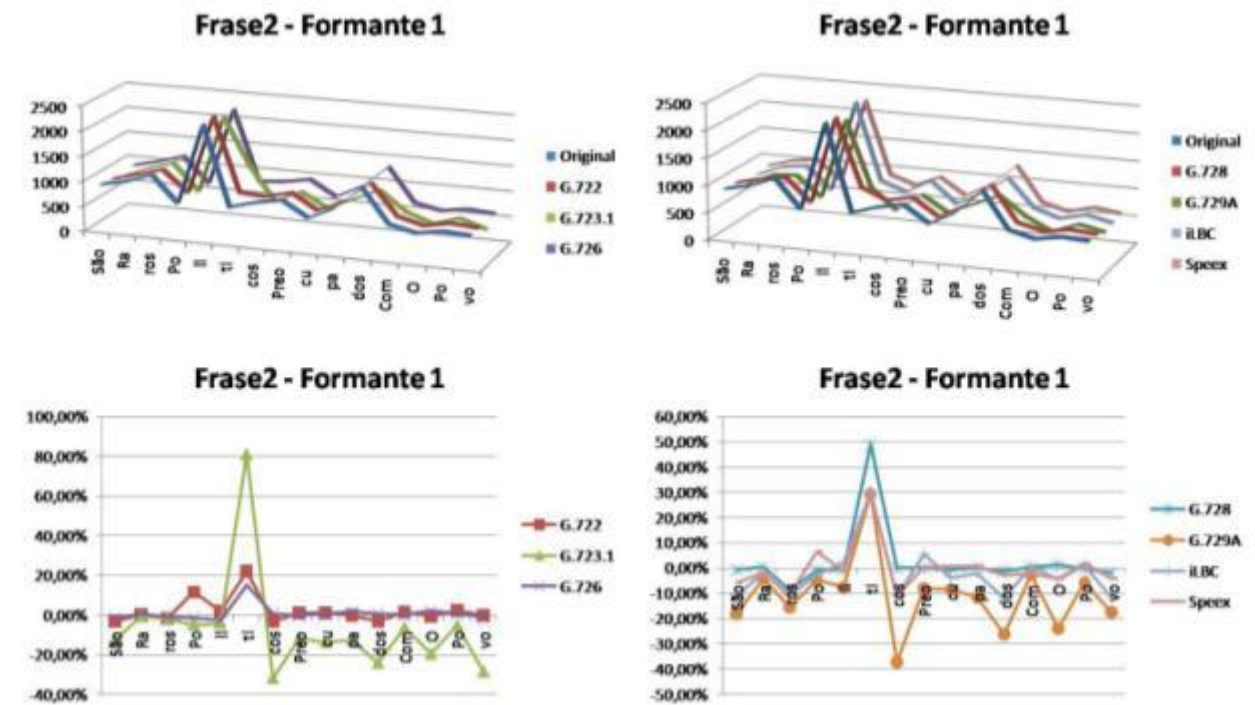


Figura 27 - Variações da frequência da primeira formante da frase 2 do indivíduo H3-14 do grupo HG1.

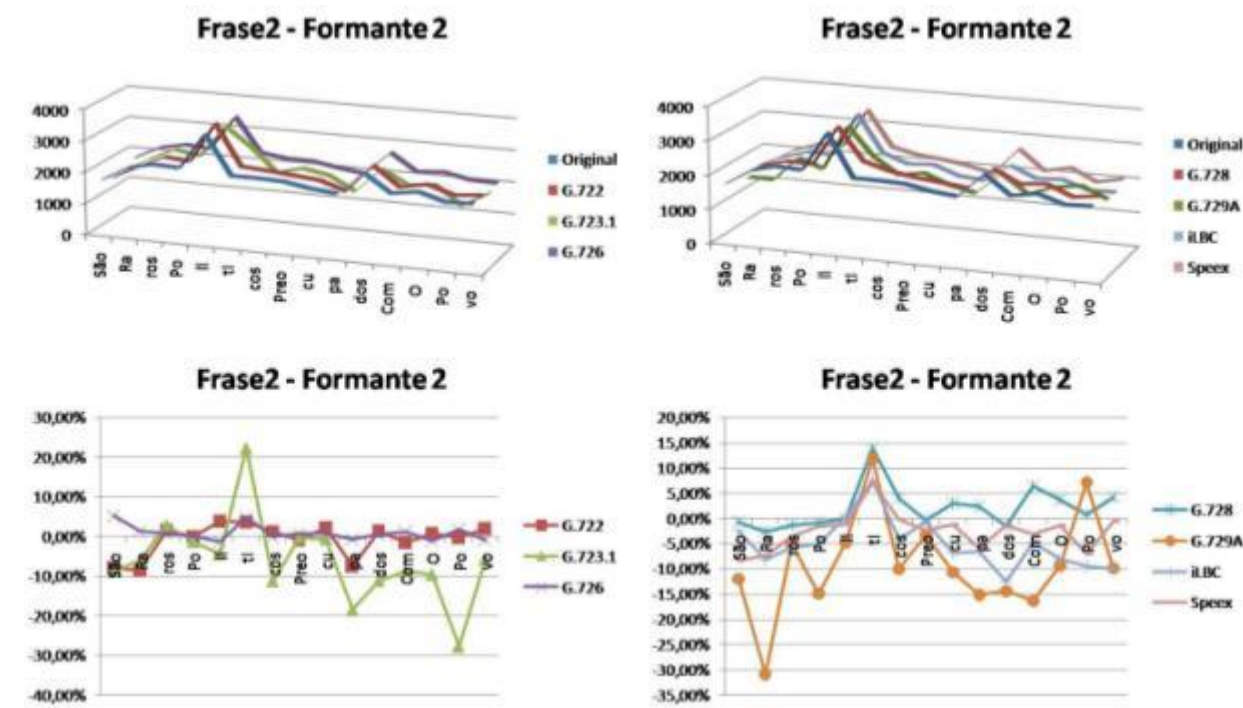


Figura 28 - Variações da frequência da segunda formante da frase 2 do indivíduo H3-14 do grupo HG1.

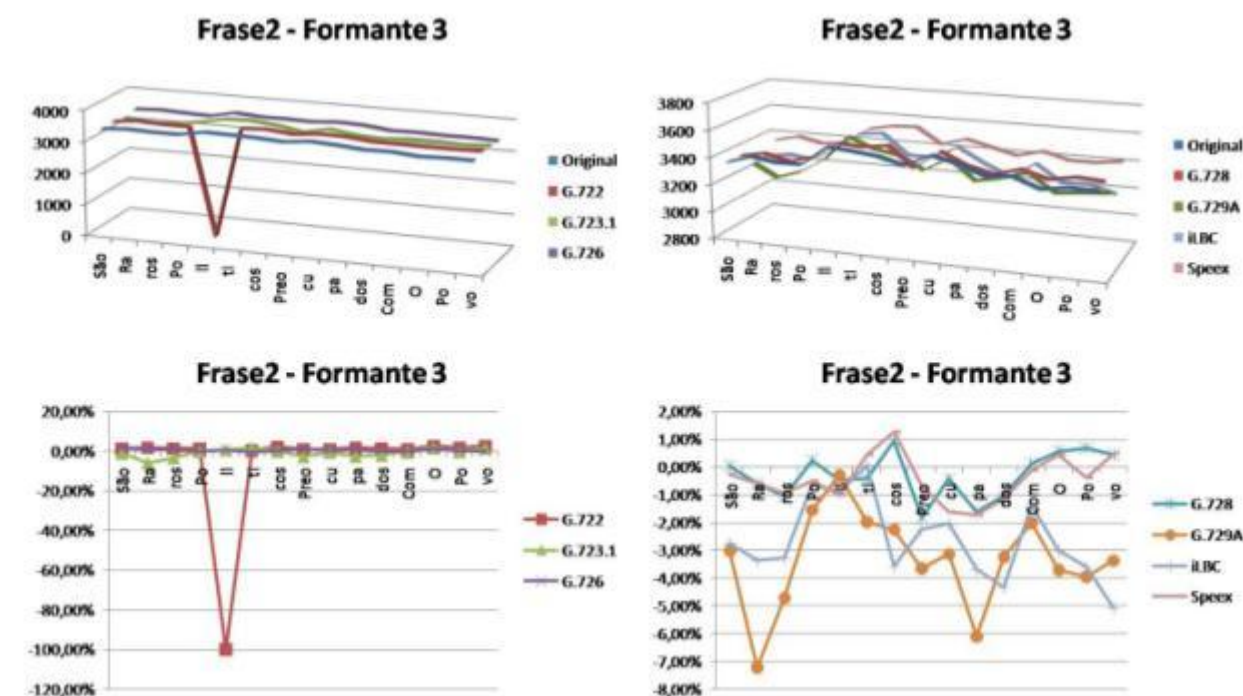


Figura 29 - Variações da frequência da terceira formante da frase 2 do indivíduo H3-14 do grupo HG1.

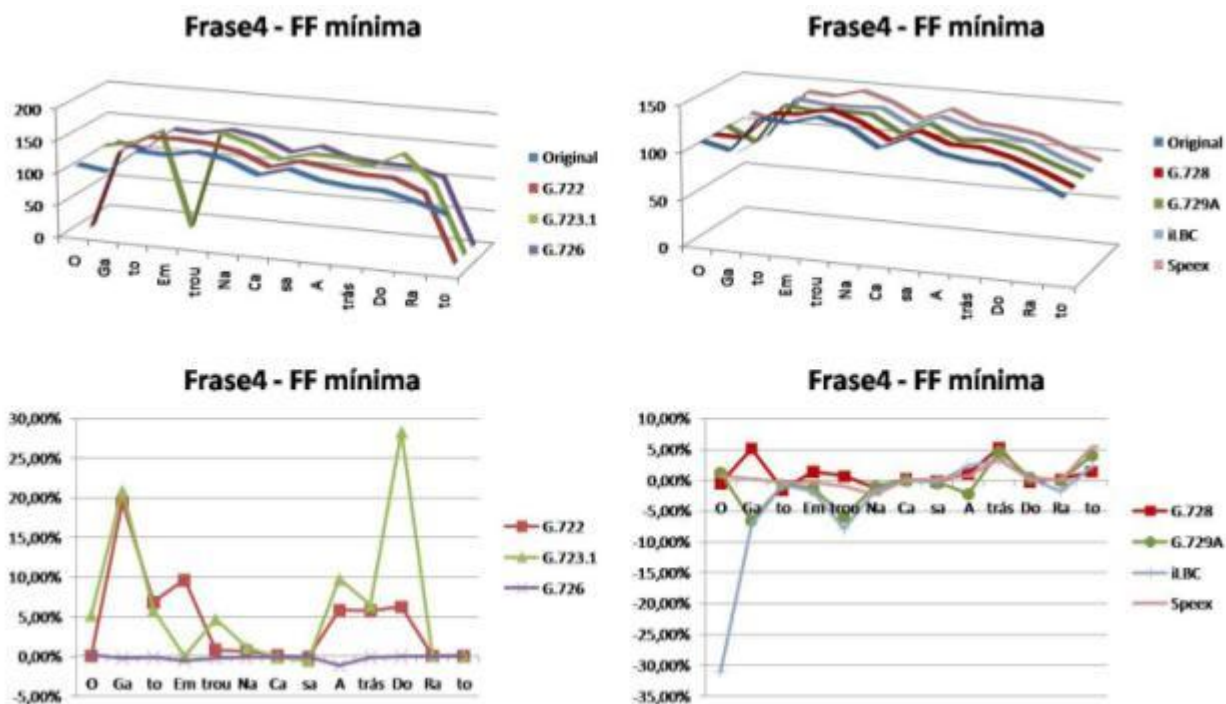


Figura 30 - Variações da frequência fundamental mínima da frase 4 do indivíduo H3-14 do grupo HG1.

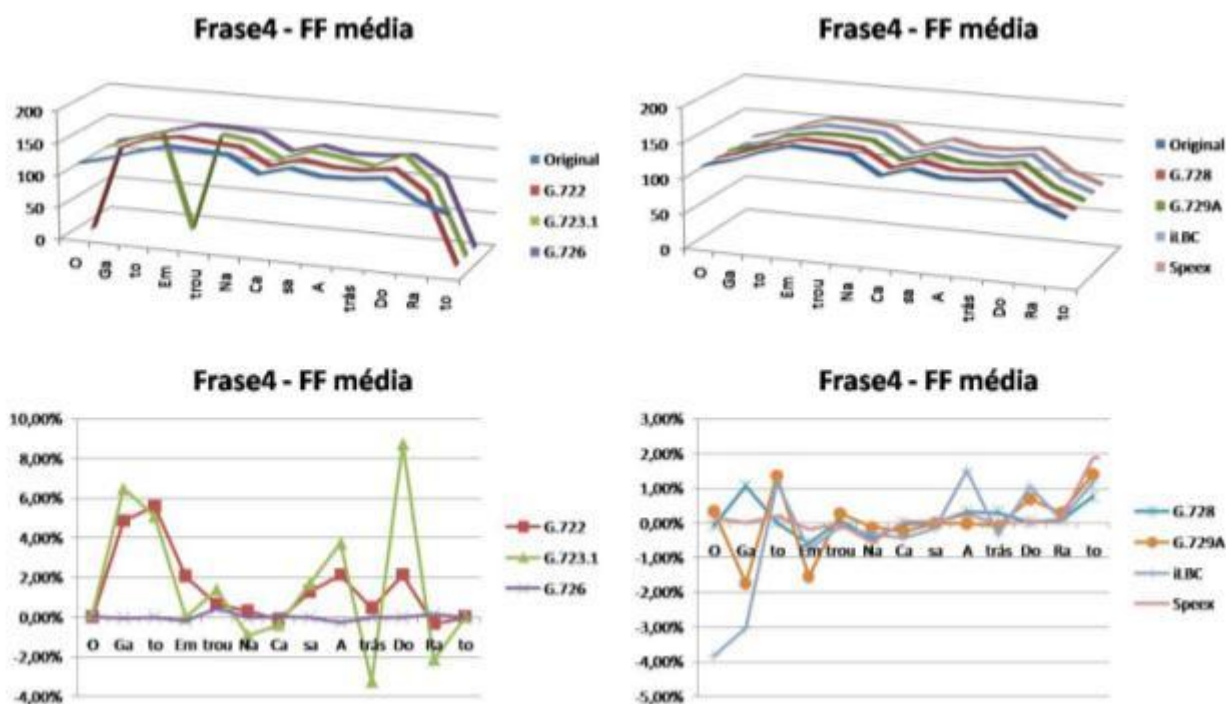


Figura 31 - Variações da frequência fundamental média da frase 4 do indivíduo H3-14 do grupo HG1.

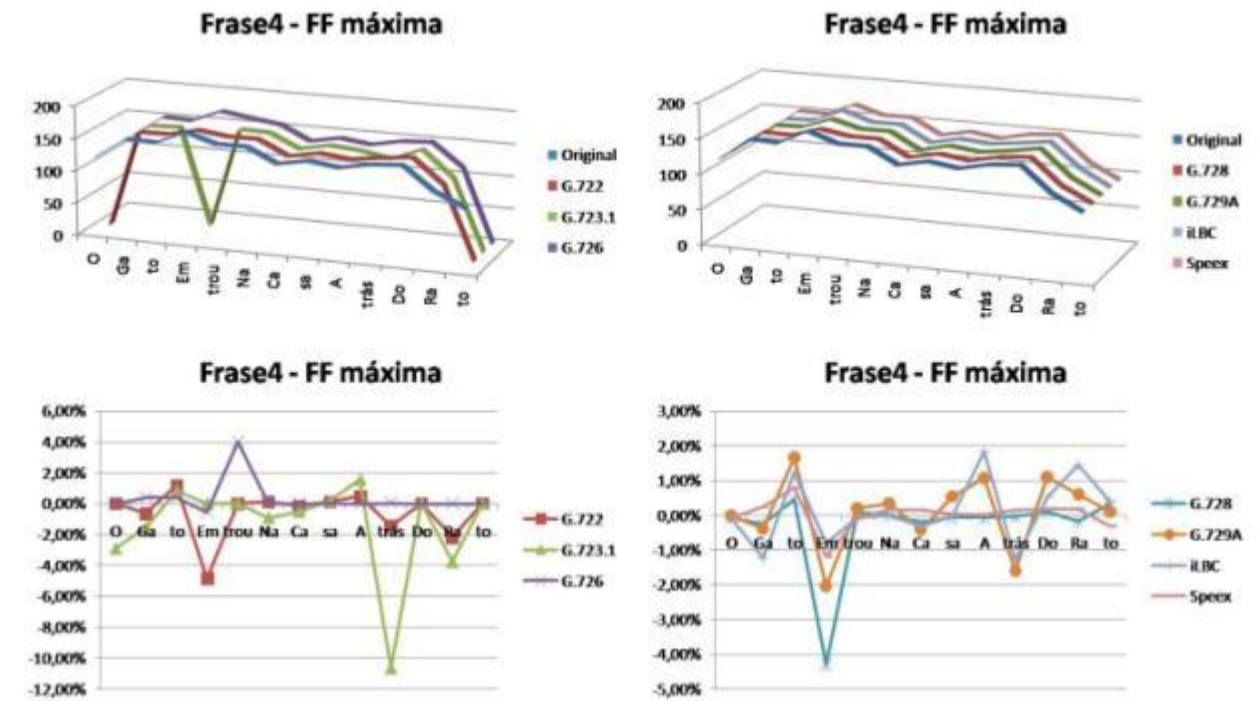


Figura 32 - Variações da frequência fundamental máxima da frase 4 do indivíduo H3-14 do grupo HG1.

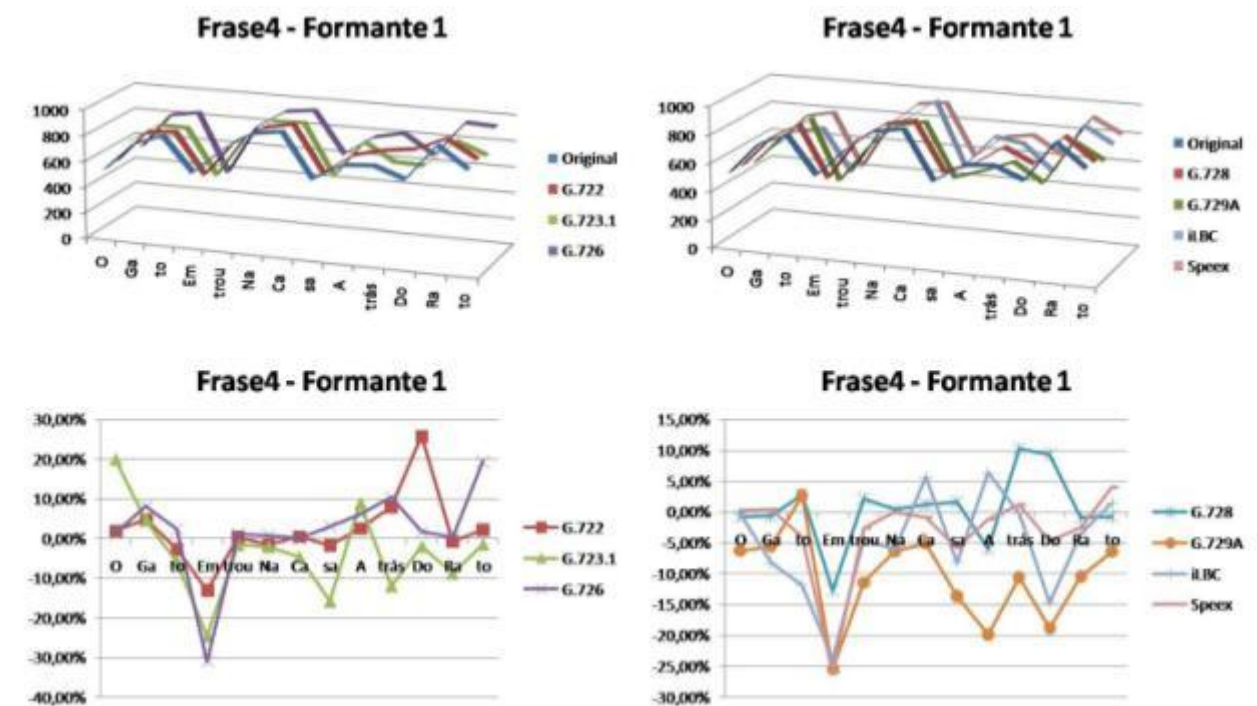


Figura 33 - Variações da frequência da primeira formante da frase 4 do indivíduo H3-14 do grupo HG1.

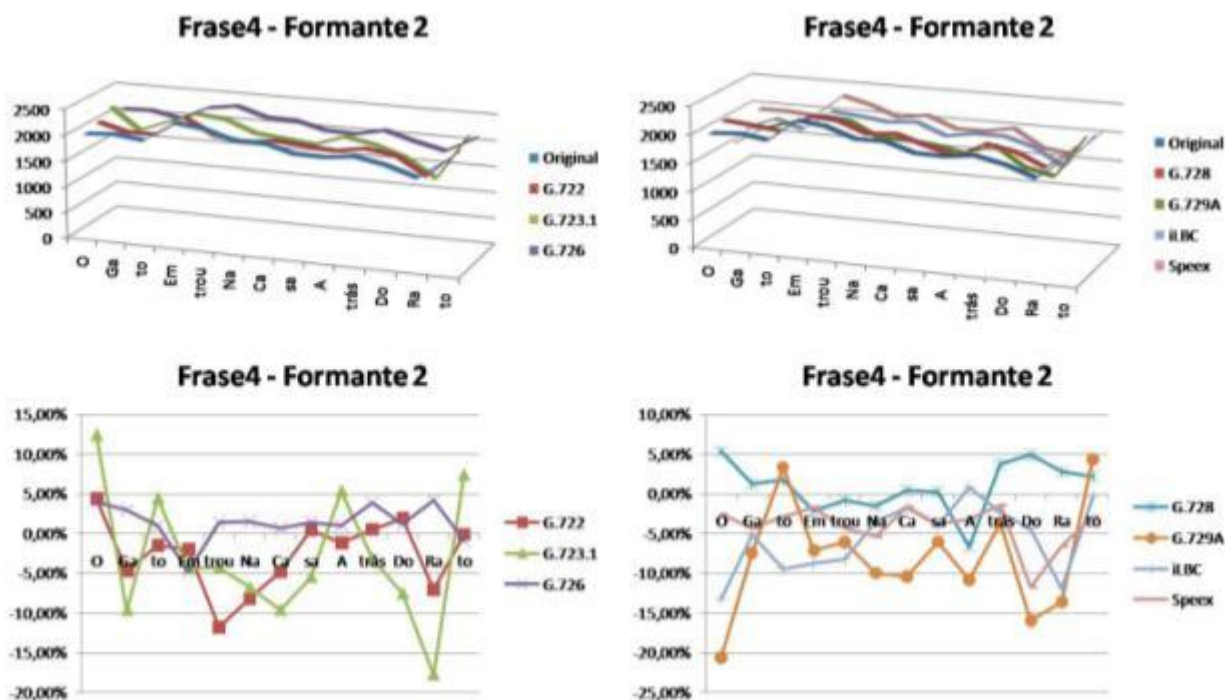


Figura 34 - Variações da frequência da segunda formante da frase 4 do indivíduo H3-14 do grupo HG1.

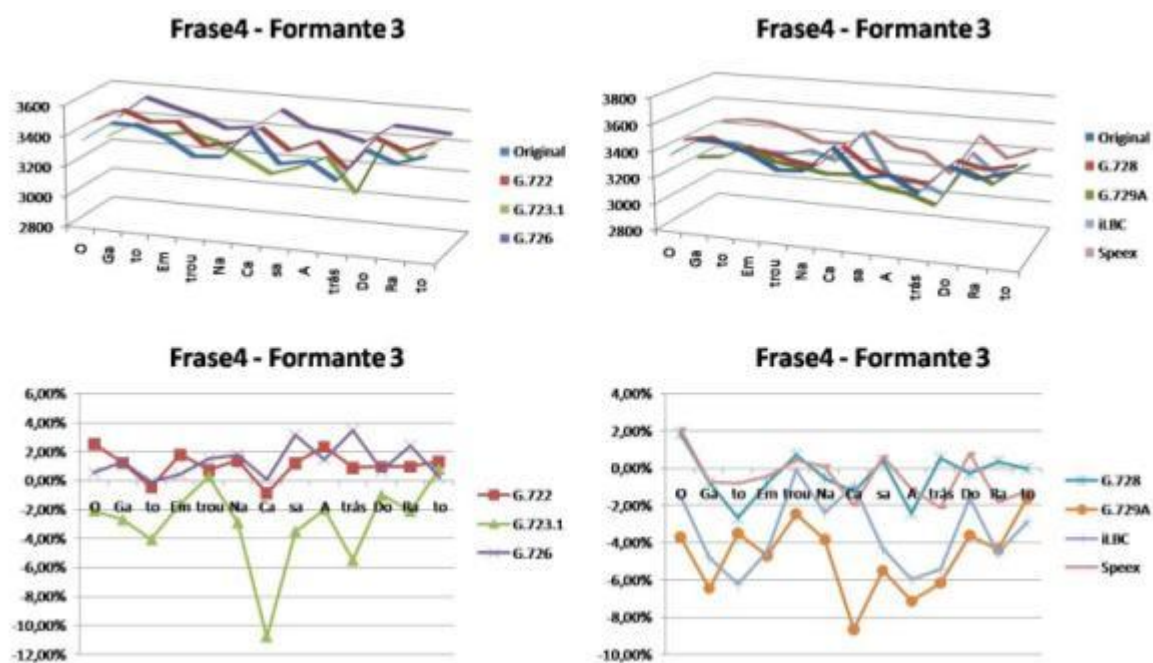


Figura 35 - Variações da frequência da terceira formante da frase 4 do indivíduo H3-14 do grupo HG1.

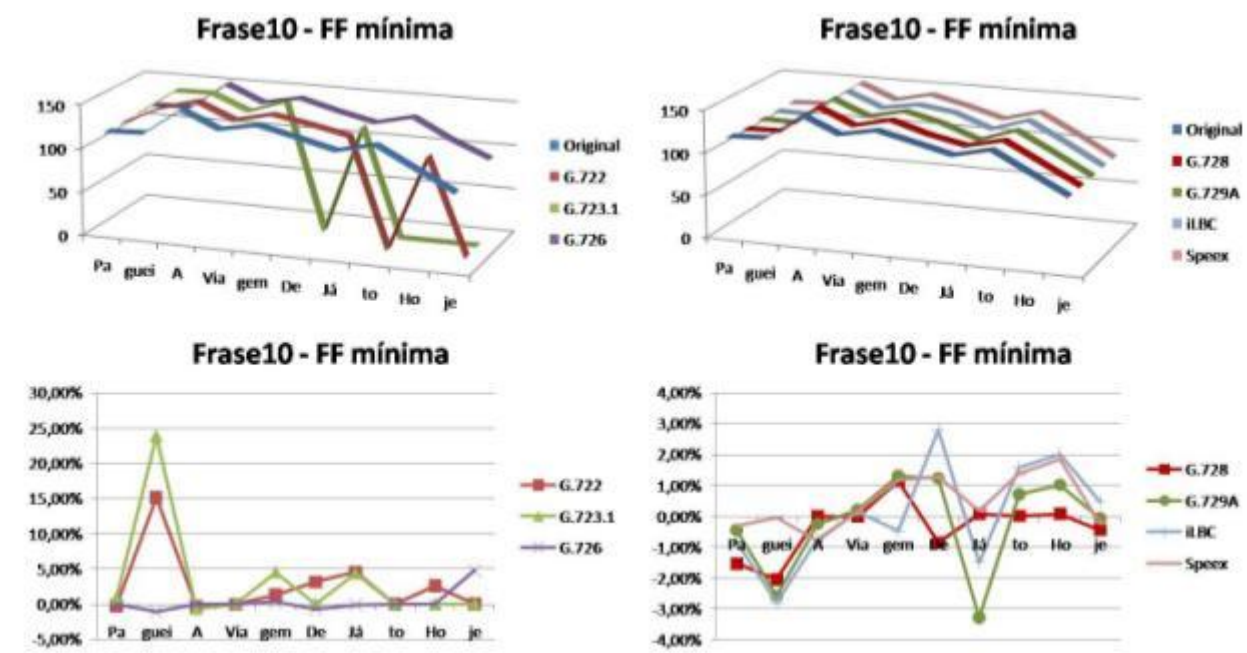


Figura 36 - Variações da freqüência fundamental mínima da frase 10 do indivíduo H3-14 do grupo HG1.

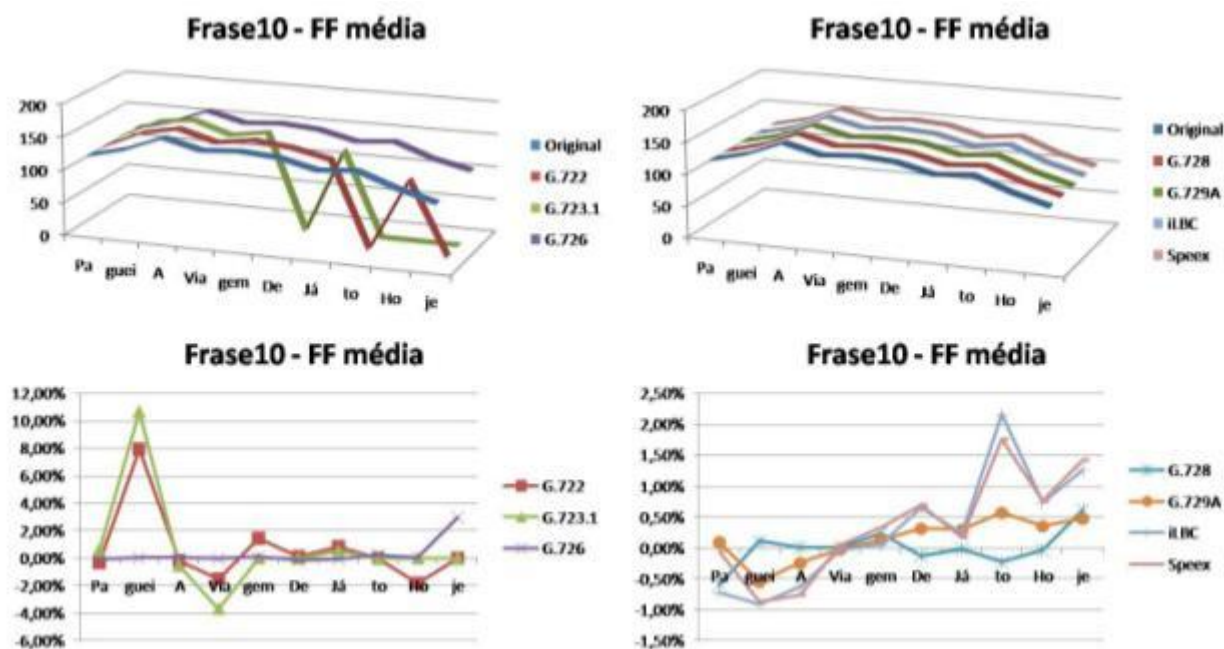


Figura 37 - Variações da freqüência fundamental média da frase 10 do indivíduo H3-14 do grupo HG1.

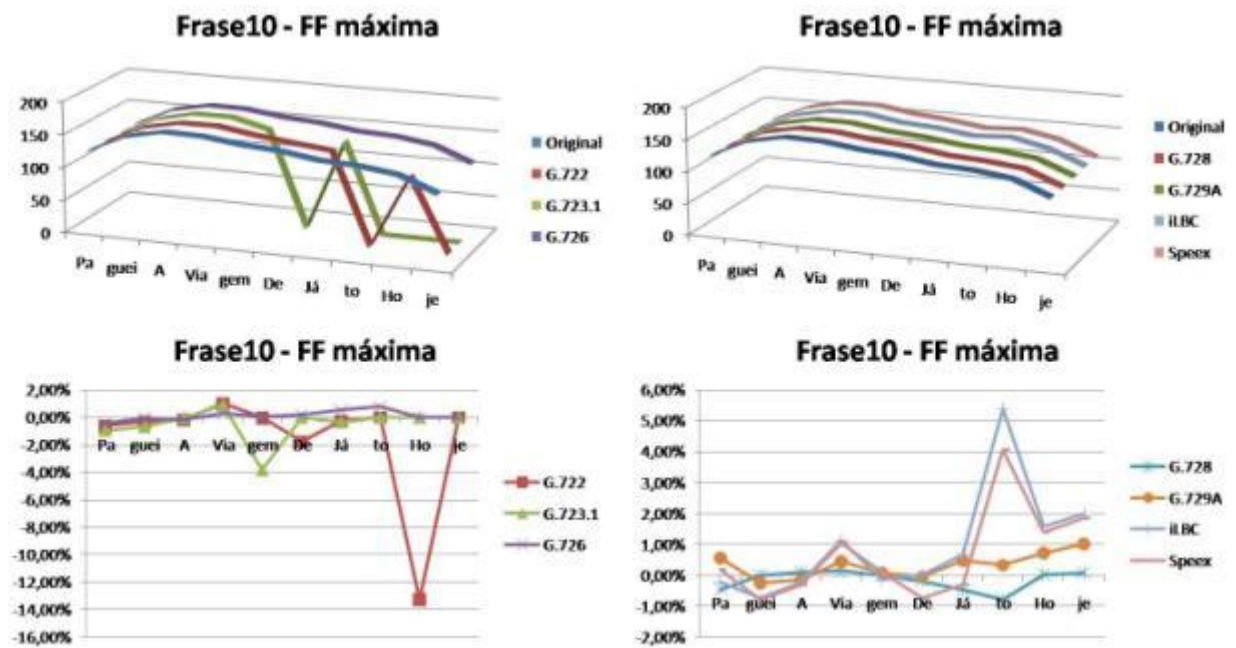


Figura 38 - Variações da frequência fundamental máxima da frase 10 do indivíduo H3-14 do grupo HG1.

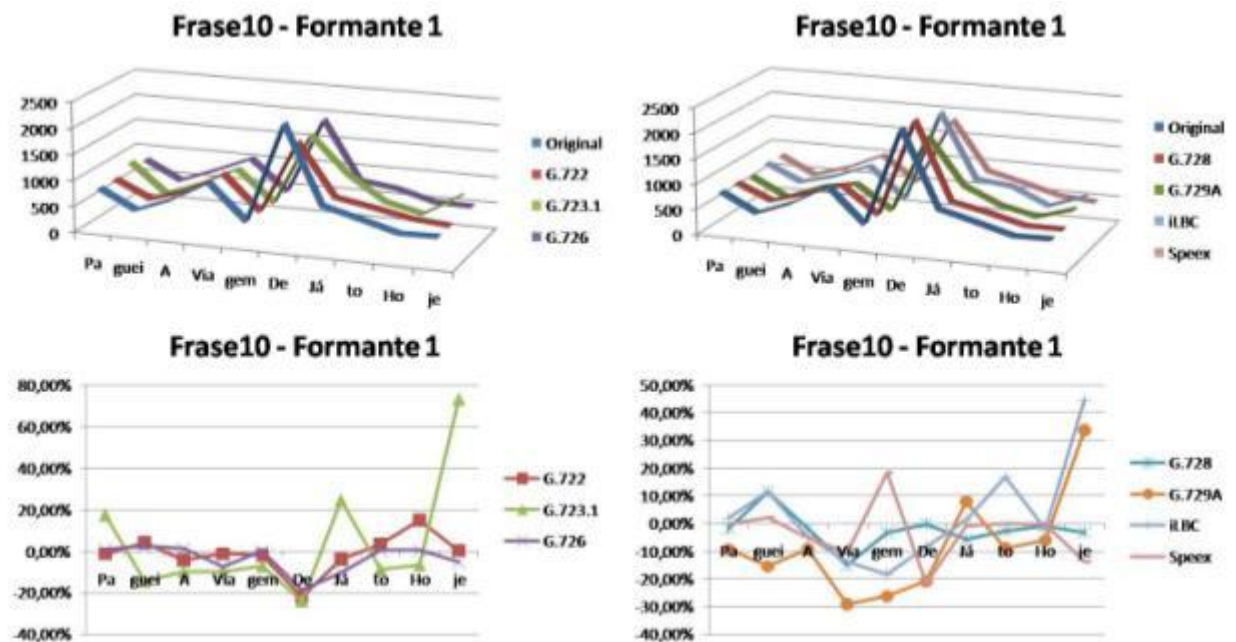


Figura 39 - Variações da frequência da primeira formante da frase 10 do indivíduo H3-14 do grupo HG1.

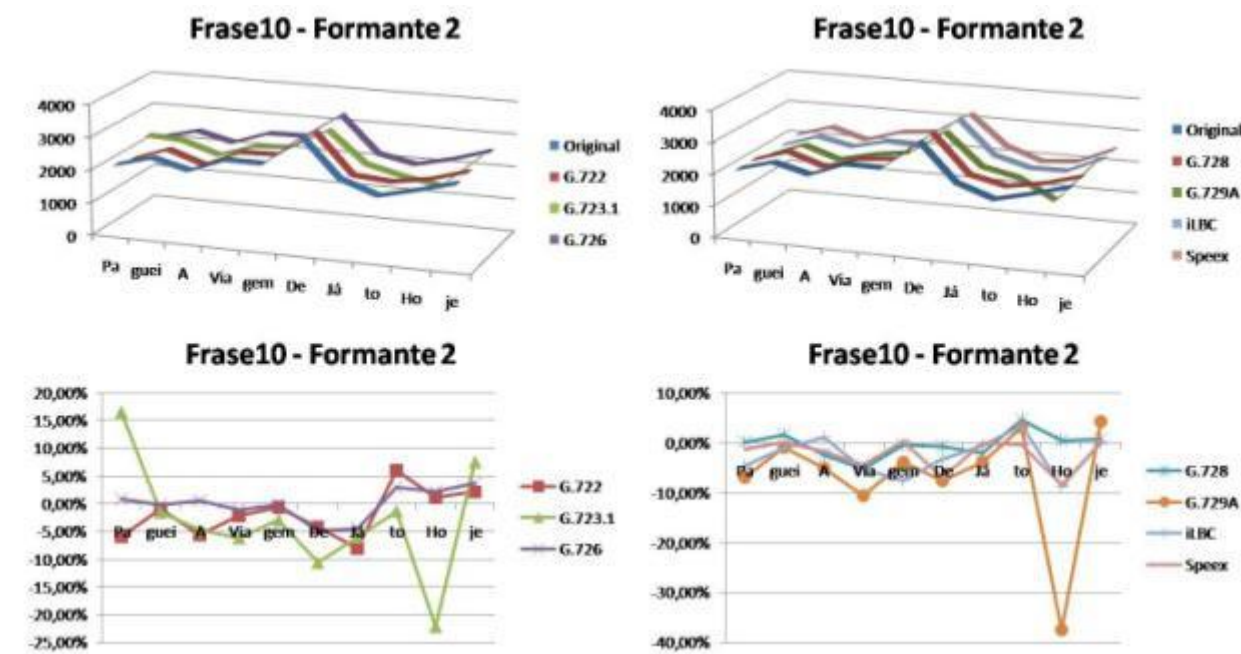


Figura 40 - Variações da frequência da segunda formante da frase 10 do indivíduo H3-14 do grupo HG1.

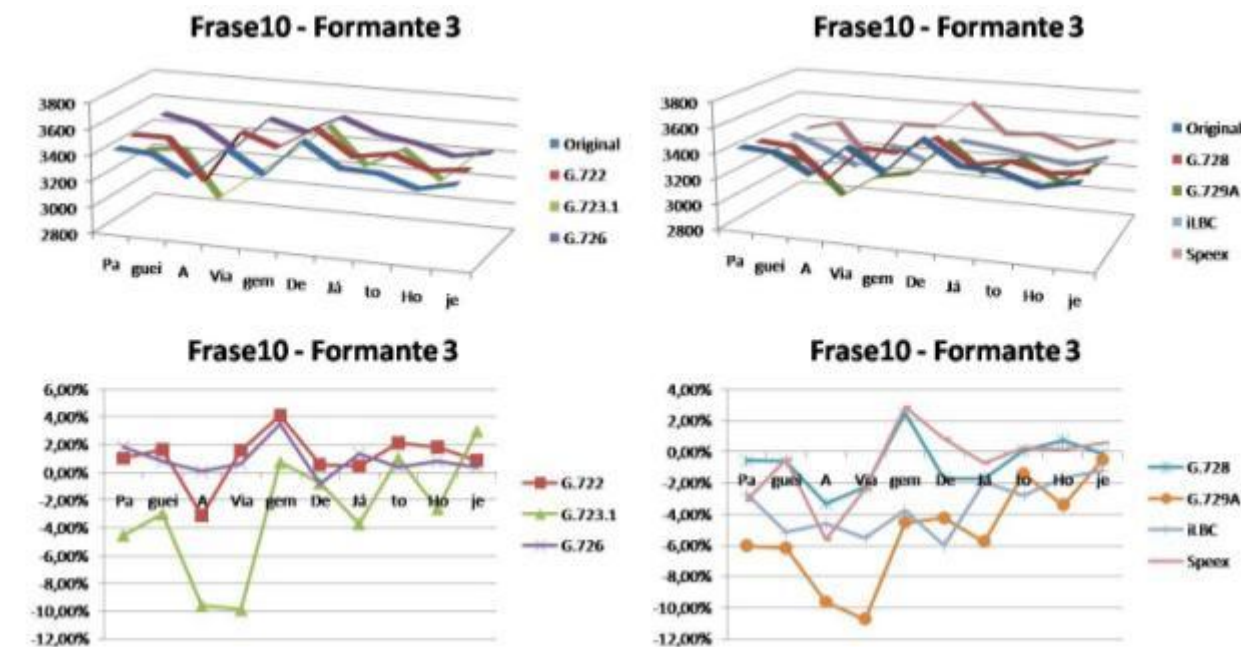


Figura 41 - Variações da frequência da terceira formante da frase 10 do indivíduo H3-14 do grupo HG1.

4.4.2 – HG2

Fazem parte do grupo os indivíduos H6-54 e H9-35. O indivíduo H6-54 foi escolhido para ter seus resultados individuais apresentados por possuir os maiores erros de processamento do grupo.

Quanto à avaliação perceptual, o indivíduo H6-54 não pronunciou o intervalo “o” da frase 2. Portanto seus gráficos possuem frequência fundamental nula em tal intervalo.

Neste grupo, nenhum dos indivíduos apresentou erros de processamento com a inserção de ruído e forte atenuação.

Na tabela 17 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo H6-54 e o grupo HG2. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento.

Tabela 17 - Duração dos intervalos médios e dos arquivos do grupo HG2 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento			
		Intervalo Médio (s)	Duração (s)	G.723.1	G.728	G.729A	iLBC
H6-54	Frase 1	0,224	2,687	22,33%	11,16%	3,72%	7,44%
	Frase 2	0,199	2,766	21,69%	14,46%	3,62%	3,62%
	Frase 3	0,292	1,420	21,12%	7,04%	0,00%	7,04%
	Frase 4	0,136	1,772	16,93%	11,29%	0,00%	5,64%
	Frase 5	0,164	1,474	20,35%	6,78%	0,00%	6,78%
	Frase 6	0,140	1,120	17,86%	8,93%	0,00%	8,93%
	Frase 7	0,176	1,939	20,63%	10,32%	0,00%	5,16%
	Frase 8	0,187	1,781	16,84%	11,23%	0,00%	5,61%
	Frase 9	0,165	1,984	25,20%	10,08%	5,04%	5,04%
	Frase 10	0,159	1,588	18,89%	6,30%	0,00%	6,30%
	Frase 11	0,173	1,730	17,34%	11,56%	0,00%	5,78%
	Frase 12	0,223	2,893	20,74%	10,37%	0,00%	3,46%
HG2	Frase 1	0,231	2,772	21,67%	10,83%	3,61%	7,22%
	Frase 2	0,195	2,813	21,33%	12,47%	3,56%	3,56%
	Frase 3	0,252	1,660	21,09%	8,79%	0,00%	6,15%
	Frase 4	0,183	2,384	18,48%	10,65%	1,67%	6,16%
	Frase 5	0,223	2,005	22,00%	7,33%	1,97%	7,33%
	Frase 6	0,164	1,312	18,90%	11,11%	0,00%	7,79%
	Frase 7	0,209	2,304	21,56%	10,78%	1,87%	4,45%
	Frase 8	0,185	1,896	20,85%	10,59%	2,49%	5,29%
	Frase 9	0,184	2,212	24,90%	11,19%	4,57%	4,57%
	Frase 10	0,202	2,018	21,70%	9,28%	2,04%	5,19%
	Frase 11	0,204	2,040	19,31%	12,16%	2,13%	5,02%
	Frase 12	0,236	3,073	21,13%	9,80%	1,54%	4,80%

Nas tabelas 18 e 19 temos os erros de processamento das frequências fundamentais mínimas, médias e máximas para cada frase para o indivíduo H6-54 e médios do grupo HG2. Pode-se observar que os codificadores G.723.1, G.729A, iLBC e Speex possuem os maiores erros de processamento da frequência fundamental.

Tabela 18 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H6-54 e o grupo HG2.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
H6-54	Frase 1	0,41%	0,17%	0,82%	-1,07%	-0,37%	-0,65%	-0,46%	-0,14%	0,51%
	Frase 2	-1,75%	1,74%	19,25%	-1,79%	0,32%	0,29%	-0,26%	0,15%	0,39%
	Frase 3	0,69%	0,44%	1,50%	2,12%	0,07%	-0,70%	0,47%	0,53%	2,29%
	Frase 4	-0,39%	0,21%	1,80%	-1,32%	0,77%	1,50%	-0,47%	-0,17%	0,29%
	Frase 5	-2,26%	-0,12%	0,00%	0,26%	0,35%	-0,75%	-0,37%	-0,24%	-0,18%
	Frase 6	0,76%	0,85%	1,56%	-3,79%	0,68%	0,88%	0,09%	0,07%	-0,66%
	Frase 7	0,30%	0,09%	1,08%	0,17%	0,01%	0,70%	0,94%	0,43%	0,93%
	Frase 8	1,15%	0,17%	0,32%	2,87%	5,69%	24,88%	0,09%	-0,02%	0,14%
	Frase 9	-1,10%	-0,62%	0,03%	0,45%	1,89%	6,65%	-0,02%	-0,25%	-0,02%
	Frase 10	-1,11%	-0,23%	-0,17%	1,32%	-0,55%	-3,32%	-1,57%	-0,03%	0,22%
	Frase 11	-0,16%	-0,79%	-1,06%	0,34%	0,01%	0,20%	-0,44%	-0,36%	-0,28%
	Frase 12	0,31%	0,11%	0,78%	1,30%	0,29%	-0,57%	-0,18%	0,01%	-0,02%
HG2	Frase 1	-0,17%	-0,05%	0,03%	-0,54%	0,08%	-0,43%	0,04%	-0,29%	0,20%
	Frase 2	-4,29%	-3,34%	3,89%	-1,23%	-0,76%	-0,22%	-0,14%	-0,41%	-0,53%
	Frase 3	0,04%	-0,11%	-0,72%	0,36%	-0,15%	-1,53%	0,87%	-0,14%	1,16%
	Frase 4	-0,14%	-0,02%	0,41%	-0,54%	0,26%	0,26%	-0,21%	-0,17%	-0,12%
	Frase 5	-1,03%	-0,38%	0,58%	-0,16%	-0,07%	-1,16%	-0,07%	-0,22%	0,02%
	Frase 6	0,64%	-0,52%	-2,45%	-1,76%	-1,06%	-3,13%	0,13%	0,06%	-0,36%
	Frase 7	0,23%	0,05%	0,52%	0,25%	0,89%	6,94%	0,56%	0,26%	0,69%
	Frase 8	-0,22%	0,02%	0,12%	0,94%	2,82%	12,19%	0,23%	0,01%	0,09%
	Frase 9	-0,47%	-0,37%	-0,12%	0,37%	0,92%	2,12%	0,11%	0,35%	-0,25%
	Frase 10	-0,32%	-0,18%	-0,28%	4,92%	2,06%	7,51%	-1,63%	-0,46%	0,08%
	Frase 11	0,08%	-0,33%	-0,48%	0,45%	0,00%	-0,41%	-0,24%	-0,17%	-0,15%
	Frase 12	0,34%	-0,60%	0,29%	0,69%	0,48%	1,52%	0,05%	-0,13%	-0,01%

Tabela 19 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H6-54 e o grupo HG2.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
H6-54	Frase 1	-0,42%	-0,08%	0,53%	-0,36%	0,26%	1,85%	0,36%	2,93%	4,96%	0,08%	-0,11%	0,29%
	Frase 2	0,16%	0,35%	0,39%	0,26%	0,13%	-0,19%	0,52%	0,73%	0,79%	-0,12%	0,18%	0,29%
	Frase 3	-0,07%	0,49%	1,71%	1,12%	0,49%	2,27%	-0,44%	-0,49%	0,42%	2,53%	0,96%	1,71%
	Frase 4	-0,49%	-0,16%	0,36%	-1,40%	0,14%	-0,86%	-3,28%	0,22%	0,62%	-1,53%	0,26%	-0,61%
	Frase 5	-0,51%	0,05%	0,23%	-1,34%	-0,34%	0,05%	-0,31%	0,70%	1,38%	-0,23%	0,64%	2,77%
	Frase 6	0,10%	0,15%	-0,56%	-3,83%	0,22%	1,99%	-3,46%	0,93%	2,27%	-3,32%	1,26%	1,91%
	Frase 7	0,53%	0,00%	1,28%	0,19%	0,15%	1,58%	1,36%	0,05%	0,00%	-2,40%	-0,59%	1,63%
	Frase 8	1,39%	0,29%	0,12%	1,50%	0,09%	0,00%	3,14%	0,47%	0,48%	3,02%	0,61%	0,57%
	Frase 9	0,36%	0,25%	0,97%	1,36%	0,50%	1,56%	0,39%	0,70%	3,18%	1,19%	0,62%	1,99%
	Frase 10	-0,38%	0,20%	0,55%	-0,06%	-0,03%	0,01%	0,23%	0,62%	0,38%	2,23%	0,89%	0,63%
	Frase 11	0,27%	-0,06%	-0,43%	-1,85%	-0,82%	0,78%	2,11%	1,59%	2,67%	2,96%	0,73%	0,71%
	Frase 12	-0,74%	0,00%	-0,13%	0,35%	0,66%	1,19%	-2,04%	0,22%	2,63%	-0,98%	0,11%	0,50%
HG2	Frase 1	0,06%	-0,26%	0,21%	-0,73%	0,05%	0,67%	0,24%	1,85%	2,06%	-0,26%	-0,06%	-0,28%
	Frase 2	0,07%	-0,31%	-0,53%	-0,03%	-4,01%	-10,72%	-1,41%	-5,33%	-11,41%	-1,91%	-5,67%	-11,69%
	Frase 3	0,60%	-0,16%	0,87%	-0,10%	0,01%	1,03%	-0,97%	-0,40%	-0,49%	1,31%	0,07%	0,50%
	Frase 4	-0,22%	-0,16%	-0,08%	-0,50%	-0,03%	-0,35%	-1,51%	0,08%	0,11%	-0,75%	0,08%	-0,15%
	Frase 5	-0,14%	-0,07%	0,23%	-2,67%	-3,18%	-4,18%	-0,72%	-0,13%	1,45%	-0,15%	0,05%	0,93%
	Frase 6	0,13%	0,10%	-0,31%	-1,75%	-1,38%	-1,85%	-1,41%	-0,93%	-1,43%	-1,54%	-0,49%	-1,78%
	Frase 7	0,36%	0,05%	0,55%	0,05%	0,13%	0,98%	1,00%	0,05%	-0,65%	-1,22%	-0,21%	1,47%
	Frase 8	0,88%	0,16%	0,09%	0,32%	0,12%	0,20%	1,14%	0,42%	0,67%	1,63%	0,30%	0,30%
	Frase 9	0,32%	0,58%	0,73%	0,75%	-0,43%	-1,47%	0,28%	-0,45%	-0,54%	-2,04%	-0,75%	-1,32%
	Frase 10	-1,03%	-0,35%	0,25%	-1,02%	-4,95%	-5,60%	-1,06%	0,31%	2,14%	0,95%	0,33%	0,03%
	Frase 11	0,11%	-0,02%	-0,22%	-1,59%	-0,47%	0,21%	1,39%	2,25%	6,22%	1,36%	0,25%	-0,34%
	Frase 12	-0,23%	-0,13%	-0,07%	1,85%	-0,79%	-1,63%	-1,78%	-0,66%	1,41%	1,90%	0,05%	0,43%

Nas tabelas 20 e 21 temos os erros de processamento das formantes para cada frase para o indivíduo H6-54 e médios do grupo HG2. Pode-se observar que os codificadores apresentaram erros de processamento maiores para as formantes do que para a frequência fundamental.

Tabela 20 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H6-54 e o grupo HG2.

		G.722			G.723.1			G.726		
		Fomantes			Fomantes			Fomantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
H6-54	Frase 1	-3,02%	-1,81%	0,66%	-43,04%	-35,46%	-34,22%	-38,74%	-32,91%	-33,41%
	Frase 2	-4,74%	-2,39%	0,52%	-46,06%	-38,49%	-33,05%	-39,73%	-36,96%	-32,21%
	Frase 3	-2,81%	-2,45%	0,36%	-31,14%	-35,18%	-34,42%	-28,93%	-32,71%	-33,91%
	Frase 4	-0,89%	-1,00%	-0,12%	-32,77%	-38,32%	-36,70%	-29,14%	-33,24%	-33,99%
	Frase 5	-44,72%	-38,53%	-33,69%	-49,56%	-38,46%	-33,66%	-45,01%	-38,65%	-31,92%
	Frase 6	-44,70%	-36,06%	-29,72%	-47,73%	-39,01%	-34,44%	-44,11%	-35,61%	-32,12%
	Frase 7	-34,48%	-38,68%	-32,85%	-35,56%	-36,86%	-31,10%	-33,27%	-38,63%	-32,62%
	Frase 8	-25,50%	-39,13%	-33,63%	-30,10%	-43,75%	-34,02%	-25,00%	-38,60%	-31,48%
	Frase 9	-29,55%	-33,12%	-35,82%	-34,23%	-36,63%	-34,61%	-29,52%	-31,58%	-32,72%
	Frase 10	-33,91%	-34,25%	-33,91%	-34,33%	-34,09%	-34,25%	-32,28%	-32,06%	-31,23%
	Frase 11	-42,64%	-37,44%	-33,56%	-45,16%	-39,49%	-32,15%	-43,43%	-35,80%	-32,34%
	Frase 12	-32,58%	-29,58%	-35,57%	-34,63%	-33,58%	-35,53%	-32,60%	-29,08%	-32,51%
HG1	Frase 1	-2,04%	-1,34%	0,79%	-28,04%	-17,95%	-17,83%	-18,35%	-15,15%	-15,91%
	Frase 2	-4,43%	-4,77%	-2,52%	-16,33%	-16,10%	-16,54%	-19,64%	-17,33%	-15,57%
	Frase 3	-0,67%	-1,42%	0,62%	-19,71%	-18,28%	-17,29%	-14,37%	-15,47%	-16,33%
	Frase 4	-0,57%	-1,68%	0,20%	-17,47%	-19,23%	-19,15%	-14,96%	-15,71%	-16,60%
	Frase 5	-21,88%	-19,77%	-16,71%	-29,76%	-18,90%	-17,29%	-22,91%	-18,71%	-15,81%
	Frase 6	-22,47%	-18,63%	-14,44%	-28,87%	-19,55%	-17,25%	-19,93%	-16,78%	-15,68%
	Frase 7	-18,37%	-20,38%	-16,36%	-20,71%	-19,25%	-16,92%	-17,00%	-18,12%	-15,69%
	Frase 8	-12,19%	-20,15%	-16,10%	-18,45%	-22,99%	-17,30%	-11,11%	-17,04%	-14,39%
	Frase 9	-14,38%	-17,16%	-17,71%	-21,47%	-19,08%	-18,98%	-13,08%	-16,21%	-15,71%
	Frase 10	-19,89%	-18,56%	-16,37%	-21,64%	-18,68%	-18,57%	-17,23%	-15,82%	-15,15%
	Frase 11	-21,76%	-19,49%	-16,55%	-28,69%	-19,69%	-17,60%	-21,99%	-18,24%	-15,83%
	Frase 12	-16,08%	-15,87%	-17,30%	-20,84%	-17,95%	-19,02%	-15,46%	-13,85%	-15,64%

Tabela 21 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H3-14 e o grupo HG1.

		G.728			G.729A			iLBC			Speex		
		Fomantes			Fomantes			Fomantes			Fomantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
H6-54	Frase 1	-38,82%	-33,17%	-34,11%	-16,52%	-5,89%	-3,33%	-37,99%	-35,90%	-36,31%	-37,10%	-33,69%	-34,36%
	Frase 2	3,67%	1,57%	-0,53%	-14,22%	-3,81%	-3,03%	-43,51%	-39,48%	-35,36%	-39,45%	-37,83%	-32,75%
	Frase 3	-3,25%	0,92%	-0,90%	-13,11%	-6,28%	-4,91%	-28,93%	-36,12%	-36,56%	-28,44%	-34,29%	-34,07%
	Frase 4	2,19%	1,50%	0,12%	-15,62%	-5,75%	-4,81%	-30,80%	-38,04%	-39,77%	-27,05%	-33,85%	-34,81%
	Frase 5	-10,69%	-1,35%	-0,29%	-18,78%	-2,78%	-3,11%	-44,18%	-39,52%	-34,17%	-41,24%	-37,92%	-32,30%
	Frase 6	-5,24%	-1,08%	-1,65%	-24,63%	-7,42%	-5,20%	-47,46%	-36,67%	-35,65%	-43,46%	-33,41%	-33,65%
	Frase 7	-7,71%	-1,25%	-0,77%	-17,31%	-3,45%	-4,56%	-35,90%	-40,30%	-34,23%	-33,74%	-37,57%	-32,11%
	Frase 8	-2,16%	1,67%	1,06%	-14,12%	-9,21%	-5,89%	-27,29%	-40,77%	-35,25%	-25,48%	-38,30%	-32,34%
	Frase 9	2,17%	0,03%	-1,36%	-14,08%	-4,48%	-5,74%	-31,34%	-38,03%	-37,46%	-30,80%	-33,05%	-34,90%
	Frase 10	-2,11%	1,32%	-1,02%	-10,63%	-9,86%	-7,47%	-36,80%	-36,05%	-35,43%	-33,13%	-34,42%	-33,69%
	Frase 11	-7,29%	0,00%	-0,97%	-22,76%	-7,96%	-5,01%	-42,80%	-39,95%	-30,99%	-41,08%	-36,69%	-32,18%
	Frase 12	-4,40%	0,67%	2,15%	-12,62%	-5,09%	-5,18%	-33,50%	-32,94%	-37,16%	-33,27%	-30,27%	-34,31%
HG1	Frase 1	-18,39%	-15,28%	-16,26%	-14,38%	-5,00%	-3,52%	-20,36%	-19,34%	-19,89%	-17,87%	-15,97%	-16,19%
	Frase 2	2,06%	1,93%	0,27%	-15,24%	-6,94%	-3,01%	-26,46%	-24,89%	-18,63%	-21,53%	-19,66%	-15,80%
	Frase 3	-1,54%	1,35%	0,17%	-7,83%	-4,89%	-4,54%	-16,33%	-20,74%	-19,65%	-15,60%	-17,05%	-16,38%
	Frase 4	0,71%	1,66%	0,46%	-10,50%	-4,44%	-4,67%	-16,03%	-21,23%	-21,49%	-12,67%	-16,89%	-17,30%
	Frase 5	-5,75%	-0,06%	0,01%	-13,65%	-4,02%	-3,01%	-25,60%	-22,03%	-18,10%	-19,86%	-18,94%	-15,95%
	Frase 6	-0,49%	0,48%	-0,45%	-17,55%	-4,17%	-4,65%	-28,15%	-20,50%	-19,30%	-23,55%	-16,79%	-16,70%
	Frase 7	-4,23%	0,57%	0,42%	-13,83%	-4,42%	-4,95%	-20,09%	-21,69%	-18,55%	-17,91%	-18,98%	-15,63%
	Frase 8	0,31%	3,09%	1,88%	-10,45%	-7,77%	-5,49%	-14,73%	-23,69%	-19,35%	-11,77%	-18,17%	-14,42%
	Frase 9	2,83%	0,58%	-0,14%	-14,79%	-4,45%	-6,17%	-20,67%	-21,61%	-20,34%	-18,86%	-16,96%	-17,17%
	Frase 10	-2,15%	0,87%	-0,05%	-17,62%	-13,25%	-11,28%	-22,65%	-19,87%	-19,29%	-17,78%	-17,47%	-16,73%
	Frase 11	-3,92%	-0,34%	-0,14%	-17,79%	-6,56%	-5,10%	-25,62%	-22,72%	-17,29%	-21,96%	-18,46%	-16,00%
	Frase 12	-1,36%	1,03%	1,69%	-8,62%	-5,24%	-4,90%	-17,11%	-18,15%	-19,95%	-16,63%	-15,48%	-16,93%

Nas figuras de 42 a 59 temos os resultados do indivíduo H6-54 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas das frequências em Hertz por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

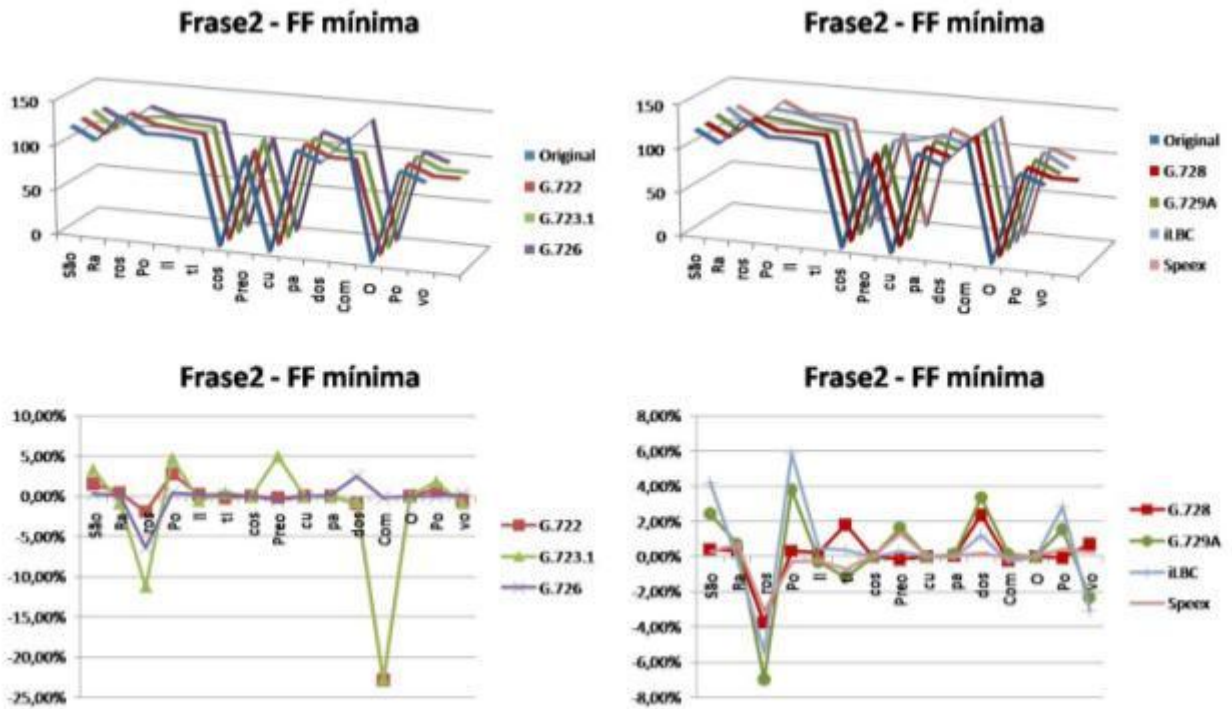


Figura 42 - Variações da frequência fundamental mínima da frase 2 do indivíduo H6-54 do grupo HG2.

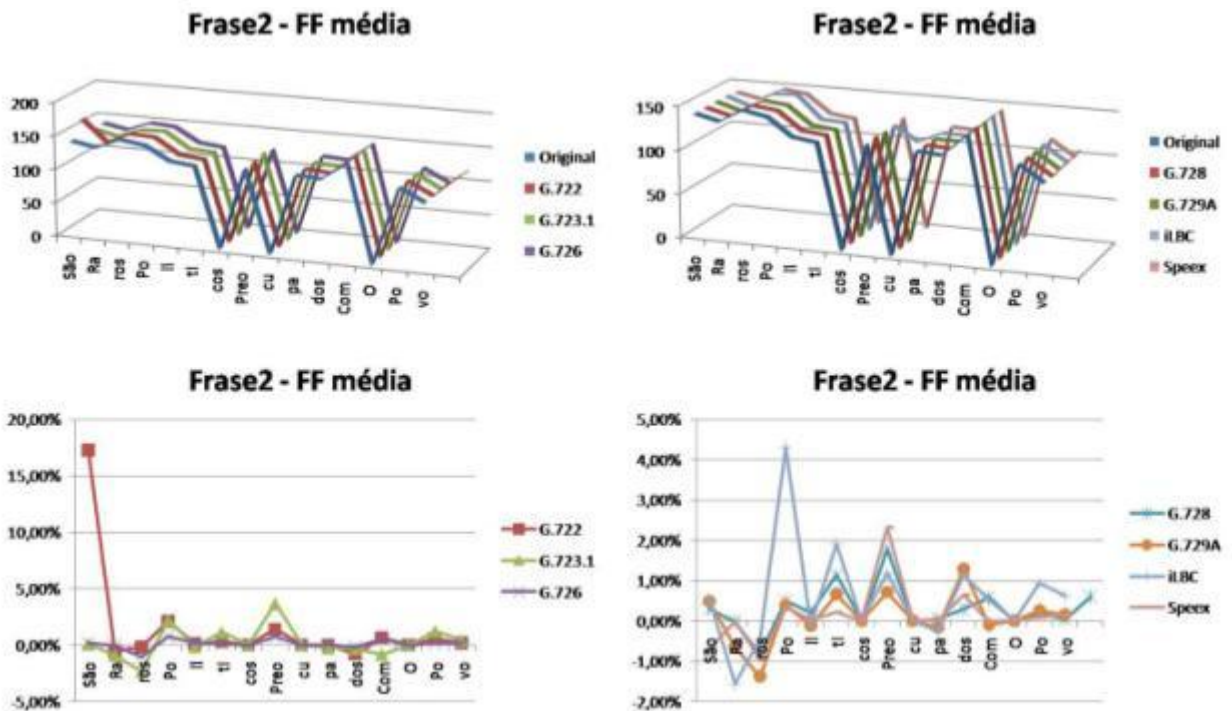


Figura 43 - Variações da frequência fundamental média da frase 2 do indivíduo H6-54 do grupo HG2.

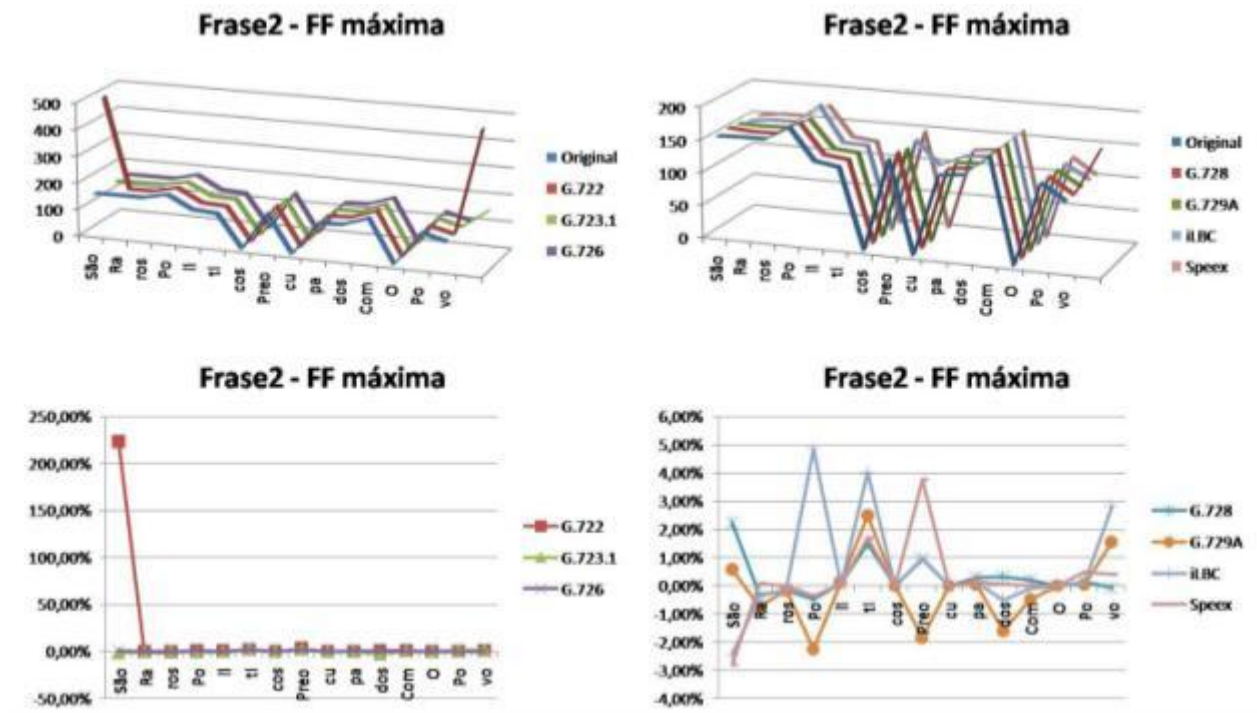


Figura 44 - Variações da frequência fundamental máxima da frase 2 do indivíduo H6-54 do grupo HG2.

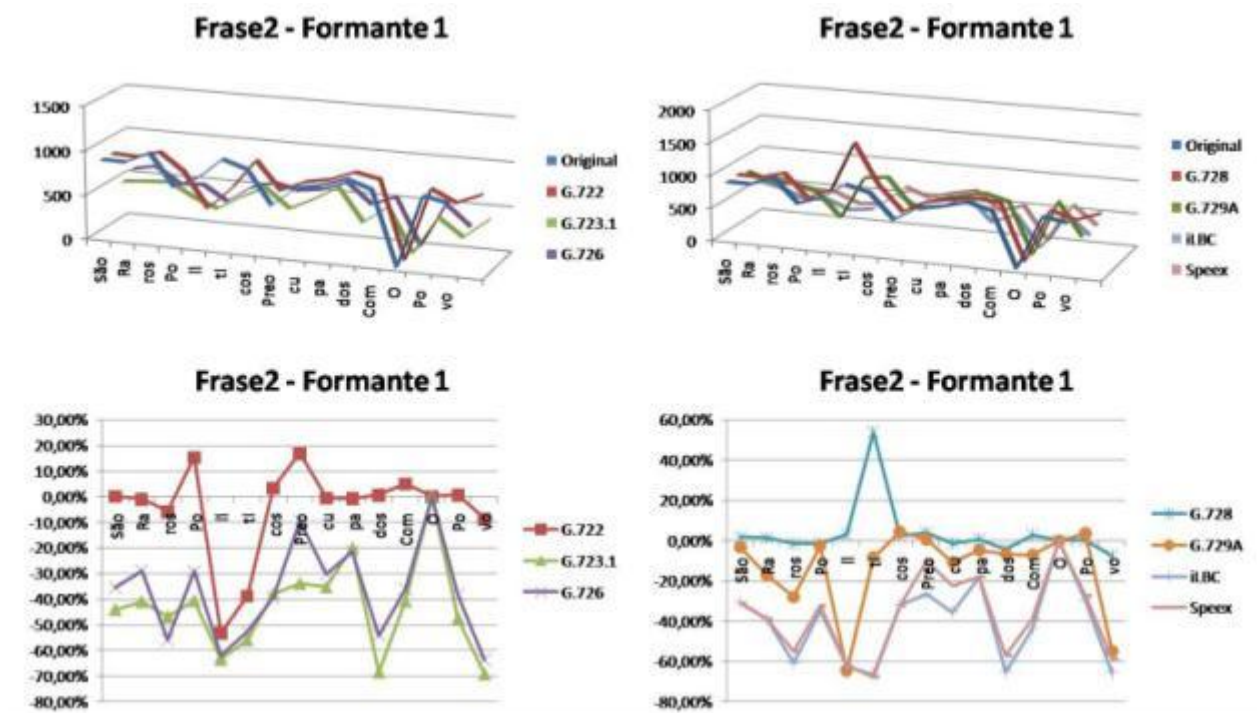


Figura 45 - Variações da frequência da primeira formante da frase 2 do indivíduo H6-54 do grupo HG2.

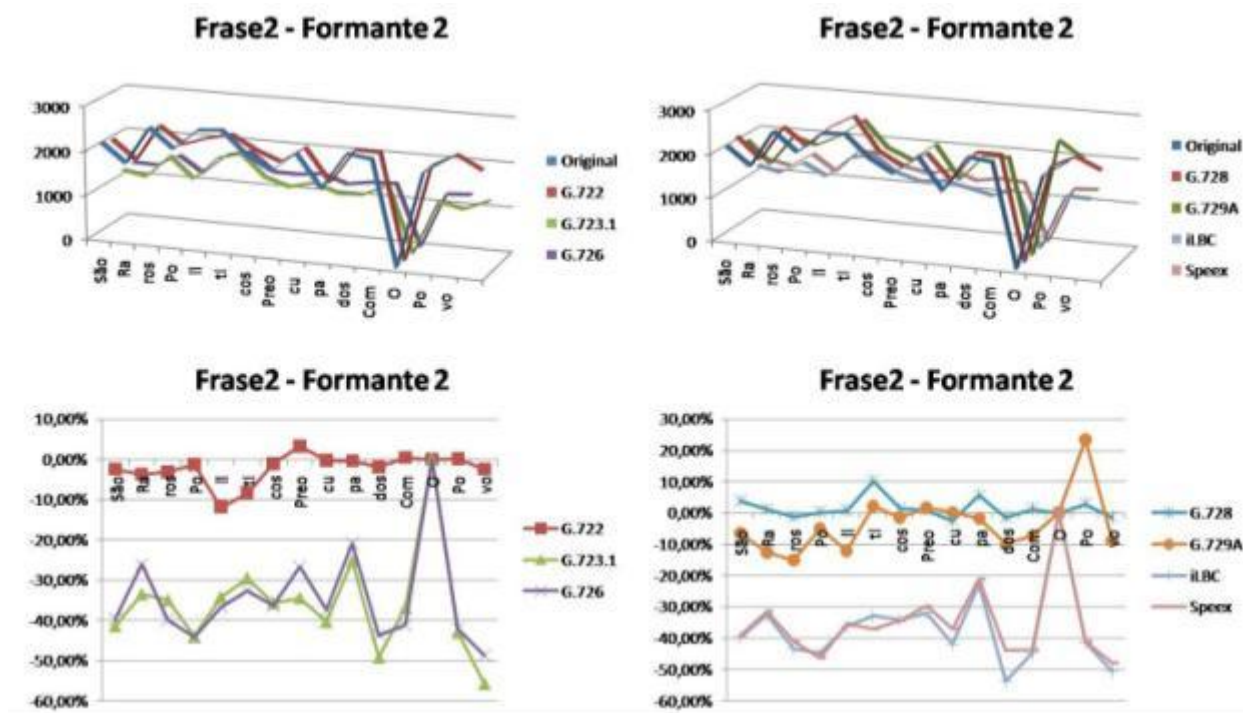


Figura 46 - Variações da frequência da segunda formante da frase 2 do indivíduo H6-54 do grupo HG2.

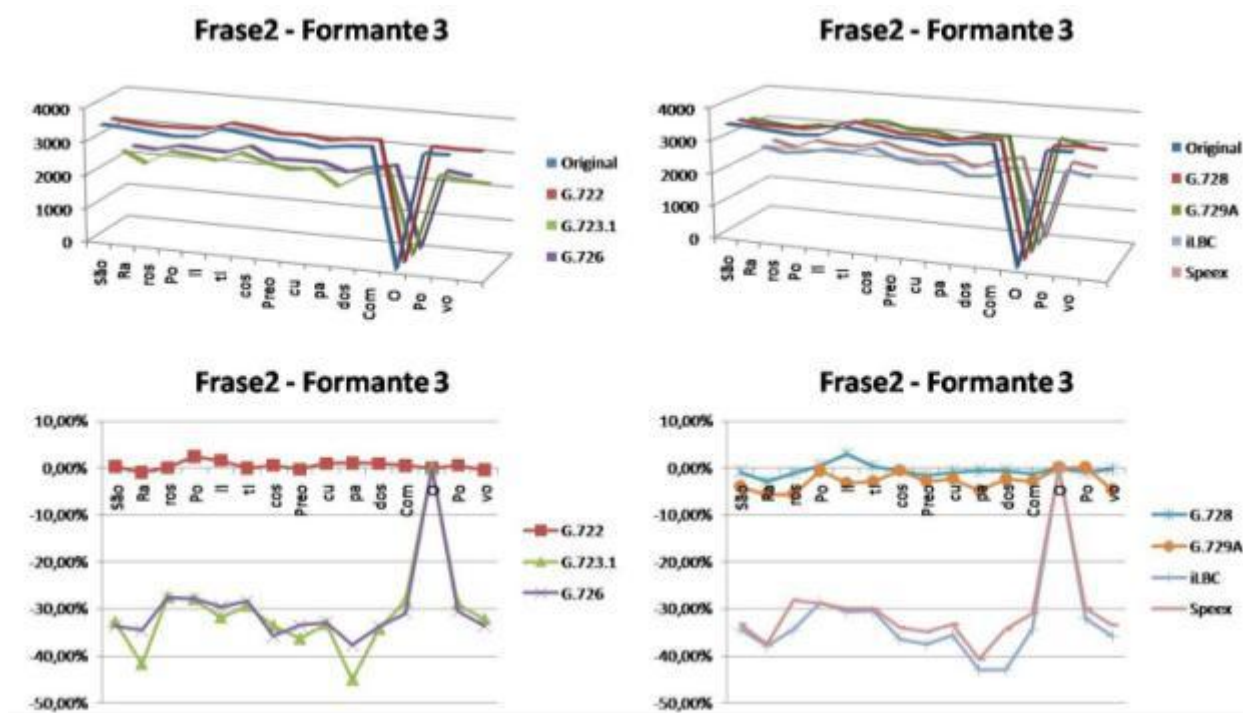


Figura 47 - Variações da frequência da terceira formante da frase 2 do indivíduo H6-54 do grupo HG2.

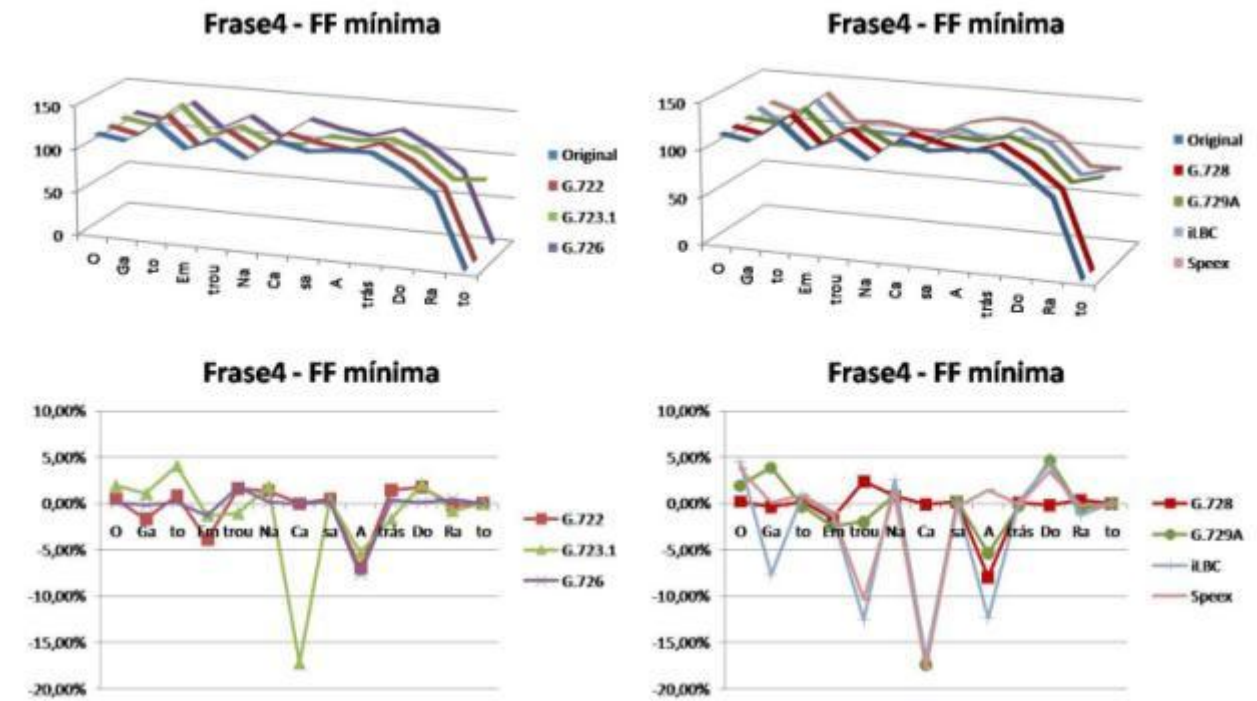


Figura 48 - Variações da frequência fundamental mínima da frase 4 do indivíduo H6-54 do grupo HG2.

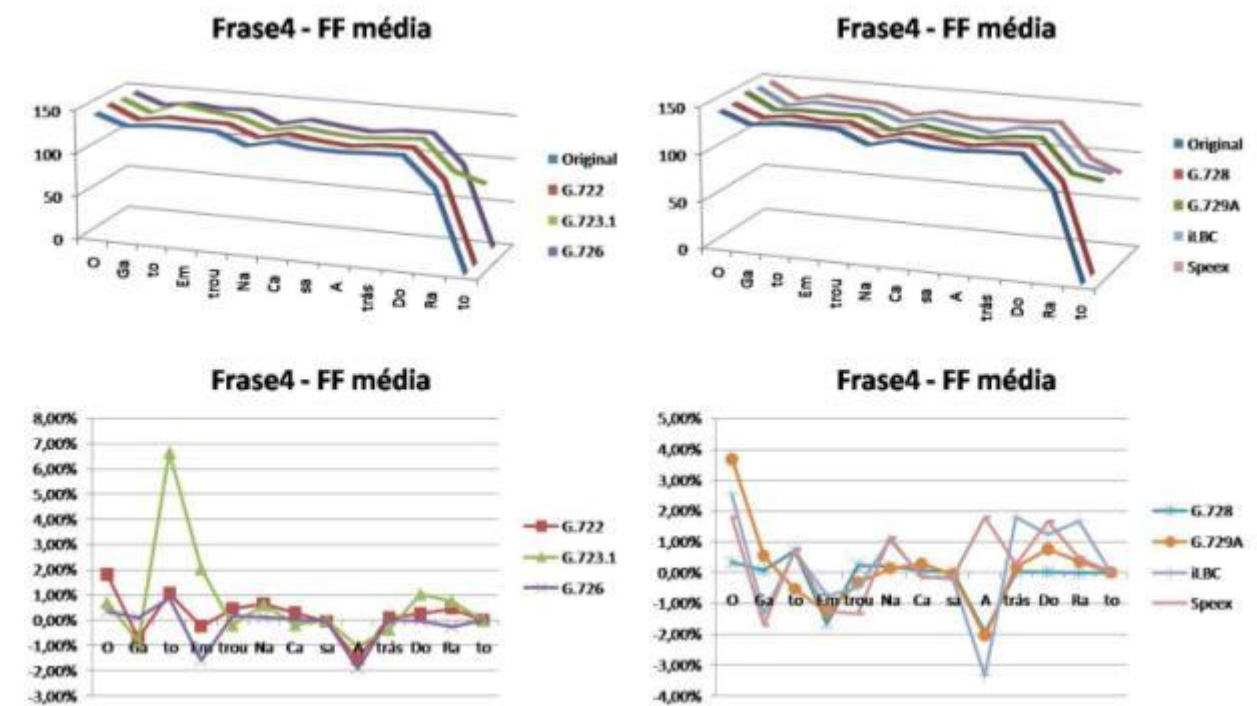


Figura 49 - Variações da frequência fundamental média da frase 4 do indivíduo H6-54 do grupo HG2.

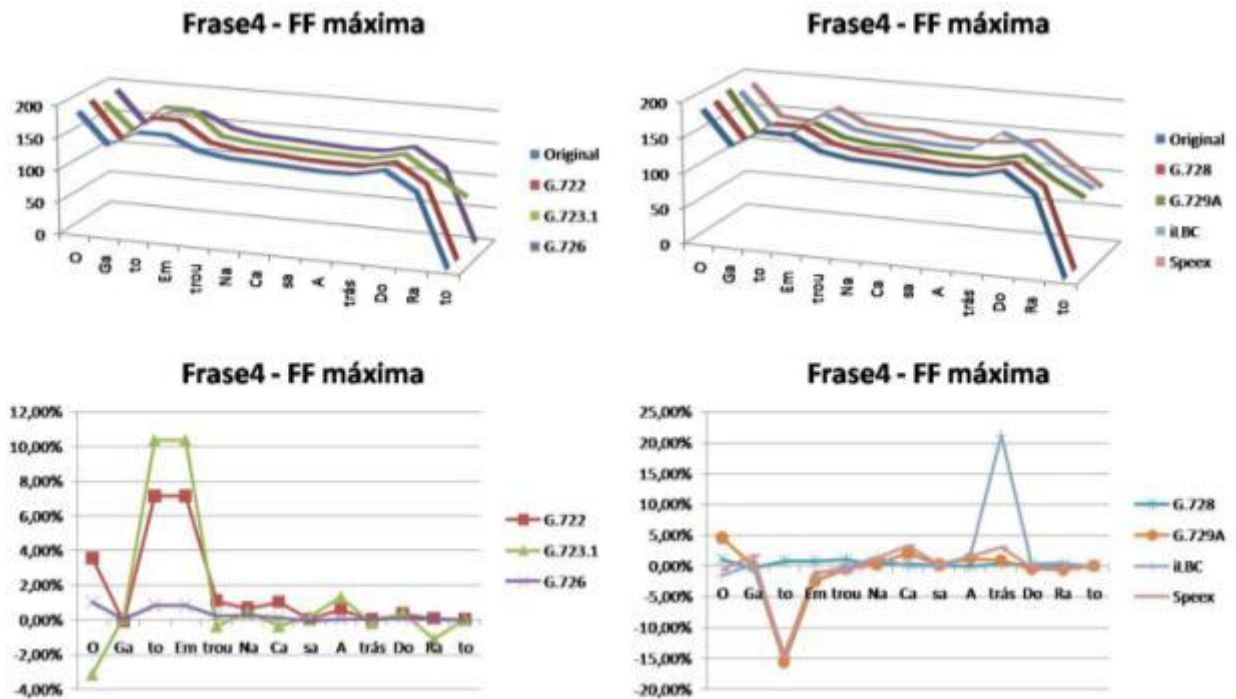


Figura 50 - Variações da frequência fundamental máxima da frase 4 do indivíduo H6-54 do grupo HG2.

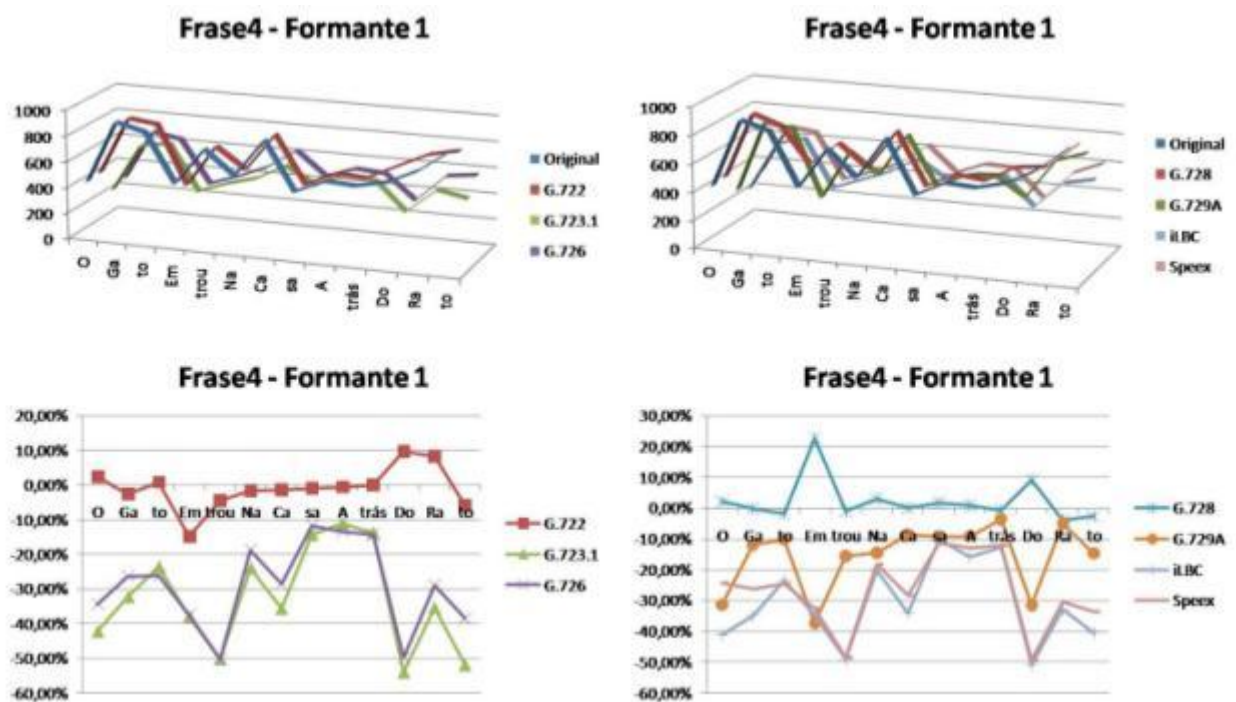


Figura 51 - Variações da frequência da primeira formante da frase 4 do indivíduo H6-54 do grupo HG2.

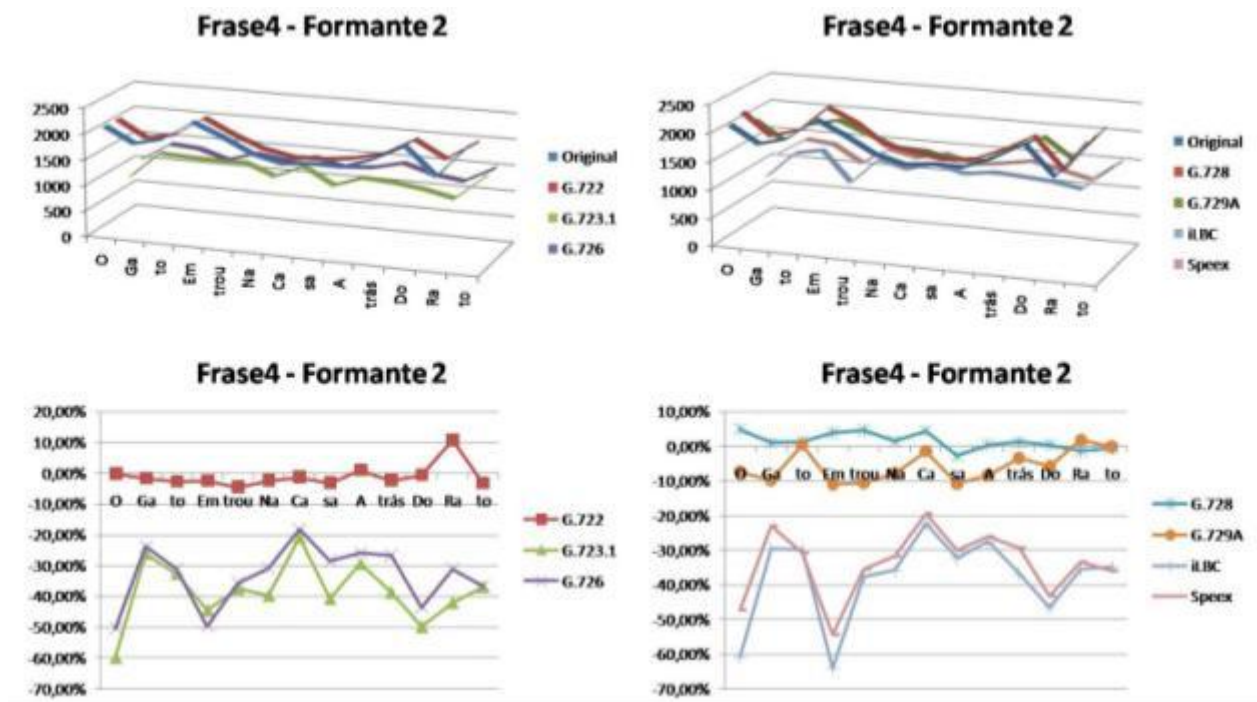


Figura 52 - Variações da frequência da segunda formante da frase 4 do indivíduo H6-54 do grupo HG2.

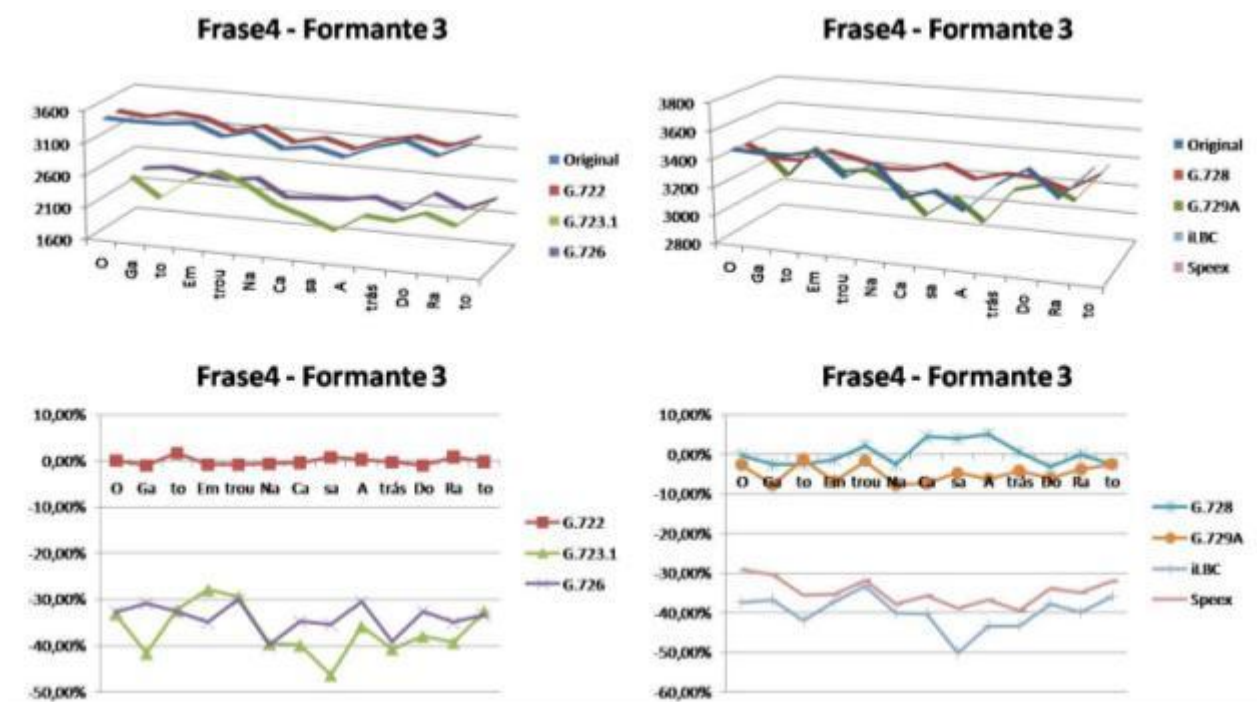


Figura 53 - Variações da frequência da terceira formante da frase 4 do indivíduo H6-54 do grupo HG2.

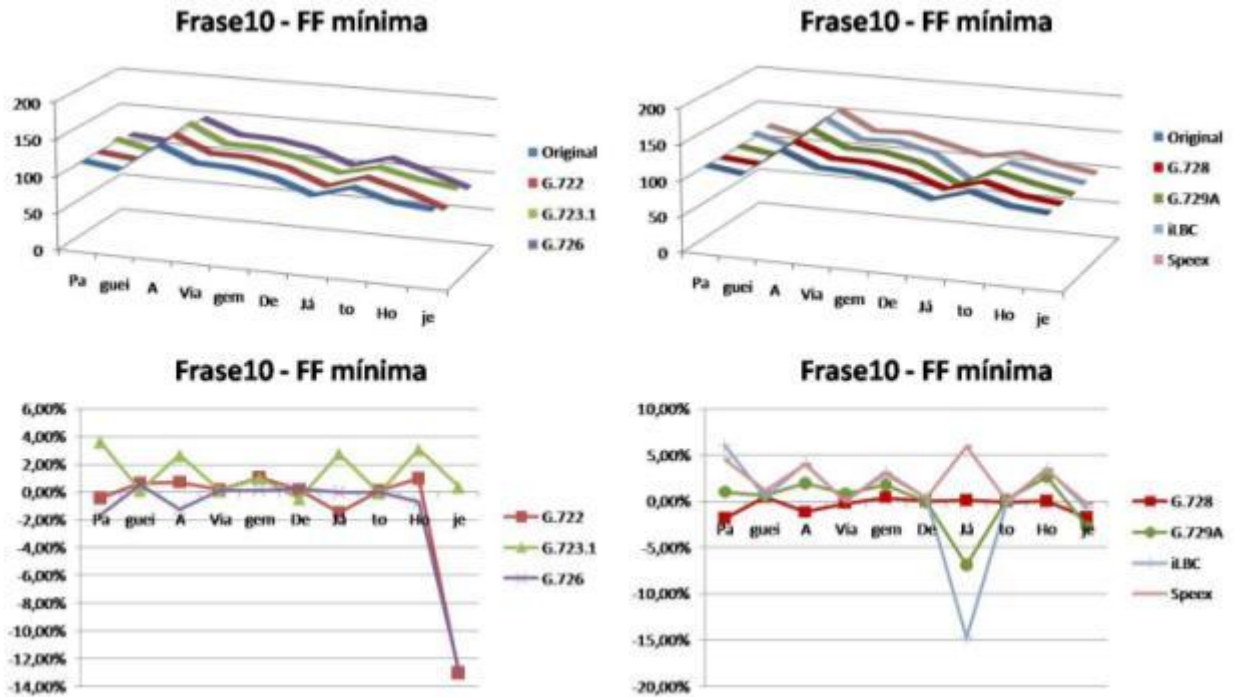


Figura 54 - Variações da frequência fundamental mínima da frase 10 do indivíduo H6-54 do grupo HG2.

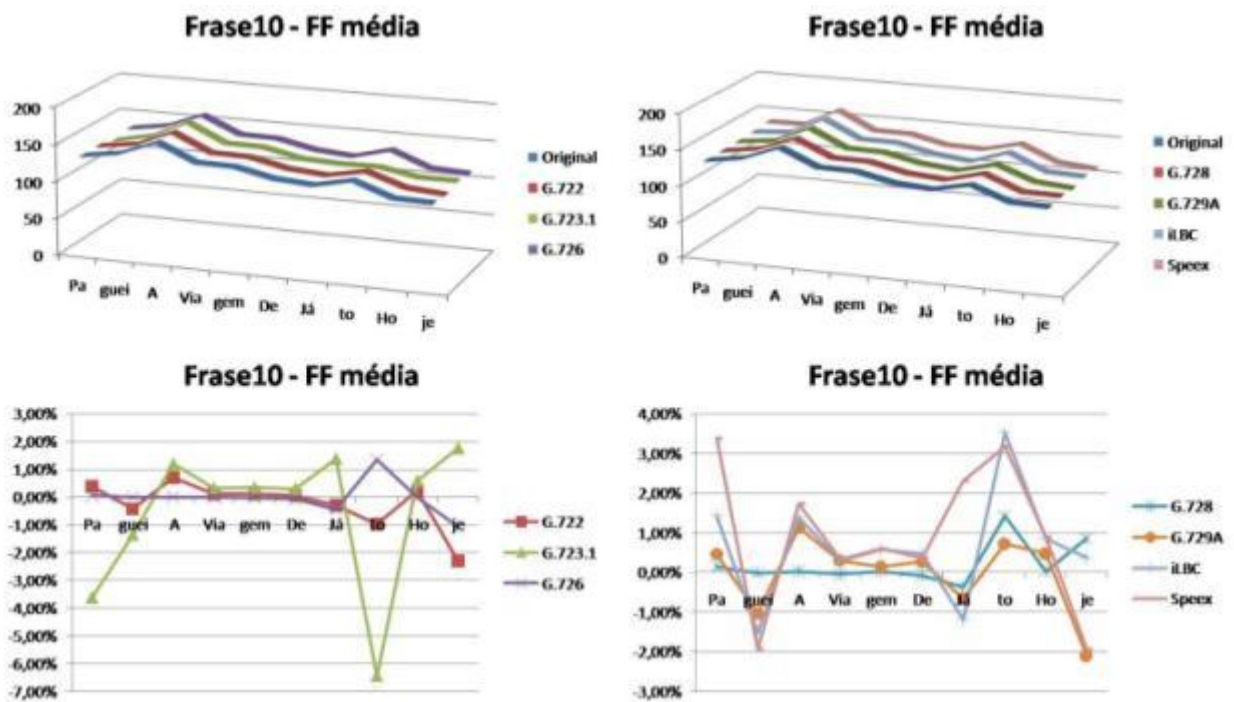


Figura 55 - Variações da frequência fundamental média da frase 10 do indivíduo H6-54 do grupo HG2.

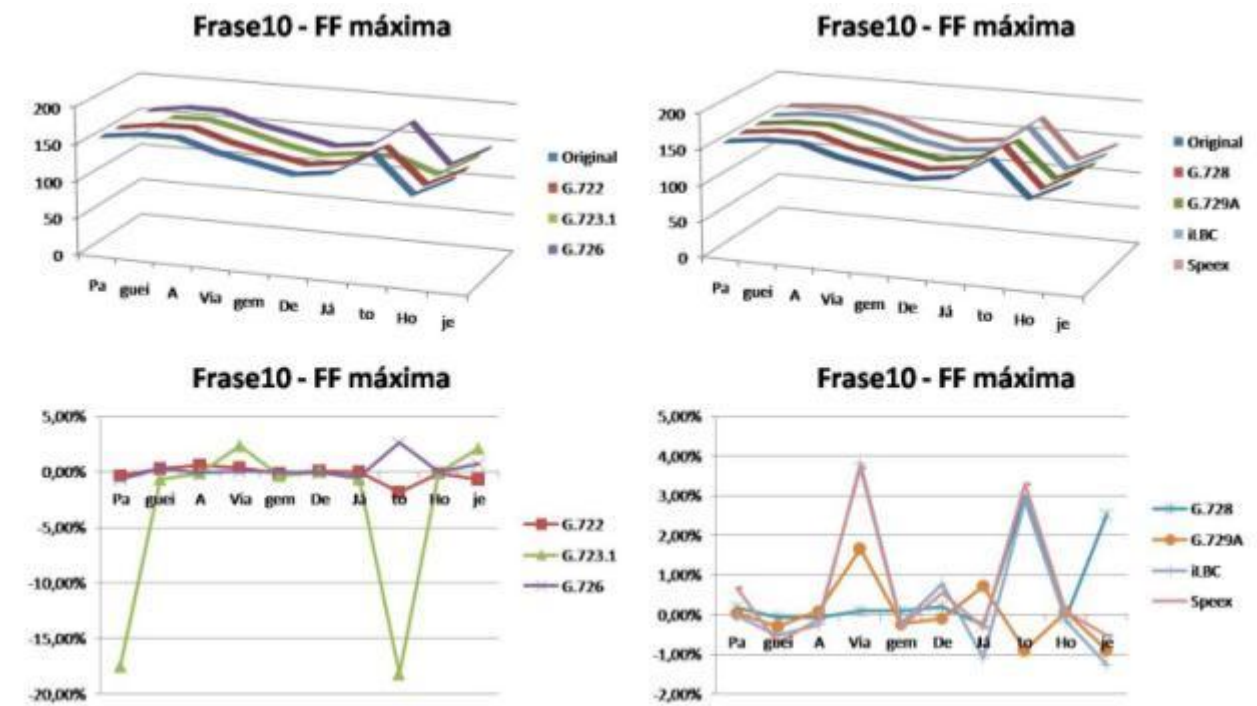


Figura 56 - Variações da frequência fundamental máxima da frase 10 do indivíduo H6-54 do grupo HG2.

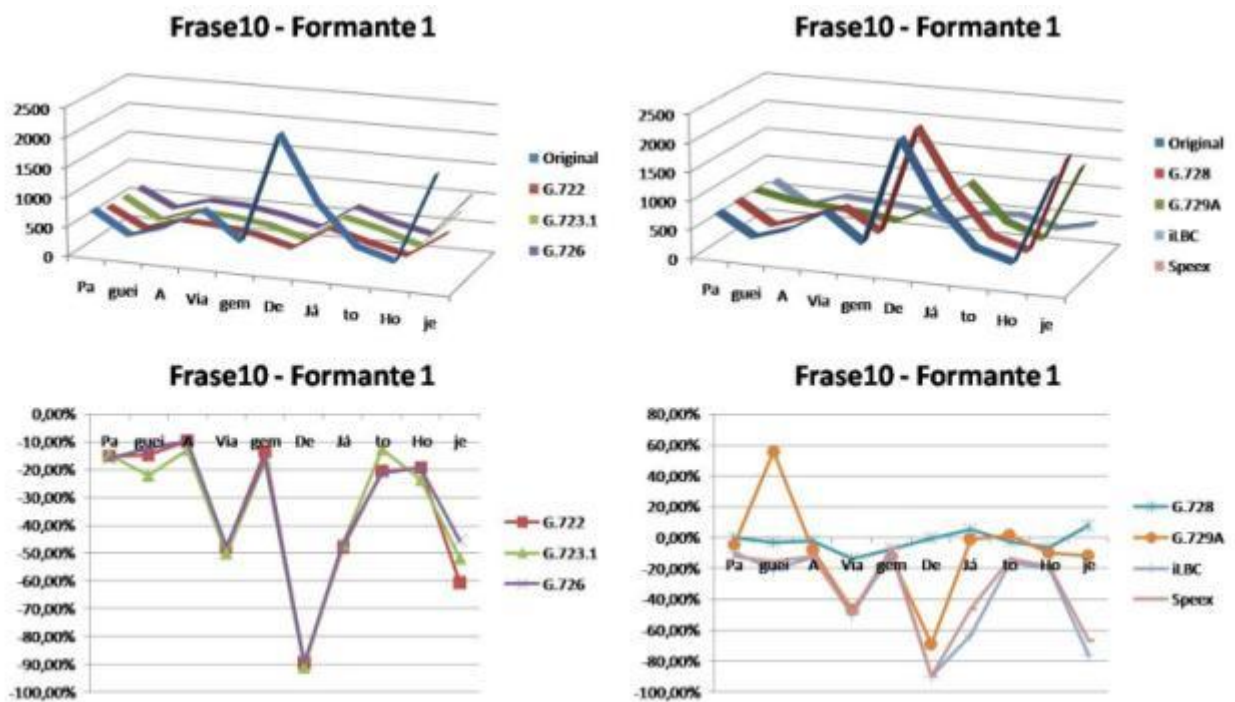


Figura 57 - Variações da frequência da primeira formante da frase 10 do indivíduo H6-54 do grupo HG2.

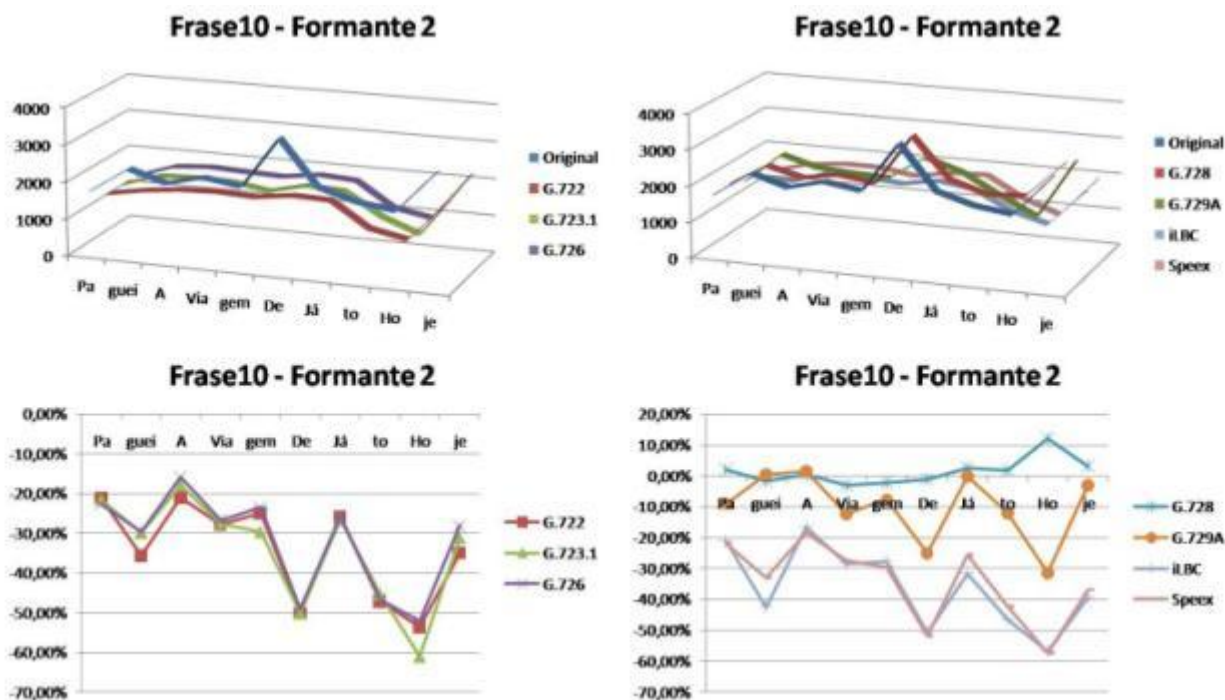


Figura 58 - Variações da frequência da segunda formante da frase 10 do indivíduo H6-54 do grupo HG2.

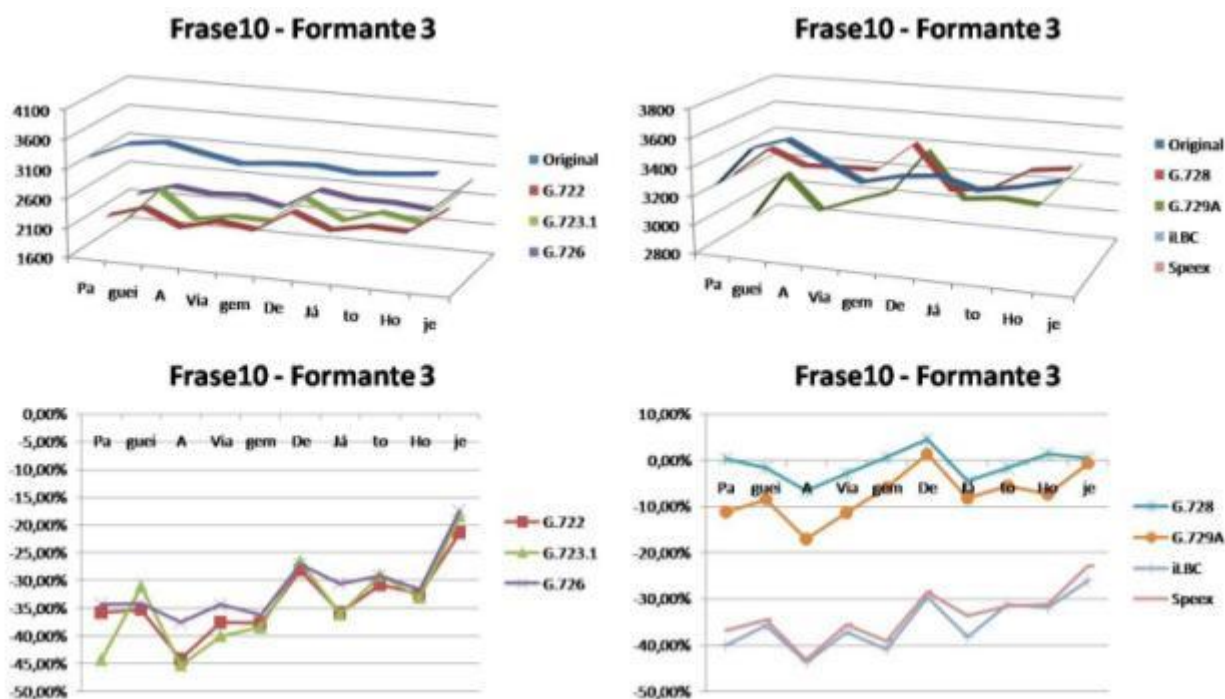


Figura 59 - Variações da frequência da terceira formante da frase 10 do indivíduo H6-54 do grupo HG2.

4.4.3 – HG3

Fazem parte do grupo os indivíduos H7-79 e H8-79. O indivíduo H8-79 foi escolhido para ter seus resultados individuais apresentados devido a sua boa dicção.

Quanto à avaliação perceptual, o indivíduo H7-79 não pronunciou corretamente as palavras “*Deslumbra*” e “*Araras*”, contidas nas frases 1 e 12, substituindo-as por “*Delubra*” e “*Arara*” devido à uma dislexia.

Neste grupo, nenhum dos indivíduos apresentou erros de processamento com a inserção de ruído e forte atenuação.

Na tabela 22 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo H8-79 e o grupo HG3. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento. Para este grupo, o codificador G.726 apresentou tempos de processamentos dentro da resolução observada para algumas frases.

Tabela 22 - Duração dos intervalos médios e dos arquivos do grupo HG3 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento				
		Intervalo Médio (s)	Duração (s)	G.723.1	G.726	G.728	G.729A	iLBC
H8-79	Frase 1	0,398	4,772	16,76%	0,00%	8,38%	4,19%	2,10%
	Frase 2	0,412	6,184	17,79%	3,23%	8,09%	3,23%	3,23%
	Frase 3	0,326	2,930	20,48%	0,00%	10,24%	3,41%	3,41%
	Frase 4	0,354	4,605	19,54%	0,00%	8,69%	2,17%	4,34%
	Frase 5	0,369	3,322	21,07%	0,00%	9,03%	3,01%	6,02%
	Frase 6	0,291	2,327	21,49%	0,00%	8,60%	4,30%	4,30%
	Frase 7	0,306	3,367	20,79%	0,00%	8,91%	2,97%	5,94%
	Frase 8	0,288	3,454	20,27%	0,00%	8,69%	2,90%	2,90%
	Frase 9	0,287	3,449	20,30%	0,00%	8,70%	2,90%	5,80%
	Frase 10	0,349	3,488	20,07%	0,00%	8,60%	5,73%	5,73%
	Frase 11	0,312	3,116	19,26%	0,00%	9,63%	3,21%	3,21%
	Frase 12	0,323	4,196	19,07%	0,00%	7,15%	2,38%	4,77%
HG3	Frase 1	0,423	5,075	17,68%	0,00%	7,91%	3,95%	2,91%
	Frase 2	0,386	5,792	18,15%	1,62%	8,67%	3,47%	3,47%
	Frase 3	0,313	2,816	21,34%	0,00%	10,67%	3,56%	3,56%
	Frase 4	0,359	4,672	19,27%	0,00%	8,56%	3,20%	4,28%
	Frase 5	0,434	3,905	20,56%	0,00%	7,86%	2,62%	5,24%
	Frase 6	0,303	2,425	22,64%	0,00%	10,24%	4,13%	4,13%
	Frase 7	0,332	3,650	20,57%	0,00%	8,27%	4,03%	5,51%
	Frase 8	0,299	3,588	19,54%	0,00%	8,37%	2,79%	4,13%
	Frase 9	0,274	3,287	21,35%	0,00%	9,15%	3,05%	6,10%
	Frase 10	0,354	3,537	19,79%	0,00%	8,48%	4,26%	5,66%
	Frase 11	0,306	3,060	19,62%	0,00%	9,81%	3,27%	3,27%
	Frase 12	0,337	4,382	19,38%	0,00%	7,95%	3,38%	4,57%

Nas tabelas 23 e 24 temos os erros de processamento das frequências fundamentais mínimas, médias e máximas para cada frase para o indivíduo H8-79 e médios do grupo HG3.

Tabela 23 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo H8-79 e o grupo HG3.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M3-48	Frase 1	-0,82%	0,08%	1,36%	1,80%	-0,62%	-4,79%	-0,25%	-0,34%	-0,89%
	Frase 2	0,27%	0,21%	0,97%	6,43%	1,67%	0,09%	0,28%	0,12%	0,46%
	Frase 3	0,61%	0,17%	-0,33%	11,72%	2,32%	-1,31%	0,85%	0,81%	1,77%
	Frase 4	1,92%	0,89%	0,53%	12,94%	2,78%	-1,57%	1,45%	1,02%	1,41%
	Frase 5	0,02%	0,11%	1,39%	-0,72%	-0,16%	0,51%	0,06%	0,11%	0,09%
	Frase 6	0,53%	0,00%	-0,16%	6,43%	1,73%	2,21%	0,16%	-0,09%	0,19%
	Frase 7	1,76%	0,58%	0,61%	4,10%	2,53%	-1,93%	0,61%	0,69%	0,75%
	Frase 8	0,21%	0,24%	0,89%	4,78%	2,13%	0,43%	-0,03%	-0,01%	-0,03%
	Frase 9	5,08%	0,70%	0,22%	8,60%	3,21%	1,05%	2,76%	0,16%	0,73%
	Frase 10	0,86%	0,19%	0,16%	10,24%	3,82%	1,19%	0,31%	0,20%	2,01%
	Frase 11	0,63%	0,00%	-0,13%	1,26%	0,31%	-1,94%	-1,15%	-0,24%	-0,19%
	Frase 12	1,05%	0,02%	-0,94%	8,51%	1,60%	-0,69%	-0,89%	-0,32%	0,02%
MG2	Frase 1	-1,19%	-1,02%	-2,27%	-1,23%	-0,61%	8,35%	-1,84%	-0,96%	10,74%
	Frase 2	0,87%	-0,38%	0,94%	3,33%	0,16%	0,33%	0,25%	0,14%	1,05%
	Frase 3	0,34%	-0,87%	-4,26%	6,27%	0,54%	-4,34%	0,41%	0,39%	0,83%
	Frase 4	1,47%	0,48%	0,92%	6,74%	1,46%	-1,07%	-0,14%	0,48%	0,69%
	Frase 5	-0,20%	0,17%	1,62%	3,85%	0,18%	0,59%	4,39%	0,41%	-1,02%
	Frase 6	0,81%	0,24%	0,56%	1,97%	0,81%	1,00%	-0,25%	-0,02%	0,19%
	Frase 7	1,24%	0,35%	-0,57%	1,95%	1,19%	-1,71%	0,04%	0,36%	0,77%
	Frase 8	-0,40%	0,11%	0,28%	2,66%	2,99%	13,29%	-0,33%	-0,05%	-0,01%
	Frase 9	3,00%	0,37%	0,06%	3,28%	1,54%	0,55%	1,90%	0,80%	8,37%
	Frase 10	0,73%	-0,52%	0,93%	5,33%	1,81%	0,39%	0,20%	-0,06%	0,99%
	Frase 11	0,48%	-0,18%	0,09%	-1,34%	1,31%	8,74%	-0,75%	-0,13%	0,01%
	Frase 12	0,41%	-0,12%	-0,93%	3,44%	0,67%	-0,56%	-2,49%	-0,32%	-0,05%

Tabela 24 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H8-79 e o grupo HG3.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M3-48	Frase 1	-1,15%	0,79%	-0,62%	1,18%	-1,01%	-2,66%	0,34%	0,92%	1,55%	-0,96%	-0,10%	-0,89%
	Frase 2	0,11%	0,11%	-0,02%	1,16%	0,62%	1,04%	2,36%	0,67%	0,35%	2,72%	0,54%	-0,18%
	Frase 3	-0,25%	0,11%	0,73%	1,94%	0,80%	1,27%	2,27%	1,13%	2,45%	0,89%	0,49%	2,01%
	Frase 4	0,10%	0,19%	0,42%	5,29%	1,67%	0,29%	7,61%	2,54%	1,51%	3,87%	1,03%	0,01%
	Frase 5	-0,11%	0,06%	0,83%	0,18%	-0,01%	-0,28%	-0,78%	-1,45%	1,05%	-4,50%	-0,51%	0,31%
	Frase 6	-1,25%	-0,12%	-0,08%	0,90%	0,24%	0,75%	1,70%	0,58%	1,74%	-1,40%	-0,29%	1,56%
	Frase 7	-0,07%	0,41%	0,83%	4,86%	1,72%	-0,63%	2,93%	0,76%	-0,13%	0,22%	0,33%	-0,07%
	Frase 8	-0,38%	-0,19%	0,07%	1,20%	0,35%	0,40%	-3,14%	-0,04%	0,81%	-0,03%	0,01%	0,26%
	Frase 9	3,23%	0,03%	0,32%	1,50%	0,93%	1,08%	5,15%	0,82%	1,26%	2,19%	0,49%	0,38%
	Frase 10	-0,54%	0,23%	2,46%	0,86%	0,31%	0,48%	2,42%	0,17%	1,18%	0,23%	0,36%	3,55%
	Frase 11	-1,20%	-0,24%	-0,23%	-0,48%	-0,10%	-0,39%	-0,18%	1,68%	12,97%	0,00%	0,33%	0,35%
	Frase 12	-0,54%	0,45%	9,57%	2,42%	0,20%	-0,71%	0,46%	-0,40%	-0,74%	0,46%	1,03%	10,45%
MG2	Frase 1	-1,65%	0,31%	-0,19%	1,75%	-1,41%	-4,41%	-1,15%	-0,53%	-2,14%	1,37%	-0,88%	-3,46%
	Frase 2	0,73%	0,19%	0,93%	0,32%	-0,68%	-0,98%	2,14%	-0,69%	-2,70%	1,75%	-0,66%	-1,66%
	Frase 3	-1,12%	-0,93%	-3,99%	1,35%	-0,44%	-3,25%	0,53%	-0,34%	-2,52%	1,48%	-0,62%	-3,11%
	Frase 4	0,70%	0,30%	0,67%	3,56%	0,75%	-0,53%	5,17%	1,30%	0,36%	3,26%	0,55%	-0,56%
	Frase 5	3,83%	0,31%	1,10%	0,14%	-0,96%	-2,76%	-0,12%	-0,63%	0,54%	-2,08%	-0,87%	-1,99%
	Frase 6	-0,91%	-0,02%	0,06%	-0,76%	0,22%	0,81%	-1,52%	0,28%	1,03%	-0,42%	-0,15%	0,95%
	Frase 7	0,06%	0,33%	1,01%	2,29%	0,94%	-0,36%	1,93%	0,42%	-1,29%	0,34%	0,15%	-0,99%
	Frase 8	-0,35%	-0,07%	-0,02%	0,47%	0,24%	0,72%	-1,38%	1,29%	14,46%	-0,38%	-0,13%	-0,09%
	Frase 9	1,78%	0,04%	0,07%	-0,24%	0,39%	0,54%	3,14%	0,51%	0,44%	1,71%	0,77%	3,66%
	Frase 10	-0,24%	0,02%	1,23%	0,70%	-0,83%	0,97%	1,73%	-0,40%	-2,25%	0,30%	0,15%	1,54%
	Frase 11	-2,89%	-0,38%	-0,02%	1,79%	-0,98%	-3,26%	0,27%	-0,17%	3,63%	-2,35%	2,18%	12,28%
	Frase 12	-0,45%	0,15%	4,73%	-0,72%	-0,61%	-1,16%	-1,28%	-0,39%	-0,33%	-1,00%	0,38%	4,98%

Nas tabelas 25 e 26 temos os erros de processamento das formantes para cada frase para o indivíduo H8-79 e médios do grupo HG3. Percebe-se que os maiores erros se encontram na primeira e segunda formantes.

Tabela 25 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo H8-79 e o grupo HG3.

		G.722			G.723.1			G.726		
		Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M3-48	Frase 1	2,93%	2,11%	3,24%	-2,63%	2,47%	6,16%	3,03%	5,61%	6,75%
	Frase 2	5,99%	3,57%	3,25%	-5,25%	2,00%	4,53%	6,81%	6,10%	6,00%
	Frase 3	-0,26%	1,28%	3,73%	-2,59%	2,87%	4,52%	2,96%	3,61%	6,45%
	Frase 4	3,50%	2,96%	4,35%	-2,45%	2,83%	4,68%	3,68%	5,28%	7,30%
	Frase 5	4,62%	2,63%	3,24%	-4,60%	-1,57%	3,14%	10,39%	5,52%	4,95%
	Frase 6	12,91%	4,70%	5,04%	-12,51%	0,30%	3,36%	15,18%	7,06%	8,50%
	Frase 7	3,79%	2,41%	4,19%	-2,59%	-2,50%	0,75%	5,73%	6,37%	5,23%
	Frase 8	3,46%	2,73%	4,38%	-2,59%	0,51%	1,88%	5,05%	7,55%	8,39%
	Frase 9	6,74%	2,05%	5,28%	-4,37%	-1,54%	2,75%	5,46%	5,78%	8,09%
	Frase 10	20,66%	6,28%	5,54%	-1,86%	-1,31%	1,37%	24,94%	9,12%	7,22%
	Frase 11	4,84%	2,95%	3,82%	-0,79%	1,34%	-0,21%	12,13%	4,95%	4,16%
	Frase 12	6,43%	2,56%	4,51%	2,05%	0,29%	1,37%	11,64%	6,22%	7,13%
MG2	Frase 1	3,03%	-0,02%	2,38%	-0,28%	2,49%	4,35%	4,29%	3,61%	4,73%
	Frase 2	3,38%	1,50%	2,31%	-3,89%	3,16%	3,44%	3,83%	4,59%	4,47%
	Frase 3	-0,09%	0,31%	2,47%	-4,33%	1,25%	1,88%	1,52%	2,59%	4,39%
	Frase 4	1,24%	0,21%	2,02%	-5,60%	1,94%	1,82%	1,91%	3,96%	4,60%
	Frase 5	2,15%	1,14%	2,35%	-4,00%	1,18%	1,86%	6,13%	4,12%	3,56%
	Frase 6	6,41%	1,58%	2,81%	-10,77%	-0,07%	1,07%	8,15%	4,38%	5,05%
	Frase 7	2,02%	0,52%	2,47%	-4,17%	-2,33%	-1,33%	2,99%	3,52%	3,40%
	Frase 8	2,63%	0,63%	2,67%	-5,17%	-1,50%	1,62%	2,24%	5,54%	5,57%
	Frase 9	3,35%	0,07%	2,65%	-4,26%	-1,09%	-1,08%	1,84%	3,58%	4,47%
	Frase 10	10,72%	2,40%	3,20%	-5,12%	-0,60%	-0,35%	13,09%	4,39%	4,64%
	Frase 11	2,63%	0,39%	2,15%	-2,06%	3,38%	-2,23%	6,55%	3,22%	2,36%
	Frase 12	3,26%	0,26%	1,91%	6,47%	1,28%	-0,81%	5,05%	3,12%	4,45%

Tabela 26 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo H8-79 e o grupo HG3.

		G.728			G.729A			iLBC			Speex		
		Formantes			Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M3-48	Frase 1	1,10%	6,63%	7,46%	-8,07%	-4,50%	-1,60%	-4,11%	-3,91%	-1,41%	-0,59%	3,07%	7,31%
	Frase 2	4,46%	7,71%	6,47%	-9,35%	-6,38%	-0,12%	-4,79%	-5,01%	-1,67%	3,06%	4,69%	5,84%
	Frase 3	3,41%	4,98%	6,42%	-6,52%	-3,60%	-1,63%	-1,80%	-3,83%	-2,11%	1,37%	2,44%	6,42%
	Frase 4	1,82%	6,51%	7,49%	-8,91%	-5,58%	-1,59%	-6,05%	-6,06%	-2,57%	4,43%	3,52%	6,98%
	Frase 5	2,85%	4,81%	4,97%	-9,60%	-4,39%	-3,37%	-6,73%	-4,36%	-2,45%	1,51%	2,76%	5,02%
	Frase 6	10,92%	6,93%	7,47%	-20,90%	-5,92%	-3,25%	-17,22%	-5,49%	-1,81%	4,05%	5,79%	7,28%
	Frase 7	1,70%	4,55%	4,16%	-9,70%	-7,28%	-3,71%	-3,22%	-3,98%	-2,86%	-0,02%	1,53%	3,88%
	Frase 8	1,37%	7,63%	8,05%	-10,76%	-10,17%	-2,84%	-6,60%	-5,44%	-2,27%	0,75%	4,58%	7,33%
	Frase 9	5,71%	5,29%	6,87%	-13,86%	-7,55%	-2,71%	-6,02%	-4,26%	-1,84%	1,70%	1,81%	6,28%
	Frase 10	15,65%	7,24%	6,47%	-13,47%	-7,20%	-2,32%	-4,61%	-2,88%	-1,78%	5,17%	2,61%	5,37%
	Frase 11	3,96%	4,38%	3,50%	-6,38%	-3,31%	-3,79%	-5,74%	-3,59%	-2,04%	1,69%	2,63%	3,31%
	Frase 12	12,84%	6,29%	5,50%	-4,22%	-4,43%	-3,48%	-1,97%	-2,12%	-2,64%	9,56%	3,17%	5,42%
MG2	Frase 1	-0,31%	4,45%	5,28%	-8,15%	-5,32%	-2,22%	-2,78%	-3,83%	-2,03%	0,23%	2,56%	5,59%
	Frase 2	1,81%	6,45%	4,70%	-7,96%	-5,22%	-1,29%	-3,78%	-4,76%	-2,23%	2,18%	3,59%	4,27%
	Frase 3	1,69%	3,62%	4,09%	-7,44%	-3,96%	-3,22%	-3,26%	-4,06%	-2,96%	0,45%	1,68%	4,45%
	Frase 4	0,47%	5,19%	3,97%	-9,24%	-6,08%	-3,18%	-5,74%	-6,27%	-3,50%	1,22%	1,82%	4,32%
	Frase 5	0,08%	3,75%	3,35%	-7,84%	-3,57%	-2,81%	-4,43%	-3,72%	-2,53%	1,78%	2,43%	3,79%
	Frase 6	5,53%	4,31%	4,19%	-14,59%	-5,19%	-3,87%	-10,02%	-4,49%	-2,22%	2,14%	2,70%	4,29%
	Frase 7	-0,01%	2,62%	2,14%	-9,07%	-5,51%	-5,00%	-3,95%	-3,67%	-3,34%	-0,42%	0,21%	2,37%
	Frase 8	0,07%	6,16%	5,37%	-9,16%	-9,05%	-3,61%	-4,99%	-6,32%	-3,23%	0,83%	2,97%	5,24%
	Frase 9	4,27%	3,20%	2,77%	-10,36%	-5,82%	-4,54%	-4,24%	-3,56%	-2,91%	1,00%	0,80%	3,22%
	Frase 10	8,28%	4,05%	3,68%	-12,49%	-6,80%	-3,95%	-5,29%	-3,19%	-2,51%	2,72%	1,34%	3,53%
	Frase 11	2,43%	5,11%	1,64%	-4,79%	-0,94%	-5,34%	-3,57%	-2,29%	-2,96%	1,21%	2,39%	1,52%
	Frase 12	6,33%	3,45%	2,66%	-6,42%	-5,61%	-4,35%	-2,41%	-3,51%	-3,17%	7,04%	0,76%	3,14%

Nas figuras de 60 a 77 temos os resultados do indivíduo H8-79 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas das frequências em Hertz por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

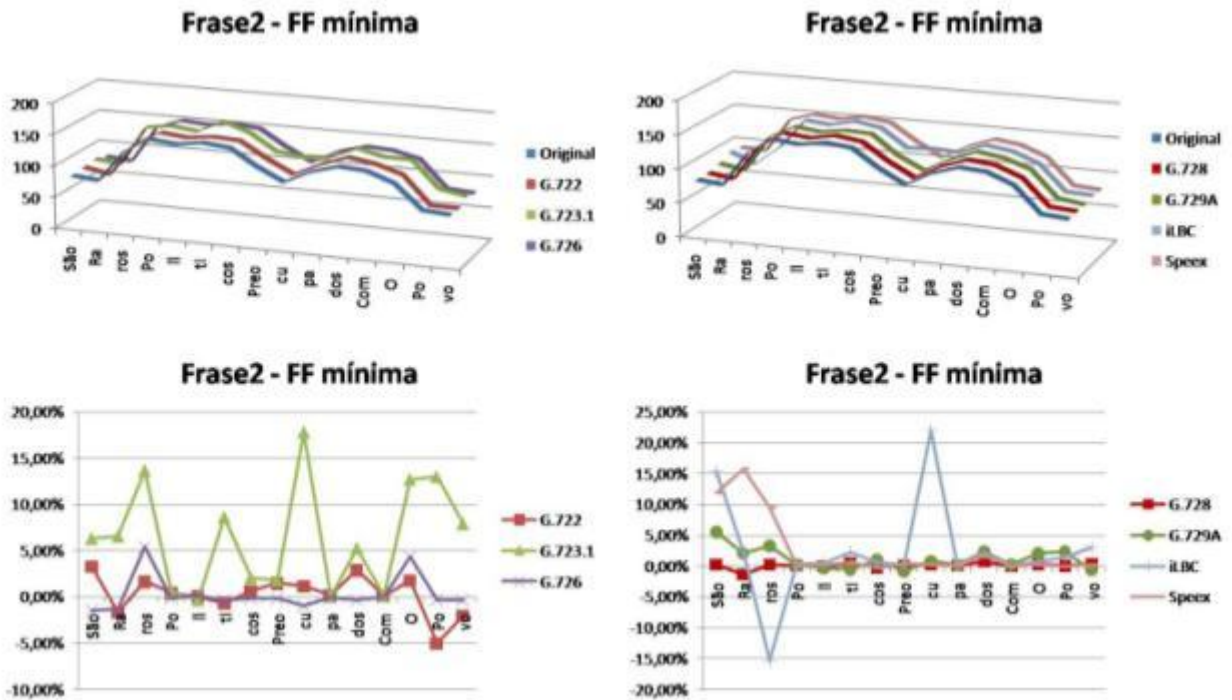


Figura 60 - Variações da frequência fundamental mínima da frase 2 do indivíduo H8-79 do grupo HG3.

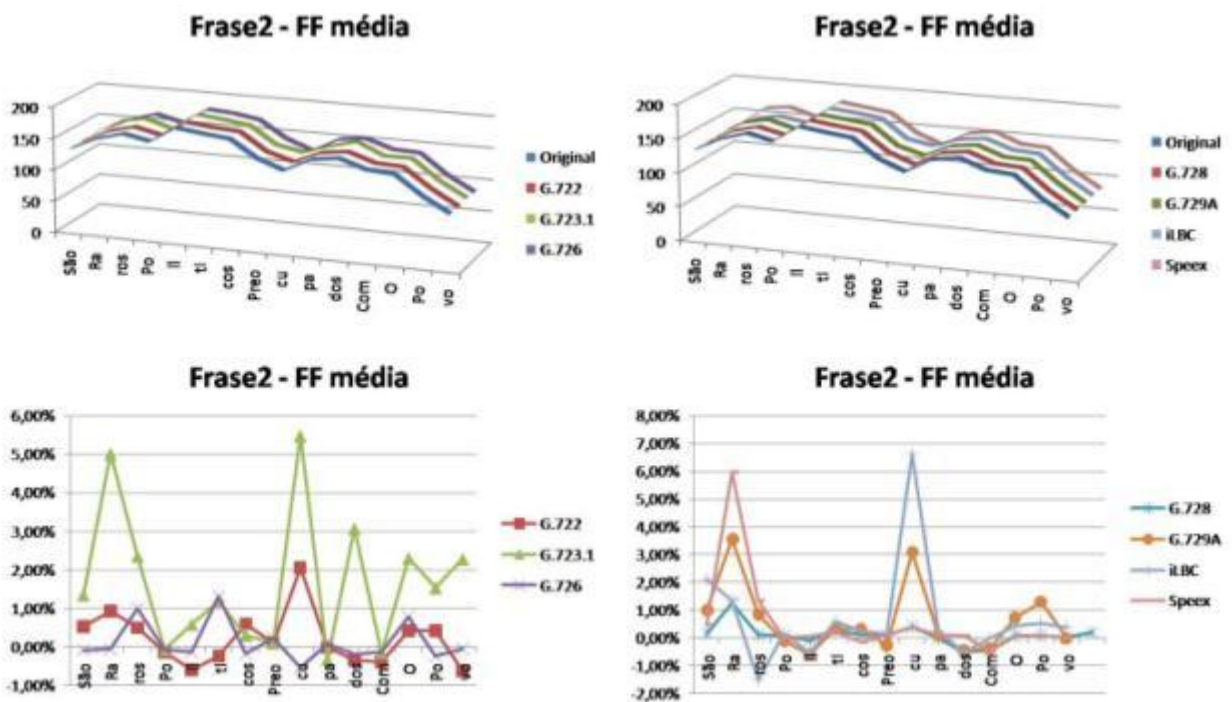


Figura 61 - Variações da frequência fundamental média da frase 2 do indivíduo H8-79 do grupo HG3.

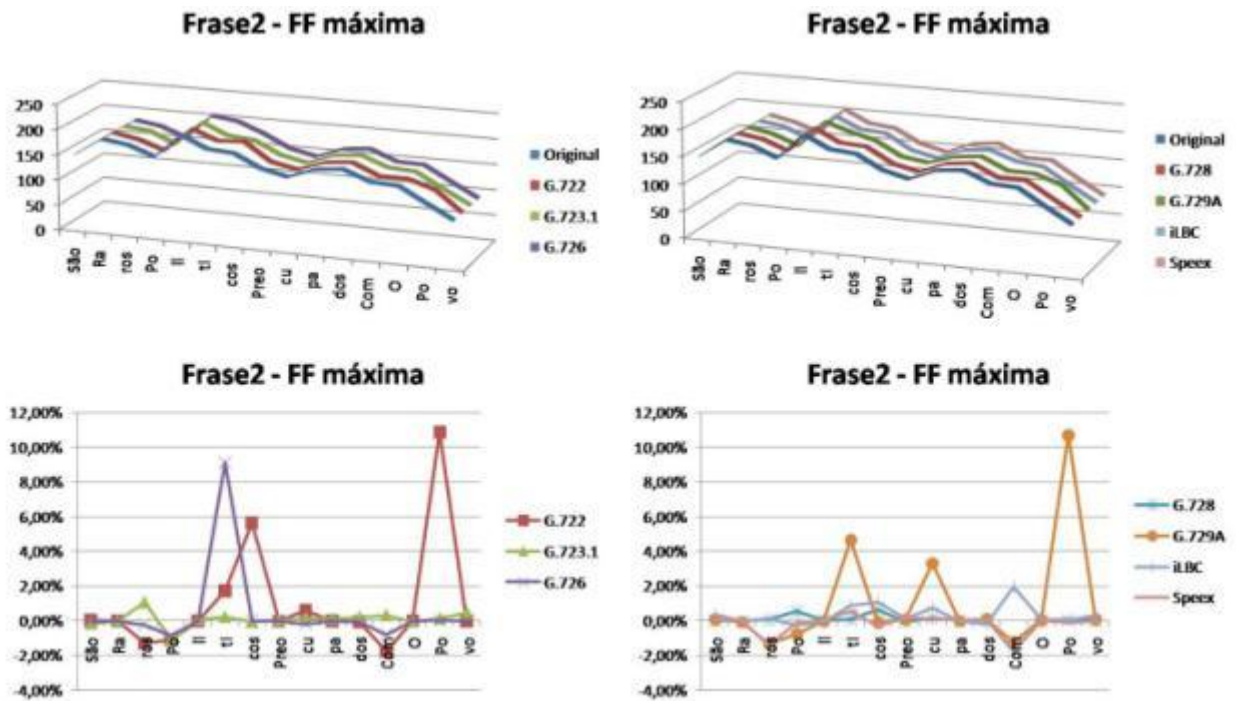


Figura 62 - Variações da frequência fundamental máxima da frase 2 do indivíduo H8-79 do grupo HG3.

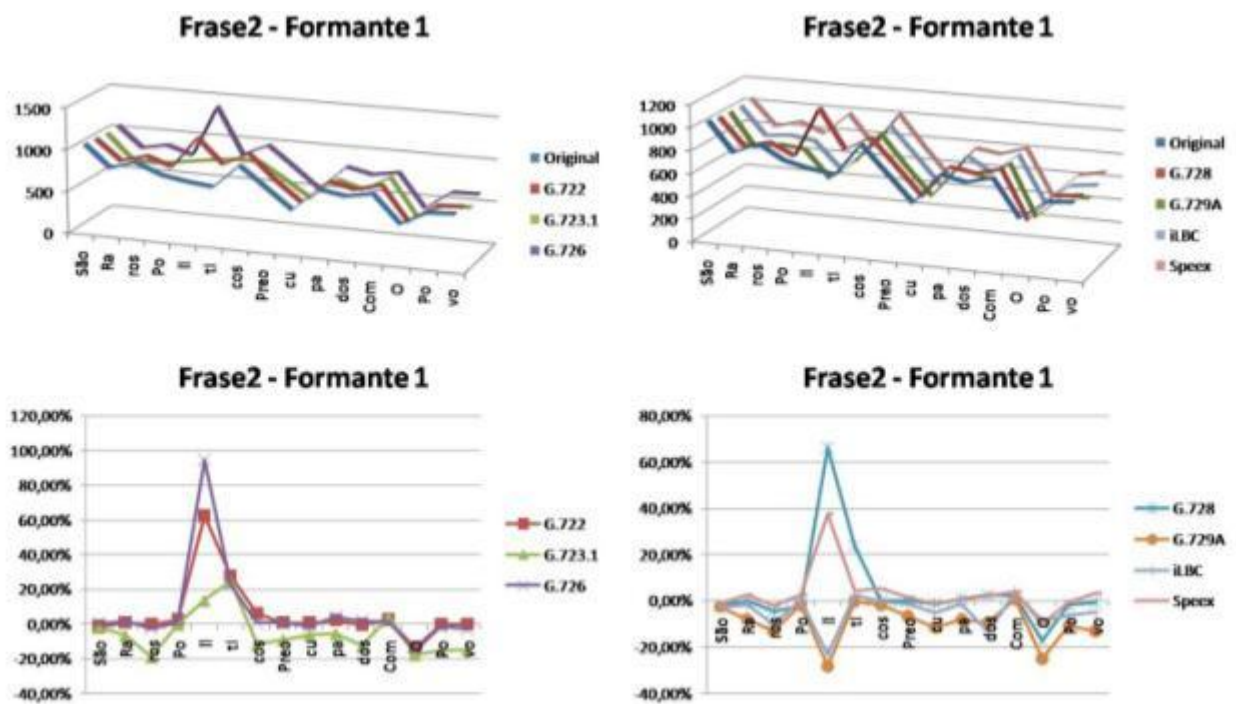


Figura 63 - Variações da frequência da primeira formante da frase 2 do indivíduo H8-79 do grupo HG3.

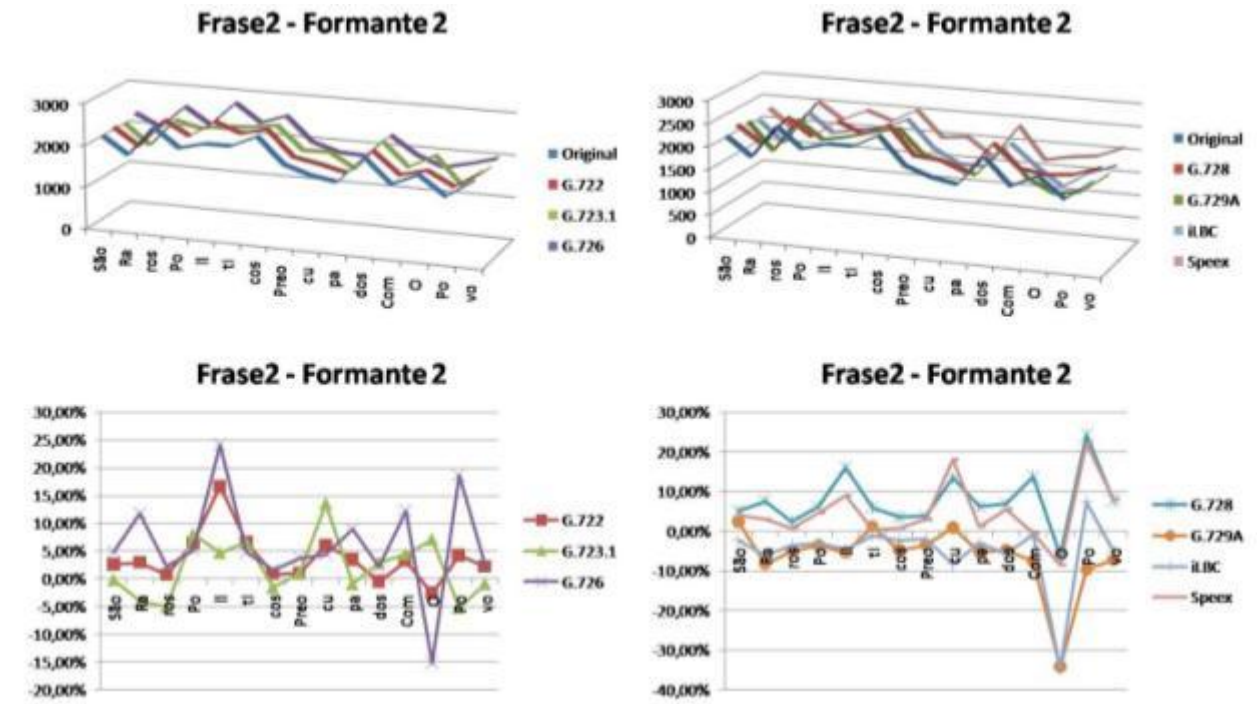


Figura 64 - Variações da frequência da segunda formante da frase 2 do indivíduo H8-79 do grupo HG3.

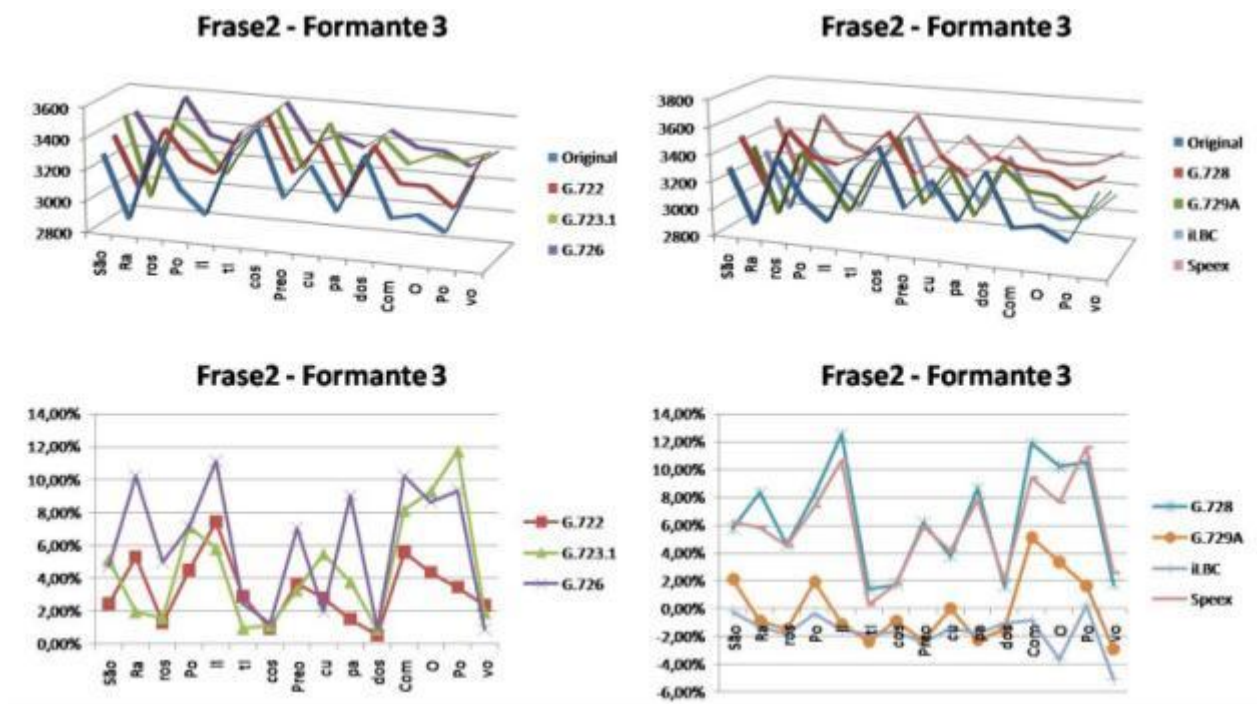


Figura 65 - Variações da frequência da terceira formante da frase 2 do indivíduo H8-79 do grupo HG3.

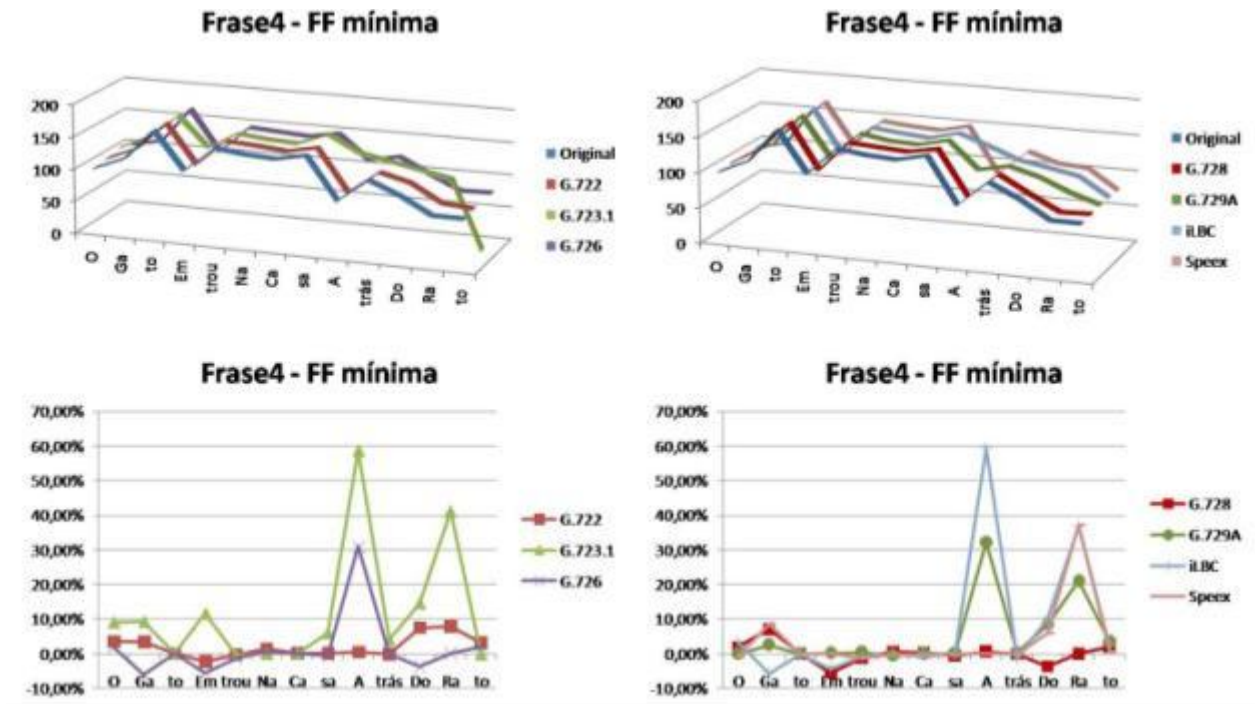


Figura 66 - Variações da frequência fundamental mínima da frase 4 do indivíduo H8-79 do grupo HG3.

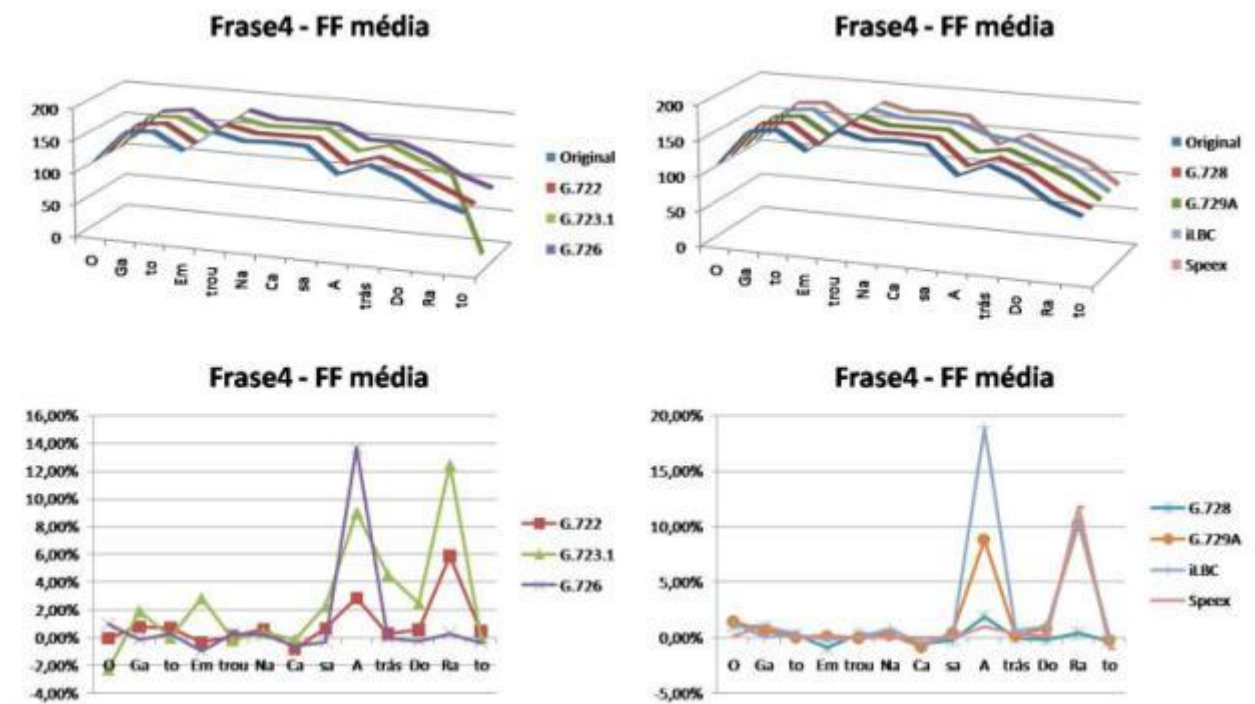


Figura 67 - Variações da frequência fundamental média da frase 4 do indivíduo H8-79 do grupo HG3.

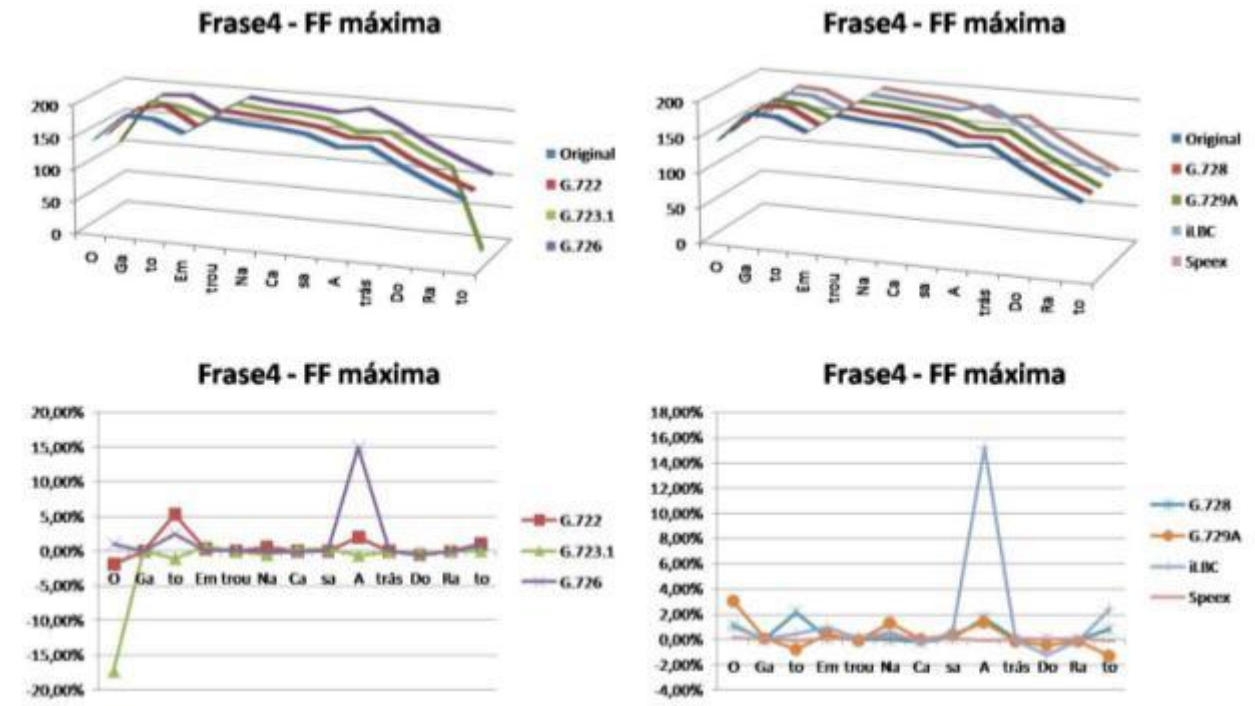


Figura 68 - Variações da frequência fundamental máxima da frase 4 do indivíduo H8-79 do grupo HG3.

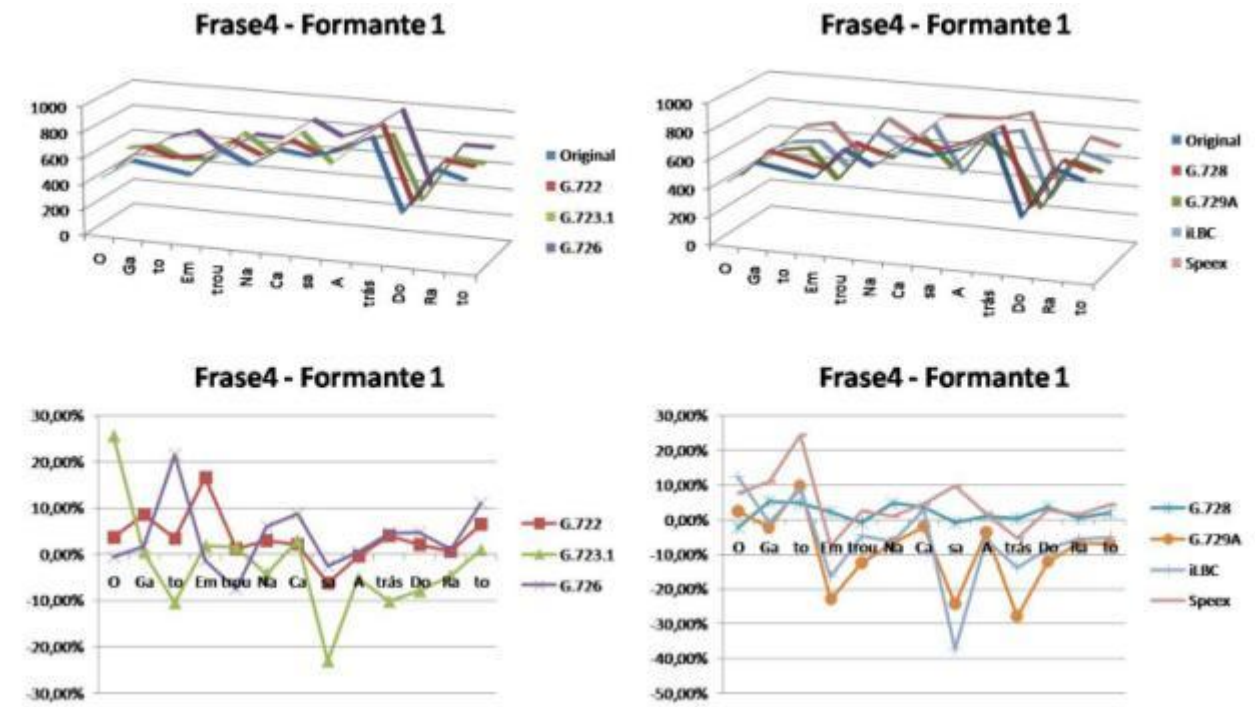


Figura 69 - Variações da frequência da primeira formante da frase 4 do indivíduo H8-79 do grupo HG3.

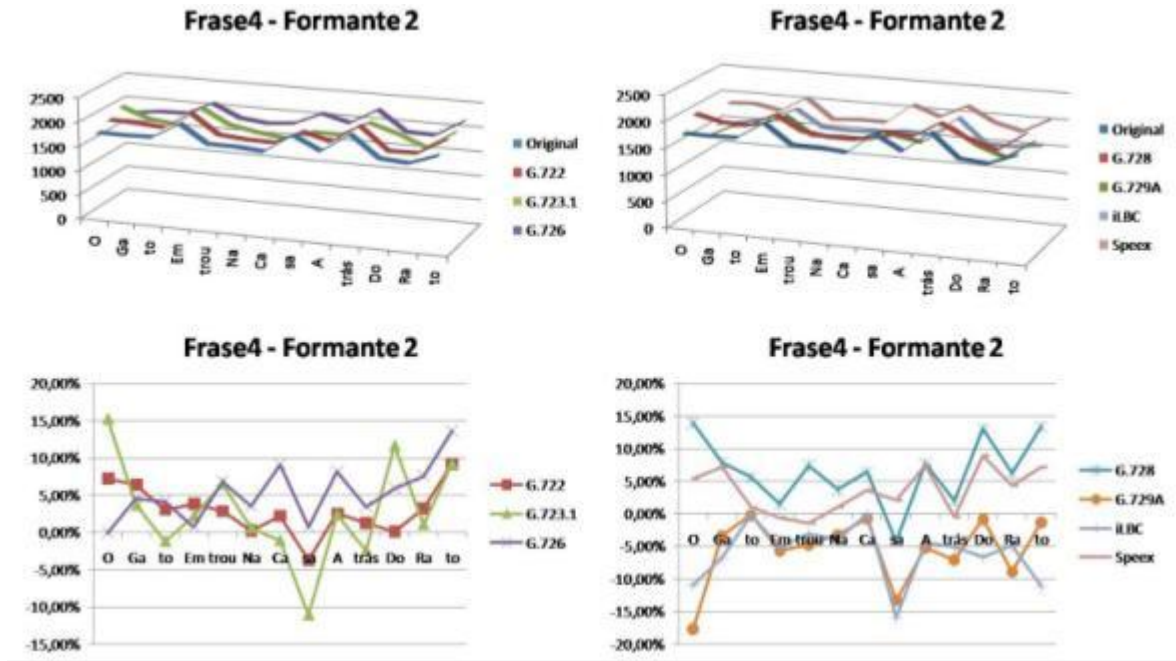


Figura 70 - Variações da freqüência da segunda formante da frase 4 do indivíduo H8-79 do grupo HG3.

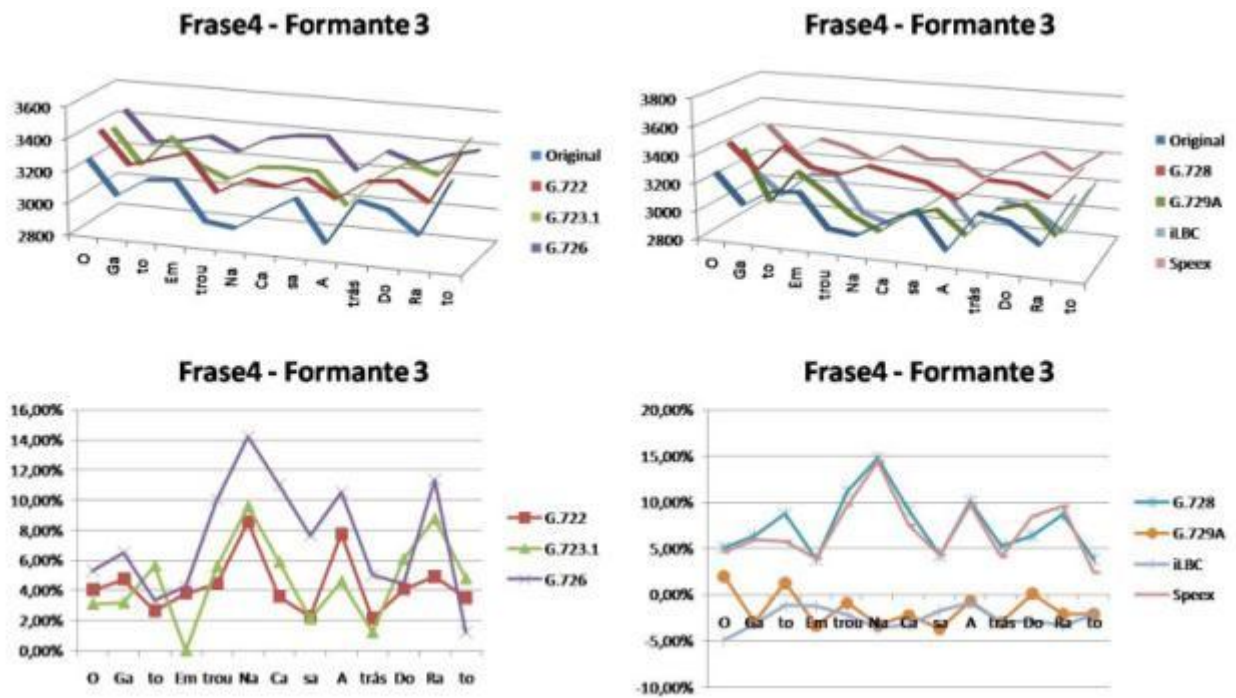


Figura 71 - Variações da freqüência da terceira formante da frase 4 do indivíduo H8-79 do grupo HG3.

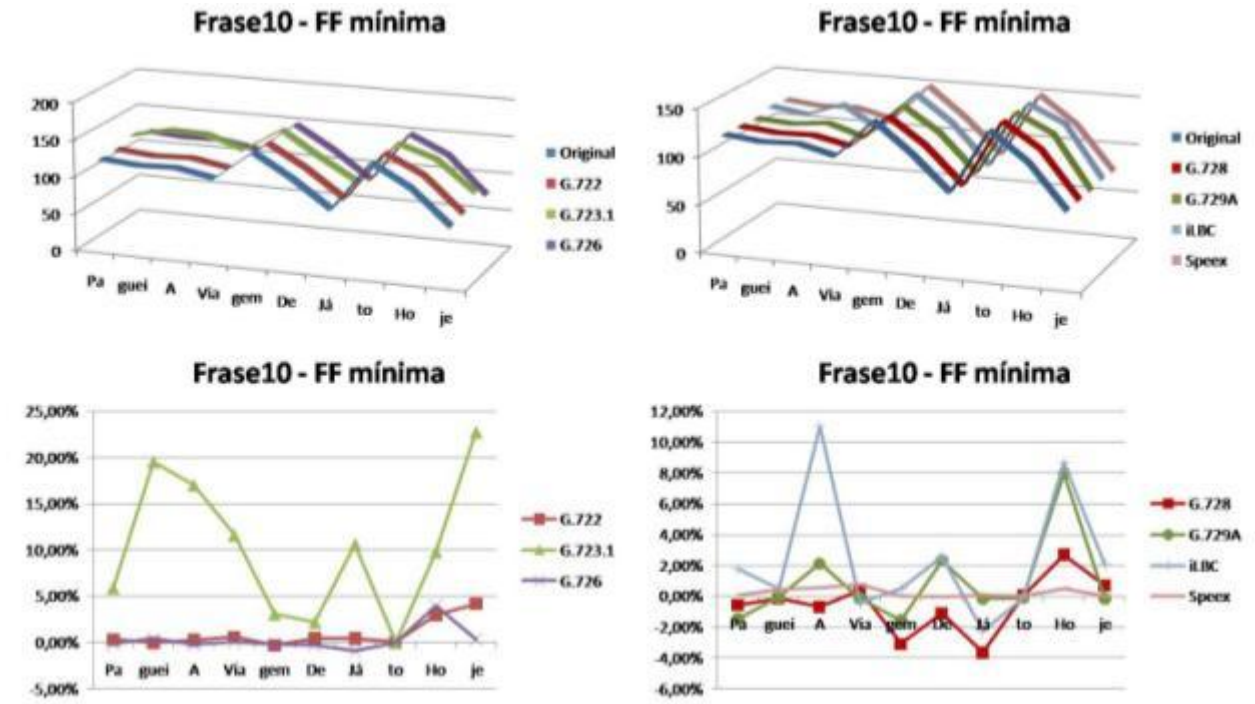


Figura 72 - Variações da frequência fundamental mínima da frase 10 do indivíduo H8-79 do grupo HG3.

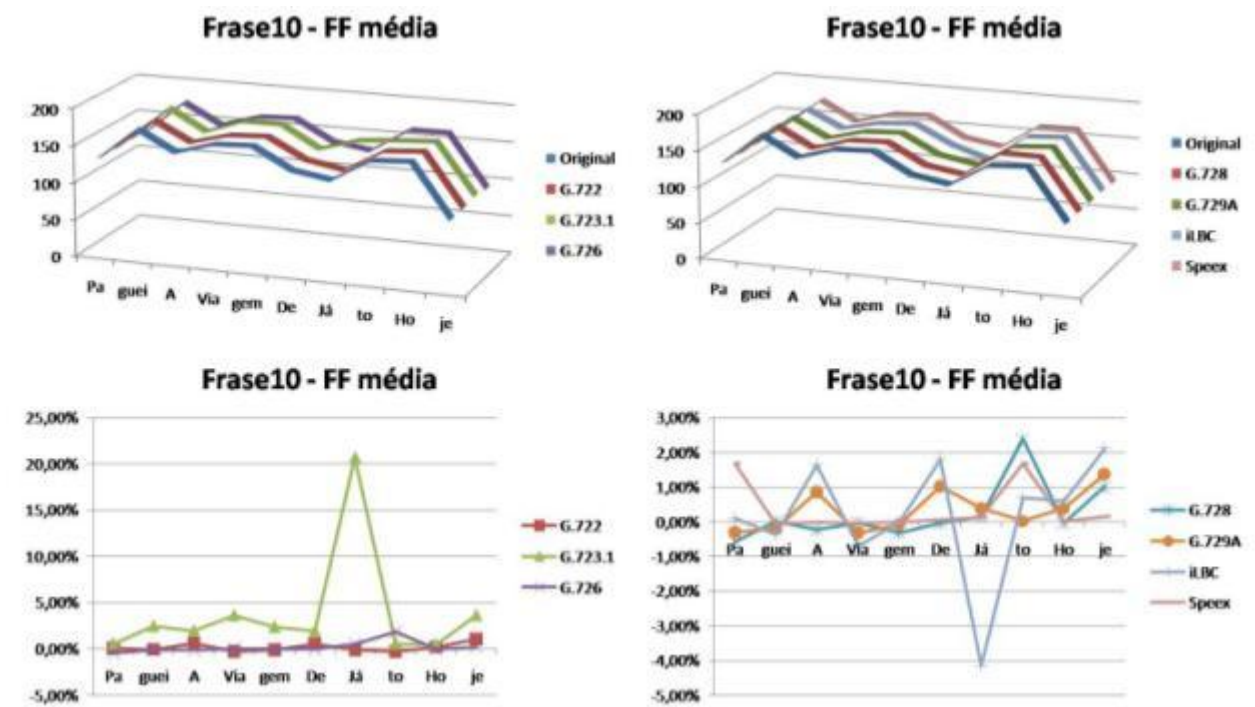


Figura 73 - Variações da frequência fundamental média da frase 10 do indivíduo H8-79 do grupo HG3.

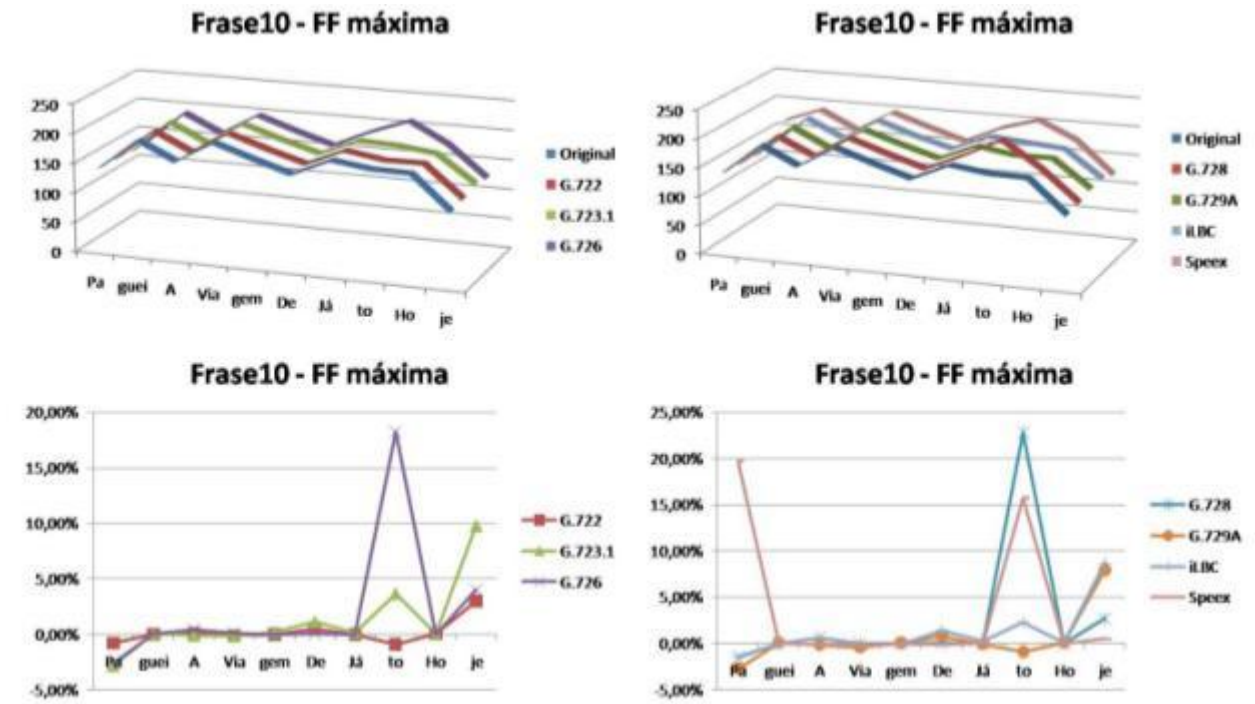


Figura 74 - Variações da freqüência fundamental máxima da frase 10 do indivíduo H8-79 do grupo HG3.

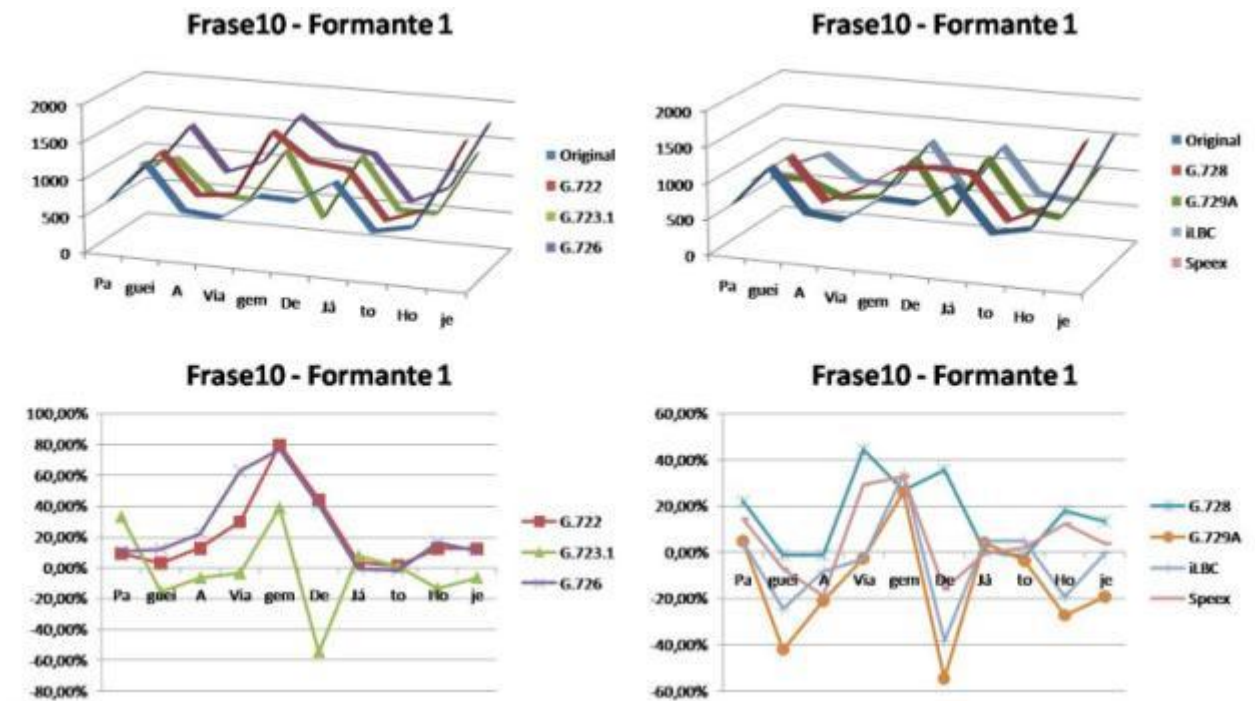


Figura 75 - Variações da freqüência da primeira formante da frase 10 do indivíduo H8-79 do grupo HG3.

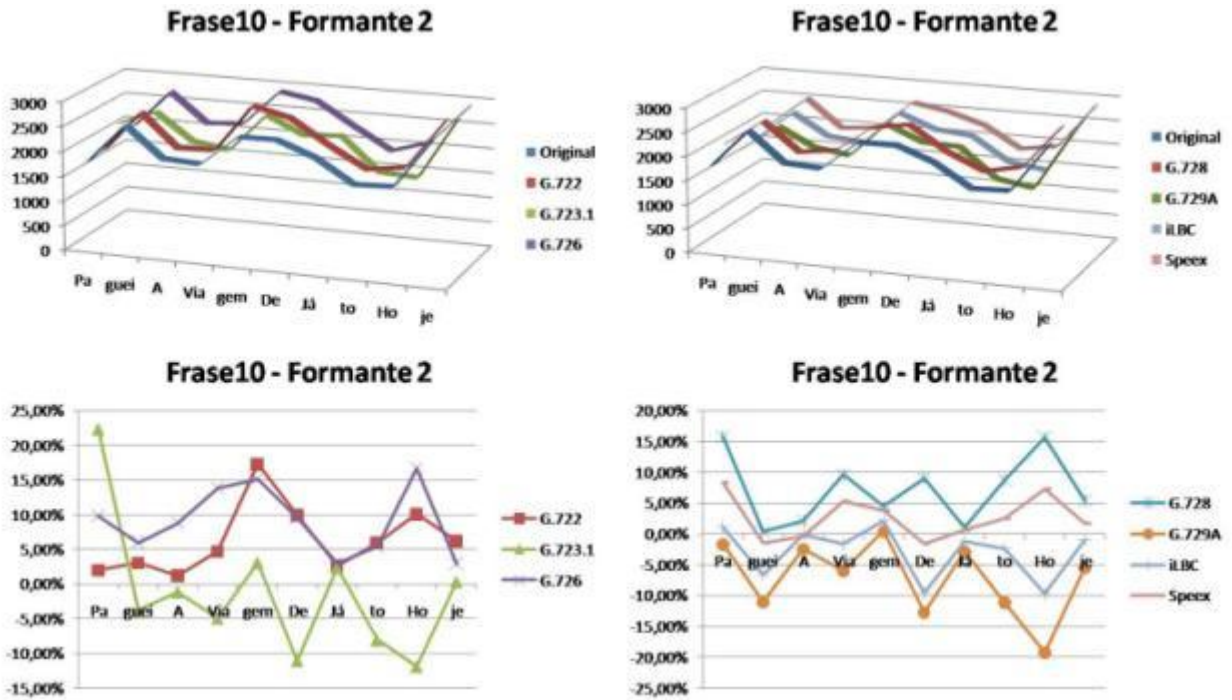


Figura 76 - Variações da frequência da segunda formante da frase 10 do indivíduo H8-79 do grupo HG3.

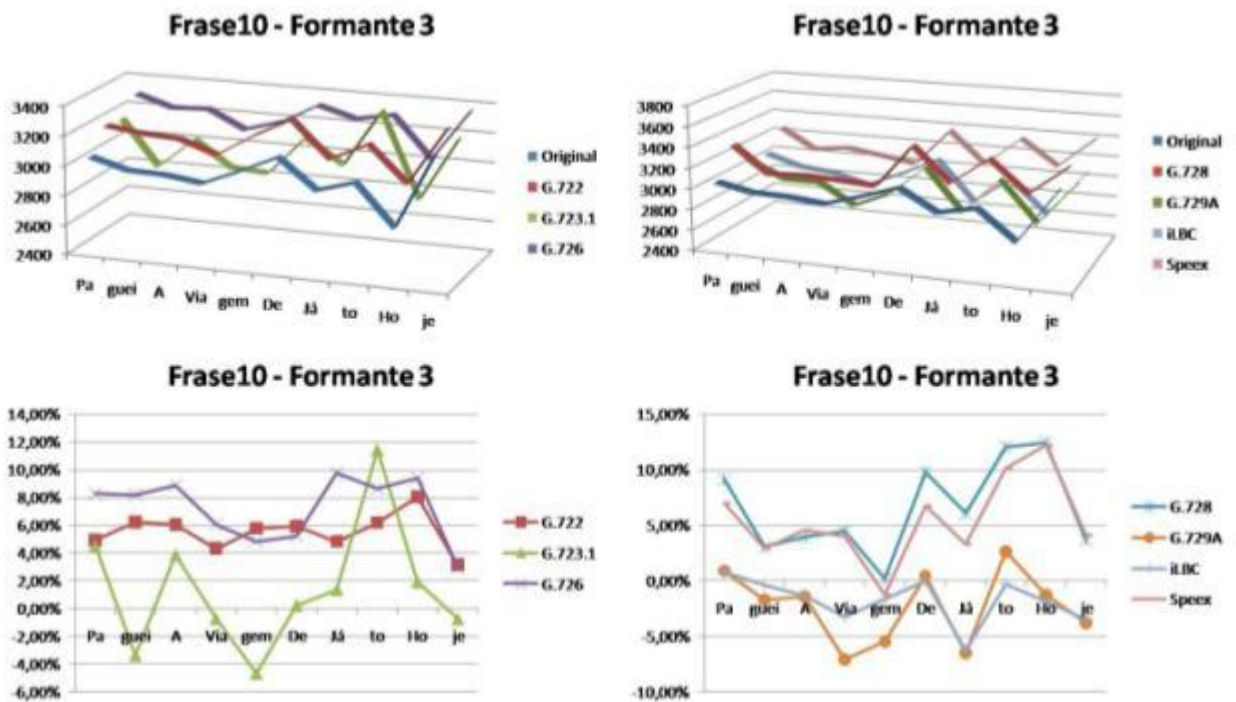


Figura 77 - Variações da frequência da terceira formante da frase 10 do indivíduo H8-79 do grupo HG3.

4.4.4 – MG1

Fazem parte do grupo os indivíduos M4-25 e M8-17. O indivíduo M8-17 foi escolhido para ter seus resultados individuais por ter a idade mais próxima do indivíduo escolhido para o grupo HG1.

Neste grupo, o indivíduo M8-17 apresentou erros de processamento com a inserção de ruído e forte atenuação para o codificador G.723.1 nas frases 6, 9, 10 e 12.

Na tabela 27 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo M8-17 e o grupo MG1. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento.

Tabela 27 - Duração dos intervalos médios e dos arquivos do grupo MG1 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento			
		Intervalo Médio (s)	Duração (s)	G.723.1	G.728	G.729A	iLBC
M8-17	Frase 1	0,192	2,303	21,71%	13,02%	4,34%	4,34%
	Frase 2	0,169	2,538	23,64%	11,82%	3,94%	3,94%
	Frase 3	0,209	1,883	21,25%	10,62%	5,31%	5,31%
	Frase 4	0,176	2,294	21,79%	8,72%	4,36%	4,36%
	Frase 5	0,216	1,943	20,59%	10,29%	5,15%	5,15%
	Frase 6	0,187	1,492	20,10%	6,70%	0,00%	6,70%
	Frase 7	0,175	1,924	20,80%	10,40%	0,00%	5,20%
	Frase 8	0,169	2,026	19,75%	9,87%	0,00%	4,94%
	Frase 9	0,176	2,111	23,69%	9,47%	4,74%	4,74%
	Frase 10	0,204	2,037	24,55%	9,82%	0,00%	4,91%
	Frase 11	0,197	1,966	20,35%	10,18%	5,09%	5,09%
	Frase 12	0,181	2,357	21,21%	12,73%	4,24%	4,24%
Grupo	Frase 1	0,213	2,556	21,54%	11,85%	3,95%	3,95%
	Frase 2	0,187	2,803	21,60%	10,80%	3,60%	5,23%
	Frase 3	0,211	1,902	23,64%	10,52%	2,66%	5,26%
	Frase 4	0,182	2,368	23,19%	8,45%	4,23%	4,23%
	Frase 5	0,219	1,973	20,28%	10,14%	2,57%	5,07%
	Frase 6	0,204	1,631	18,53%	9,00%	0,00%	6,18%
	Frase 7	0,196	2,152	20,90%	11,50%	2,10%	4,70%
	Frase 8	0,184	2,213	22,37%	9,10%	2,08%	4,55%
	Frase 9	0,185	2,225	22,53%	9,01%	4,51%	4,51%
	Frase 10	0,221	2,207	22,79%	11,22%	2,10%	4,56%
	Frase 11	0,209	2,094	21,43%	9,59%	4,79%	4,79%
	Frase 12	0,190	2,469	22,23%	10,24%	4,06%	6,00%

Nas tabelas 28 e 29 temos os erros de processamento das frequências fundamentais mínimas, médias e máximas para cada frase para o indivíduo M8-17 e médios do grupo MG1.

Tabela 28 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo M8-17 e o grupo MG1.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M8-17	Frase 1	1,48%	0,26%	-0,92%	0,53%	0,07%	-0,72%	0,01%	0,09%	0,07%
	Frase 2	0,97%	-2,00%	-5,36%	-0,01%	-1,66%	-3,97%	0,23%	-1,74%	-3,77%
	Frase 3	1,06%	0,28%	0,54%	-0,13%	-0,08%	0,26%	-0,18%	-0,10%	-0,31%
	Frase 4	1,76%	0,38%	-2,25%	0,33%	-0,07%	-1,39%	0,11%	0,18%	0,77%
	Frase 5	3,85%	0,59%	-3,85%	-0,38%	-0,27%	-3,41%	0,04%	0,20%	-0,47%
	Frase 6	0,79%	0,07%	-1,12%	3,04%	0,98%	-1,06%	0,85%	0,23%	-0,09%
	Frase 7	0,13%	-0,02%	0,11%	0,24%	0,37%	1,93%	-0,45%	-0,15%	-0,20%
	Frase 8	27,12%	2,59%	-0,46%	18,00%	0,64%	-1,69%	26,75%	0,80%	-1,49%
	Frase 9	0,75%	-0,84%	-2,06%	2,48%	0,28%	-0,69%	-0,34%	-0,18%	-1,25%
	Frase 10	0,59%	0,09%	0,12%	2,93%	0,90%	-4,72%	0,54%	0,29%	-0,20%
	Frase 11	1,09%	-0,84%	-5,44%	-0,01%	-1,10%	-4,61%	0,00%	-0,57%	-3,02%
	Frase 12	0,54%	-3,03%	-7,12%	2,70%	-2,16%	-7,88%	0,35%	-0,17%	-1,50%
MG1	Frase 1	0,70%	-0,55%	-2,92%	0,23%	-0,65%	-2,56%	-0,05%	0,11%	0,28%
	Frase 2	0,56%	-1,02%	-2,82%	0,07%	-0,85%	-2,12%	0,15%	-0,86%	-1,97%
	Frase 3	-0,13%	-0,82%	-2,19%	-0,73%	-1,00%	-2,34%	-0,08%	-0,05%	-1,58%
	Frase 4	1,23%	0,14%	-1,47%	0,51%	-0,09%	-1,04%	0,30%	0,09%	0,09%
	Frase 5	1,49%	0,38%	-1,14%	-1,03%	-0,06%	-0,92%	-0,01%	0,93%	2,04%
	Frase 6	0,34%	0,50%	1,24%	1,46%	0,96%	1,27%	0,29%	-0,03%	-0,66%
	Frase 7	0,17%	0,05%	-0,10%	0,22%	0,25%	0,81%	-0,16%	0,15%	0,79%
	Frase 8	11,18%	-0,56%	-1,80%	6,61%	-1,54%	-2,42%	13,40%	0,40%	-0,81%
	Frase 9	0,37%	-0,74%	-2,73%	1,23%	-0,18%	-2,05%	-0,13%	0,37%	2,81%
	Frase 10	0,21%	0,12%	0,28%	1,38%	0,52%	-2,14%	0,26%	0,15%	-0,08%
	Frase 11	1,69%	0,06%	-2,84%	1,14%	-0,07%	-2,43%	-0,05%	0,94%	4,65%
	Frase 12	0,07%	-1,77%	-0,13%	1,15%	-1,34%	-0,52%	0,14%	-0,47%	-0,77%

Tabela 29 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M8-17 e o grupo MG1.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M8-17	Frase 1	0,00%	0,11%	0,12%	0,40%	0,37%	0,04%	0,76%	0,25%	-0,05%	-0,16%	-0,19%	-0,13%
	Frase 2	-0,02%	-1,88%	-4,17%	0,09%	-1,92%	-4,17%	0,07%	-1,24%	-0,88%	0,09%	-1,78%	-3,89%
	Frase 3	0,31%	0,10%	0,48%	0,04%	0,59%	4,63%	0,17%	0,17%	2,32%	0,28%	0,21%	1,75%
	Frase 4	0,05%	0,68%	0,96%	0,17%	-0,21%	-1,79%	0,38%	-0,02%	-0,92%	-0,19%	-0,03%	0,55%
	Frase 5	-0,09%	0,18%	0,35%	-0,29%	0,01%	-1,19%	-0,03%	-0,22%	-0,35%	0,19%	-0,43%	-0,37%
	Frase 6	-0,14%	-0,02%	0,13%	0,07%	0,10%	0,50%	-0,09%	0,75%	3,75%	0,10%	0,16%	0,31%
	Frase 7	-0,02%	0,02%	0,33%	0,58%	0,11%	0,34%	0,32%	0,09%	-0,02%	0,57%	0,15%	0,50%
	Frase 8	26,80%	0,92%	-1,52%	17,88%	0,66%	-1,47%	27,20%	0,89%	-1,55%	27,16%	0,90%	-1,48%
	Frase 9	-0,94%	-0,08%	0,31%	-0,94%	-0,19%	-0,08%	-0,93%	-0,28%	0,43%	-0,28%	-0,10%	-0,02%
	Frase 10	0,03%	-0,06%	-0,25%	-0,04%	0,29%	1,47%	-0,16%	0,43%	1,91%	0,15%	0,21%	0,11%
	Frase 11	0,54%	-0,83%	-3,04%	0,41%	-0,91%	-3,37%	0,78%	-1,14%	-4,65%	0,98%	-0,33%	-2,96%
	Frase 12	0,10%	-0,63%	-1,52%	0,72%	-2,80%	-5,34%	0,53%	-3,05%	-7,04%	0,27%	-3,06%	-7,27%
MG1	Frase 1	-0,12%	-0,24%	-1,62%	1,86%	0,73%	-4,21%	0,01%	-0,75%	-3,57%	-0,05%	-0,95%	-3,93%
	Frase 2	0,01%	-0,94%	-2,42%	0,06%	-0,92%	-2,31%	-1,42%	-2,49%	-2,40%	0,07%	-0,84%	-1,78%
	Frase 3	0,26%	-0,07%	-1,18%	-1,34%	-2,20%	-3,94%	-0,61%	-2,12%	-4,79%	-0,50%	-2,18%	-5,04%
	Frase 4	0,08%	0,39%	0,64%	0,58%	-0,16%	-1,27%	0,33%	0,02%	-0,72%	0,16%	-0,03%	0,23%
	Frase 5	-0,07%	0,90%	2,61%	0,15%	-0,24%	-0,86%	-0,73%	-0,88%	-0,75%	-0,70%	-0,85%	-0,41%
	Frase 6	-0,26%	0,10%	0,76%	-0,07%	-0,04%	-1,06%	-0,36%	0,62%	1,78%	-0,10%	0,07%	-1,14%
	Frase 7	-0,29%	0,05%	0,29%	0,47%	0,10%	0,11%	0,27%	0,09%	0,04%	0,40%	0,07%	0,20%
	Frase 8	10,69%	-1,44%	-2,43%	8,72%	0,21%	-0,83%	11,44%	-1,00%	-1,99%	13,58%	0,36%	-0,89%
	Frase 9	-0,54%	-0,11%	0,13%	-0,55%	-0,53%	-2,42%	-0,79%	-0,57%	-2,31%	0,73%	0,39%	-1,17%
	Frase 10	-0,02%	0,01%	0,03%	-0,12%	0,16%	0,77%	-0,12%	0,45%	2,02%	-0,06%	0,08%	0,01%
	Frase 11	0,31%	0,28%	1,70%	0,00%	0,65%	3,15%	-0,78%	0,25%	3,89%	0,52%	1,10%	4,76%
	Frase 12	0,05%	-1,60%	-1,93%	0,24%	-3,35%	-7,42%	0,07%	-3,47%	-7,67%	0,09%	-3,48%	-7,91%

Nas tabelas 30 e 31 temos os erros de processamento das formantes para cada frase para o indivíduo M8-17 e médios do grupo MG1.

Tabela 30 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M8-17 e o grupo MG1.

		G.722			G.723.1			G.726		
		Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M8-17	Frase 1	-2,70%	-1,57%	1,60%	-8,81%	0,52%	-1,89%	-0,72%	1,64%	2,31%
	Frase 2	-0,62%	-2,48%	1,13%	-5,02%	2,95%	-1,26%	3,36%	0,41%	0,90%
	Frase 3	0,87%	0,36%	1,74%	-3,30%	-0,38%	-0,59%	2,09%	1,87%	2,18%
	Frase 4	-0,41%	-0,85%	2,18%	-9,03%	-0,37%	-1,07%	4,17%	1,97%	2,93%
	Frase 5	-0,18%	-0,96%	0,65%	-8,19%	-1,72%	-1,04%	-1,44%	0,06%	0,49%
	Frase 6	-0,14%	-1,13%	0,94%	-6,19%	-0,60%	-2,01%	0,43%	2,49%	1,45%
	Frase 7	0,37%	-1,13%	0,98%	-6,78%	-3,09%	-2,30%	-3,08%	0,38%	1,04%
	Frase 8	2,44%	6,13%	0,19%	-7,03%	-0,51%	-3,37%	6,68%	9,27%	0,55%
	Frase 9	-2,05%	-0,41%	1,01%	-8,96%	-2,00%	-3,28%	6,03%	0,51%	1,24%
	Frase 10	-0,50%	-7,94%	0,62%	-6,14%	-6,75%	-3,78%	4,31%	1,59%	1,51%
	Frase 11	2,13%	-0,10%	0,61%	2,83%	-0,87%	-1,08%	-0,12%	2,90%	0,67%
	Frase 12	1,59%	-0,66%	2,06%	-10,71%	-3,06%	-2,79%	3,01%	0,78%	2,32%
MG1	Frase 1	-3,83%	-1,96%	0,05%	-6,89%	-0,92%	-1,69%	0,36%	1,37%	1,71%
	Frase 2	-2,93%	2,81%	0,45%	-5,13%	5,52%	-0,75%	1,53%	1,11%	0,81%
	Frase 3	0,89%	1,23%	1,35%	-1,20%	0,86%	0,19%	1,03%	1,76%	1,69%
	Frase 4	-2,44%	-0,46%	0,65%	-6,75%	-0,22%	-0,97%	2,75%	2,30%	2,21%
	Frase 5	-1,15%	1,13%	0,27%	-5,15%	0,75%	-0,57%	-0,26%	0,51%	-1,01%
	Frase 6	-3,51%	-0,03%	-0,15%	-6,53%	0,23%	-1,71%	0,22%	1,70%	1,08%
	Frase 7	-0,49%	-2,28%	-0,83%	-4,06%	-3,26%	-2,47%	-1,56%	0,16%	0,77%
	Frase 8	1,10%	2,45%	-0,93%	-3,64%	-0,87%	-2,71%	4,14%	6,21%	0,56%
	Frase 9	-4,08%	-2,86%	-0,58%	-7,54%	-3,66%	-2,73%	3,48%	0,79%	0,97%
	Frase 10	-5,69%	-5,79%	-0,24%	-8,51%	-5,19%	-2,44%	2,05%	1,58%	1,38%
	Frase 11	0,80%	-1,08%	-0,14%	1,15%	-1,47%	-0,98%	-0,62%	1,61%	0,51%
	Frase 12	-1,98%	-1,00%	-0,38%	-8,13%	-2,20%	-2,80%	0,46%	0,76%	1,86%

Tabela 31 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M8-17 e o grupo MG1.

		G.728			G.729A			iLBC			Speex		
		Formantes			Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M8-17	Frase 1	-2,42%	1,41%	1,02%	-10,53%	-5,62%	-4,94%	-6,43%	-2,96%	-3,09%	-4,64%	-0,25%	0,82%
	Frase 2	-2,61%	-0,18%	-0,37%	-12,62%	-4,56%	-4,27%	-8,59%	-1,59%	-2,75%	-4,27%	-0,66%	0,28%
	Frase 3	2,03%	2,72%	0,95%	-6,27%	-6,32%	-4,61%	-4,16%	-5,43%	-2,72%	-1,14%	-0,27%	1,45%
	Frase 4	-0,25%	0,99%	1,26%	-9,61%	-4,91%	-4,41%	-3,62%	-3,24%	-2,86%	-5,10%	-0,67%	1,43%
	Frase 5	-3,32%	0,89%	-0,35%	-14,74%	-6,32%	-3,56%	-7,07%	-5,05%	-2,95%	-6,48%	-0,82%	-0,23%
	Frase 6	-1,86%	-0,29%	0,12%	-8,32%	-1,74%	-4,08%	-5,49%	-2,71%	-2,20%	-3,59%	-0,18%	0,17%
	Frase 7	-3,34%	-1,20%	0,05%	-7,52%	-5,60%	-5,84%	-2,59%	-2,65%	-3,06%	-1,21%	-1,70%	-0,15%
	Frase 8	2,44%	7,71%	-0,48%	-7,63%	-4,85%	-6,71%	-4,00%	-0,42%	-3,51%	0,84%	6,26%	-0,56%
	Frase 9	-2,42%	-0,14%	0,17%	-8,89%	-4,30%	-5,18%	-6,47%	-3,69%	-2,94%	-2,93%	-0,34%	0,06%
	Frase 10	0,95%	0,43%	-0,72%	-7,94%	-9,86%	-5,46%	-4,81%	-4,35%	-3,15%	-1,32%	-2,63%	-0,73%
	Frase 11	0,11%	1,75%	0,09%	-4,88%	-4,20%	-2,95%	-5,03%	-5,95%	-1,54%	1,93%	-0,29%	0,38%
	Frase 12	1,47%	0,10%	0,42%	-9,79%	-5,28%	-5,79%	-8,61%	-4,10%	-2,60%	-2,43%	-1,35%	-0,06%
MG1	Frase 1	-0,60%	1,67%	0,63%	-10,76%	-5,17%	-5,48%	-6,59%	-6,25%	-2,88%	-2,24%	-0,34%	0,54%
	Frase 2	-2,44%	1,18%	-0,03%	-9,54%	-1,52%	-3,66%	-7,71%	-0,98%	-2,73%	-3,63%	0,69%	0,41%
	Frase 3	0,38%	2,65%	0,89%	-6,69%	-7,72%	-3,94%	-5,44%	-6,24%	-2,52%	-1,82%	-0,94%	1,12%
	Frase 4	1,32%	2,17%	0,92%	-8,97%	-7,05%	-3,90%	-4,17%	-3,88%	-2,48%	-3,65%	-0,18%	1,19%
	Frase 5	-1,89%	1,40%	-0,13%	-9,03%	-4,30%	-2,91%	-6,21%	-4,67%	-2,32%	-2,37%	1,19%	0,12%
	Frase 6	-1,35%	0,91%	0,19%	-7,90%	-3,22%	-3,93%	-4,47%	-2,64%	-2,28%	-1,70%	0,68%	0,20%
	Frase 7	-1,42%	-1,36%	-0,07%	-6,11%	-7,20%	-5,55%	-1,94%	-3,11%	-2,60%	-0,38%	-1,51%	-0,33%
	Frase 8	1,16%	5,43%	-0,44%	-6,45%	-5,79%	-6,31%	-3,42%	-1,91%	-2,71%	0,19%	3,13%	-0,38%
	Frase 9	-1,87%	-0,01%	0,07%	-8,11%	-6,28%	-4,74%	-5,49%	-4,29%	-2,57%	1,76%	0,77%	0,08%
	Frase 10	-0,36%	1,22%	0,19%	-11,02%	-9,58%	-5,17%	-5,18%	-4,81%	-2,83%	-1,20%	-1,63%	0,10%
	Frase 11	-0,67%	0,35%	-0,02%	-6,13%	-5,99%	-2,69%	-4,10%	-6,96%	-1,73%	0,91%	-1,77%	0,49%
	Frase 12	0,47%	0,51%	0,41%	-9,71%	-5,08%	-5,87%	-8,89%	-3,94%	-2,39%	-3,97%	-1,24%	0,23%

Nas figuras de 78 a 95 temos os resultados do indivíduo M8-17 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas das frequências

em Hertz por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

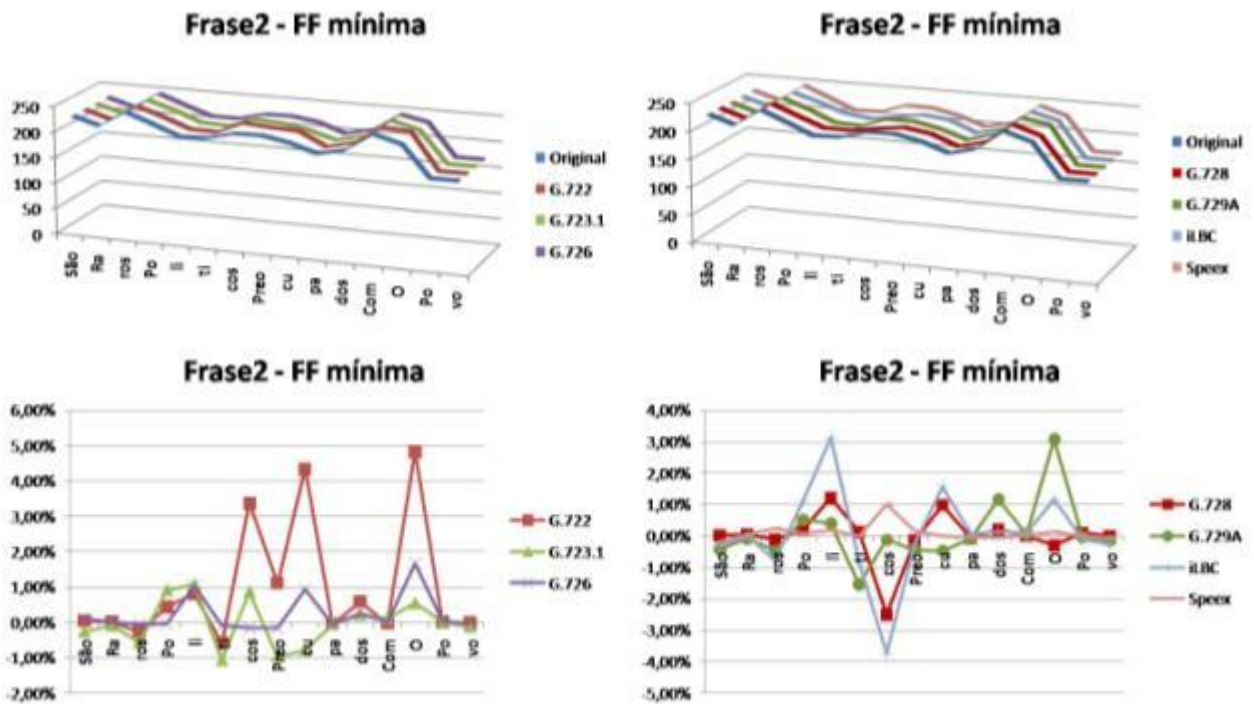


Figura 78 - Variações da frequência fundamental mínima da frase 2 do indivíduo M8-17 do grupo MG1.

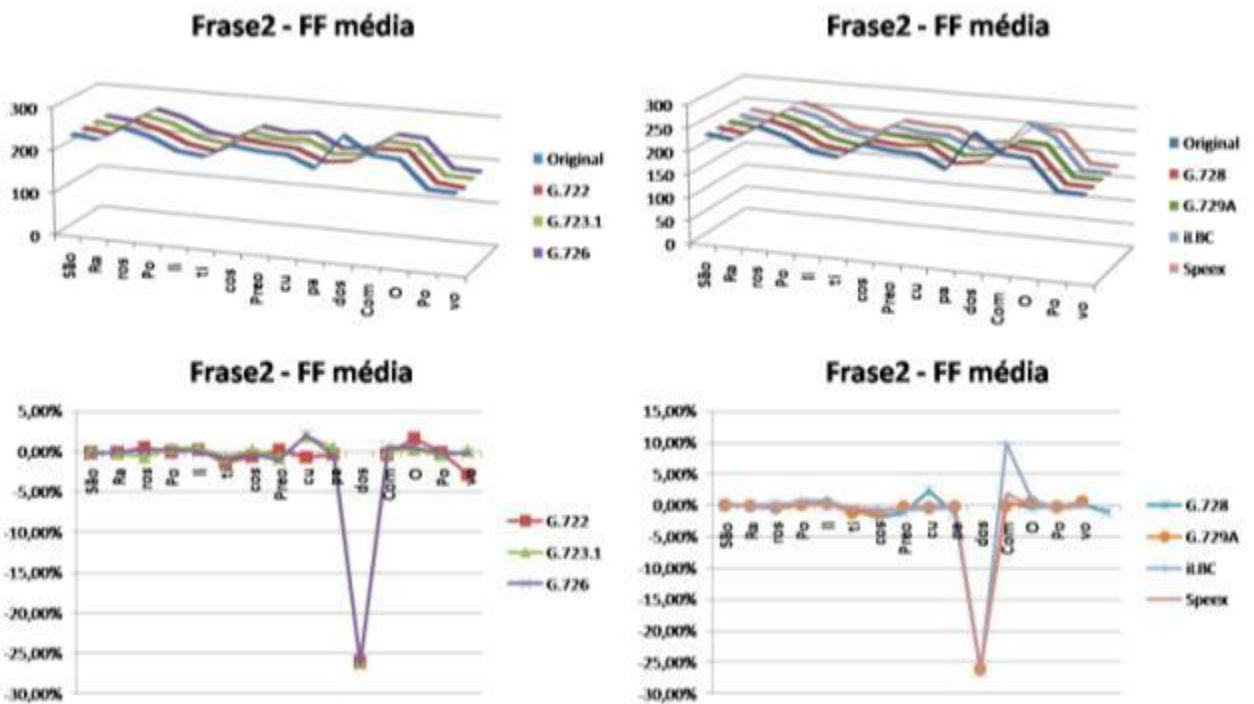


Figura 79 - Variações da frequência fundamental média da frase 2 do indivíduo M8-17 do grupo MG1.

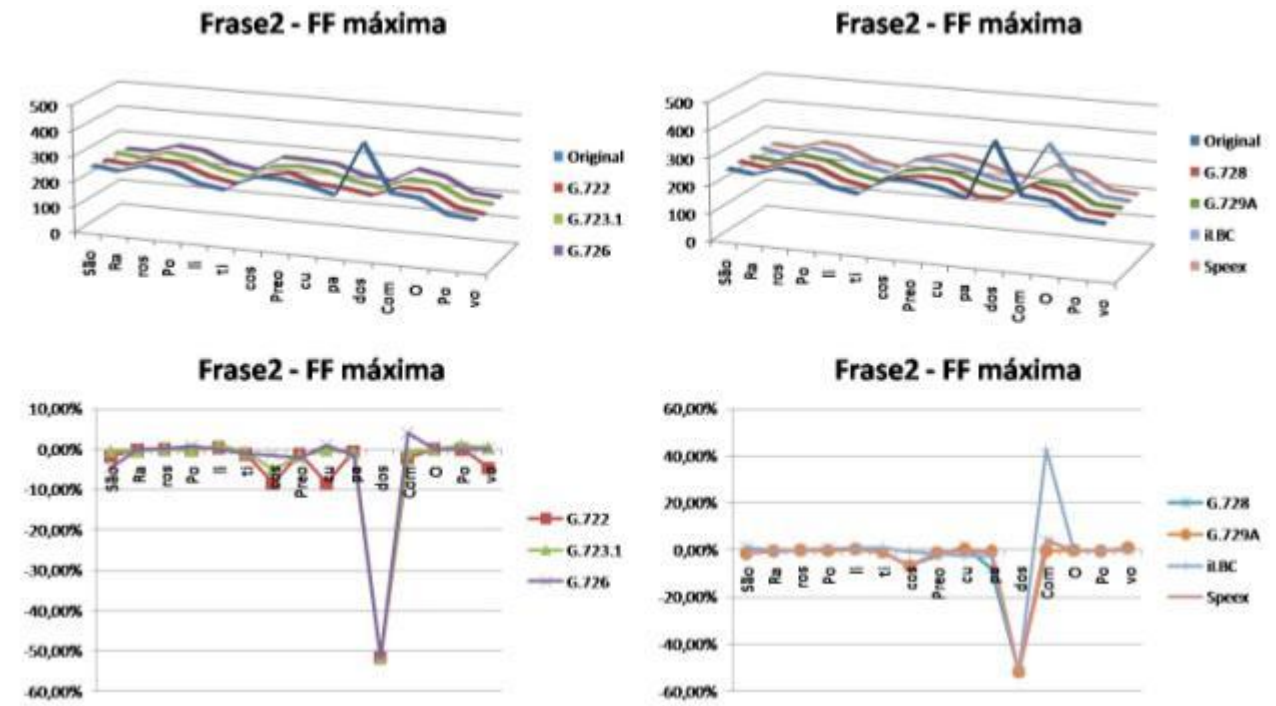


Figura 80 - Variações da frequência fundamental máxima da frase 2 do indivíduo M8-17 do grupo MG1.

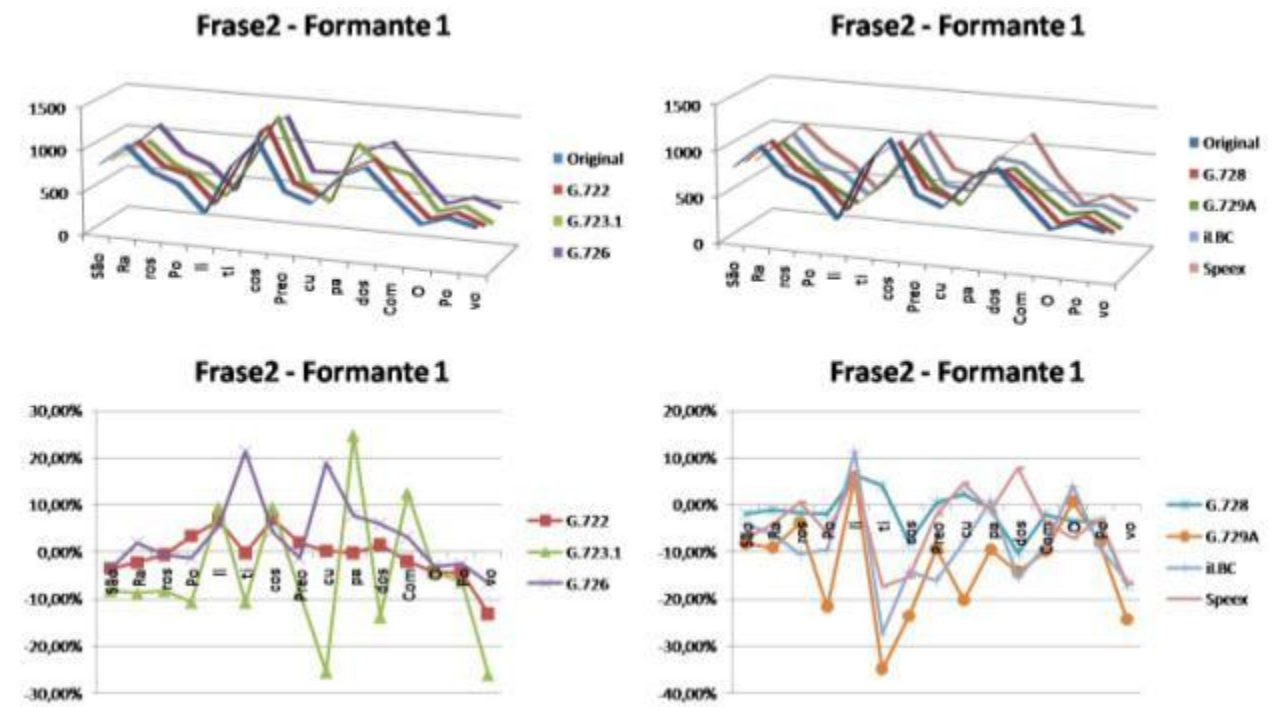


Figura 81 - Variações da frequência da primeira formante da frase 2 do indivíduo M8-17 do grupo MG1.

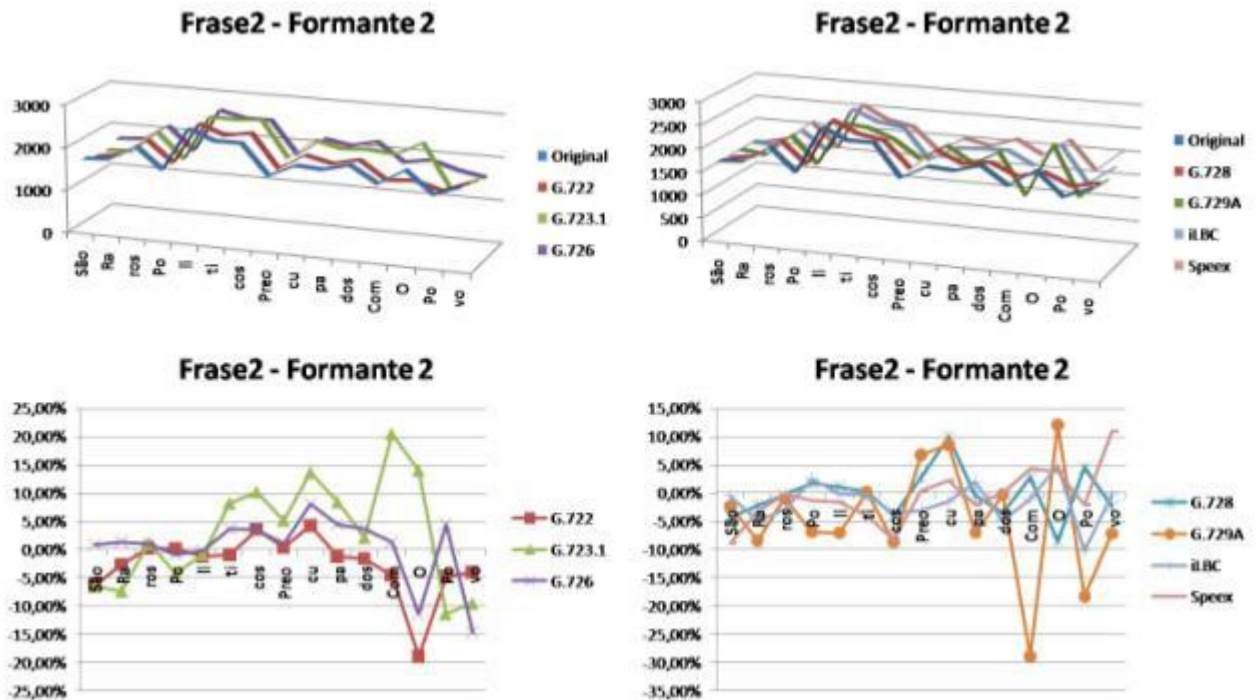


Figura 82 - Variações da freqüência da segunda formante da frase 2 do indivíduo M8-17 do grupo MG1.

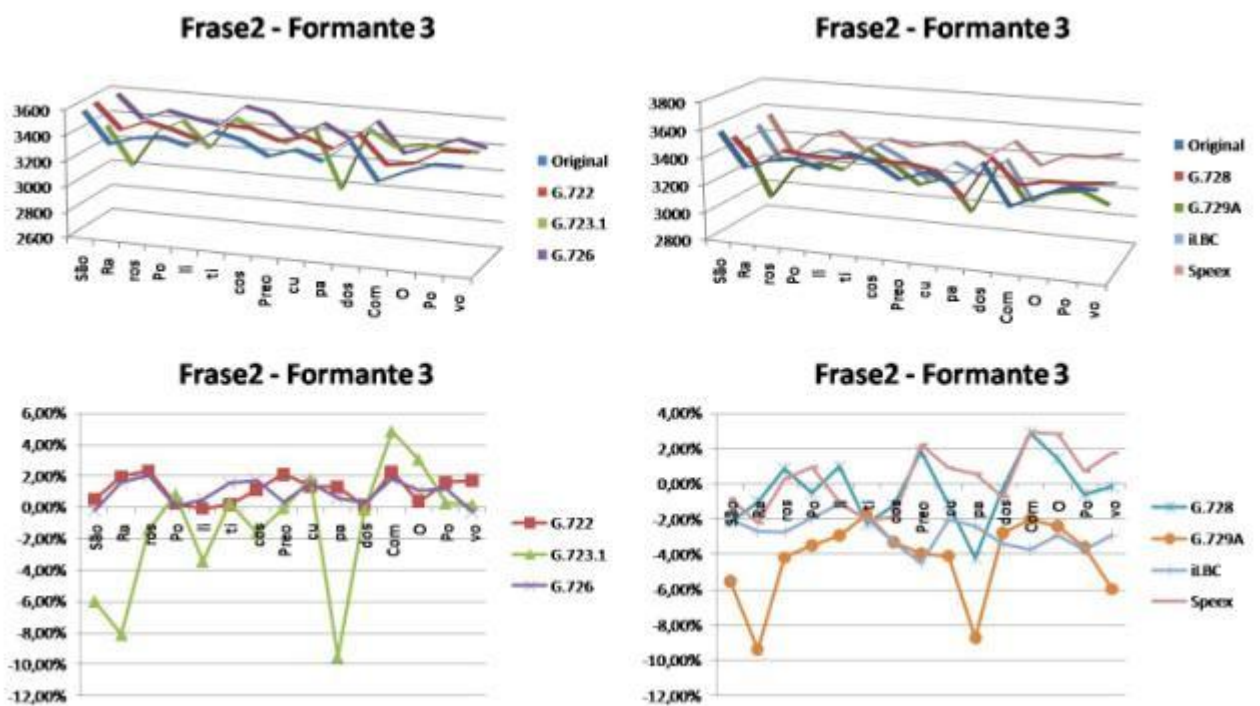


Figura 83 - Variações da freqüência da terceira formante da frase 2 do indivíduo M8-17 do grupo MG1.

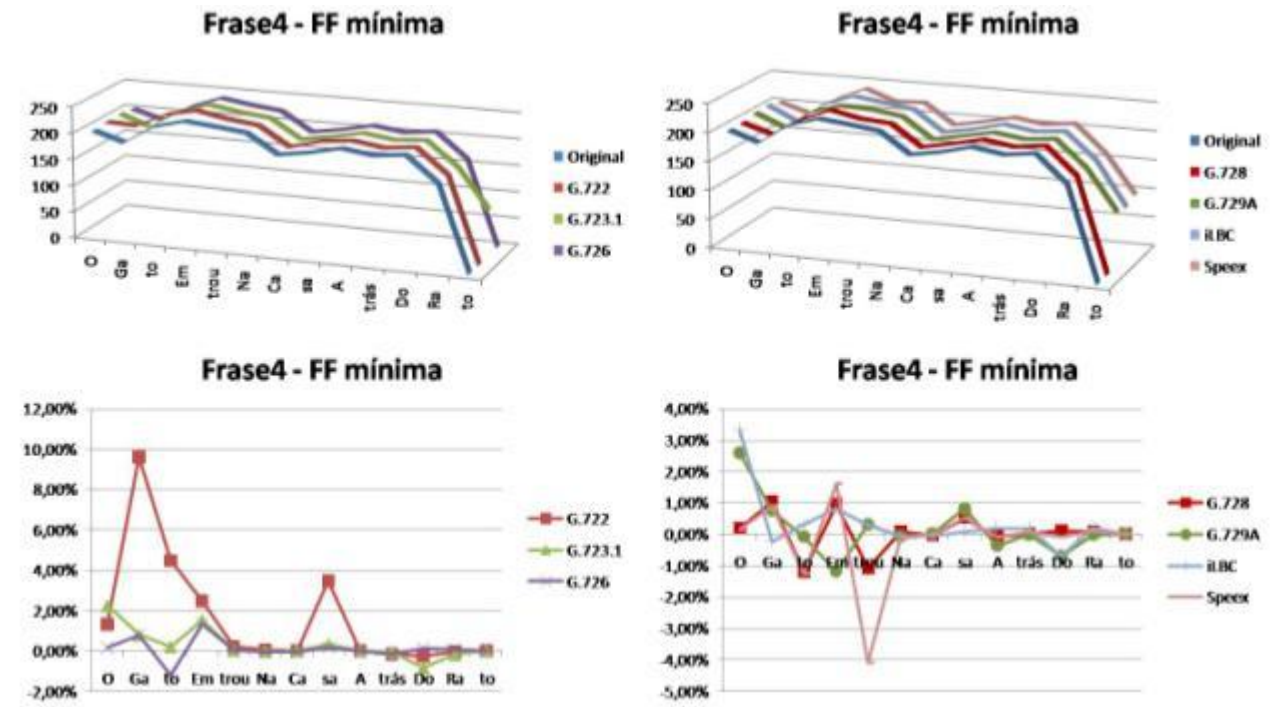


Figura 84 - Variações da frequência fundamental mínima da frase 4 do indivíduo M8-17 do grupo MG1.

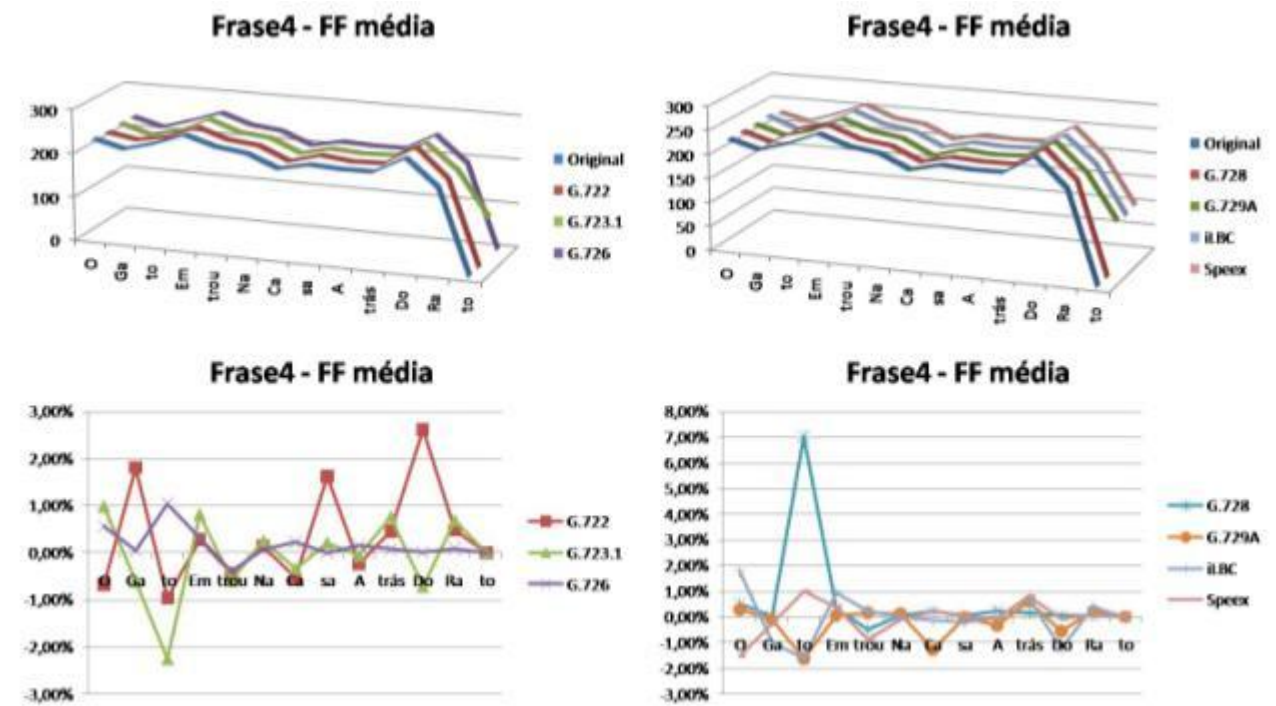


Figura 85 - Variações da frequência fundamental média da frase 4 do indivíduo M8-17 do grupo MG1.

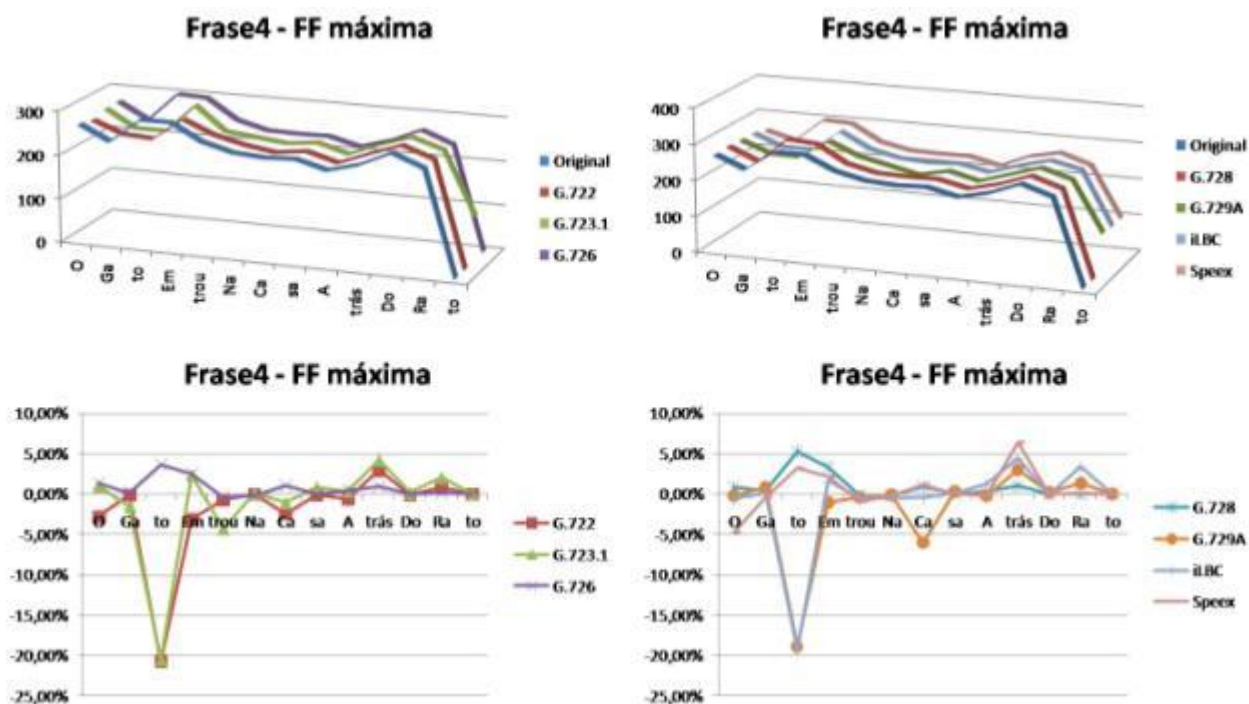


Figura 86 - Variações da frequência fundamental máxima da frase 4 do indivíduo M8-17 do grupo MG1.

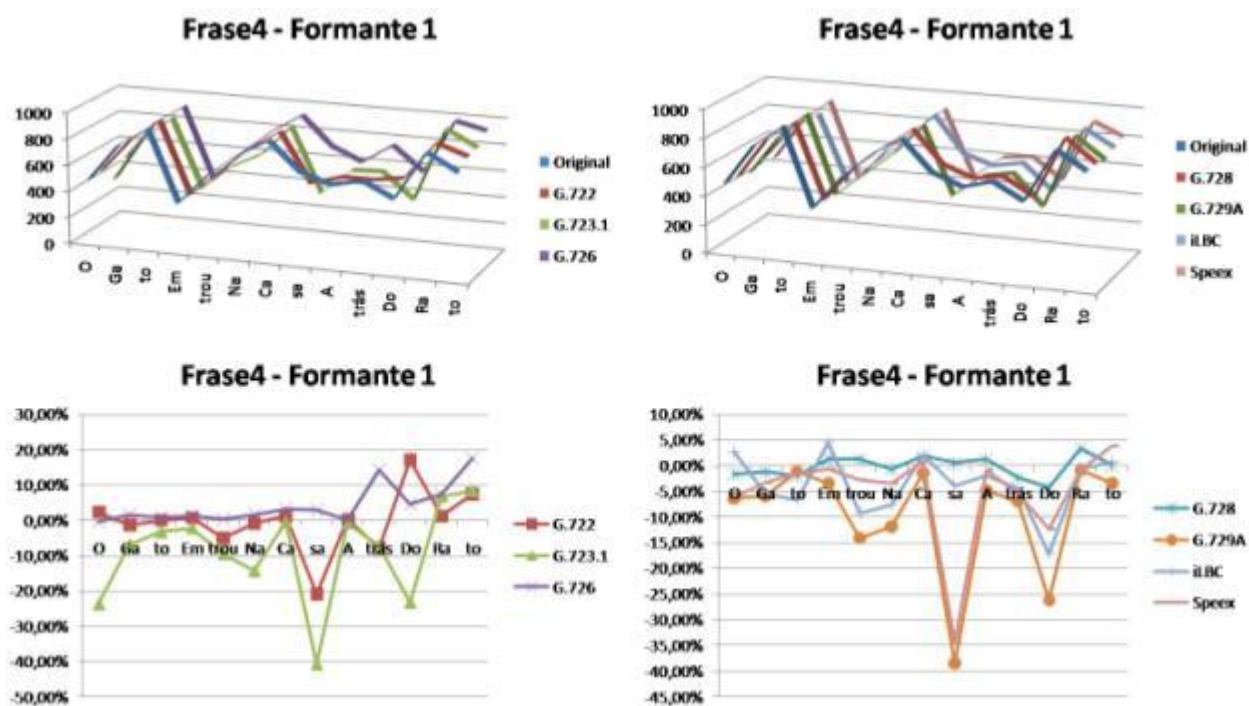


Figura 87 - Variações da frequência da primeira formante da frase 4 do indivíduo M8-17 do grupo MG1.

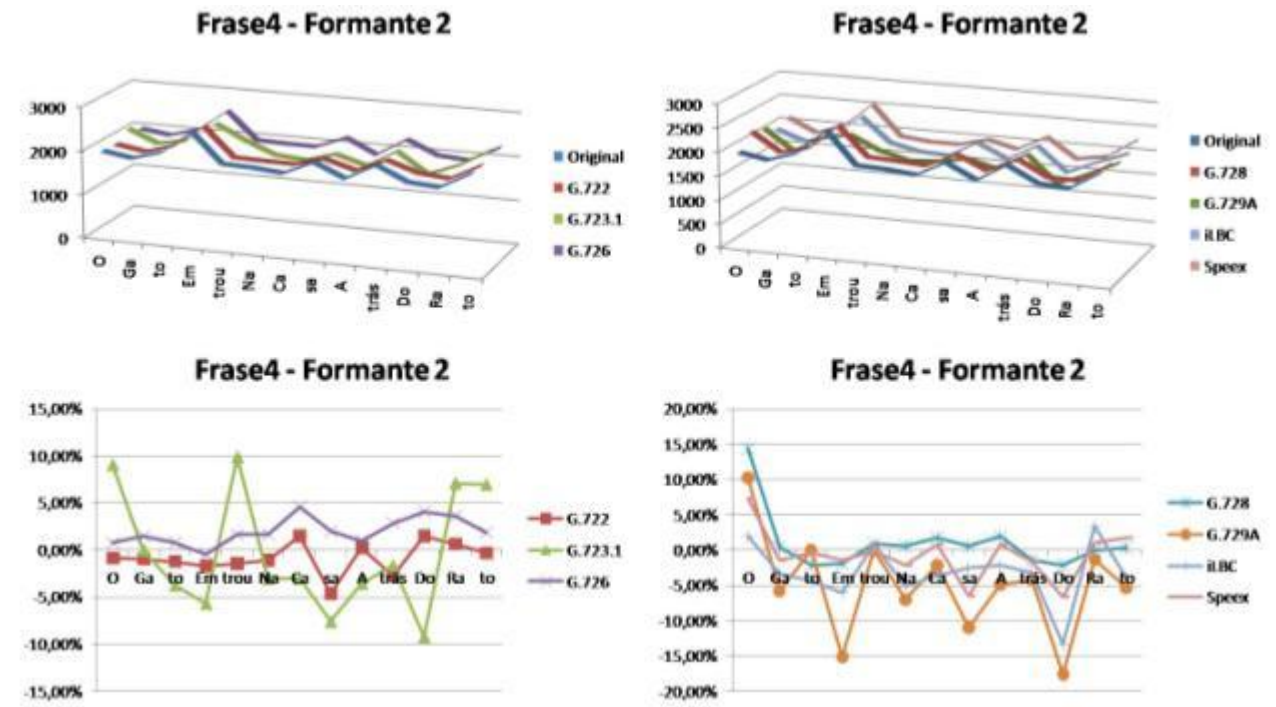


Figura 88 - Variações da frequência da segunda formante da frase 4 do indivíduo M8-17 do grupo MG1.

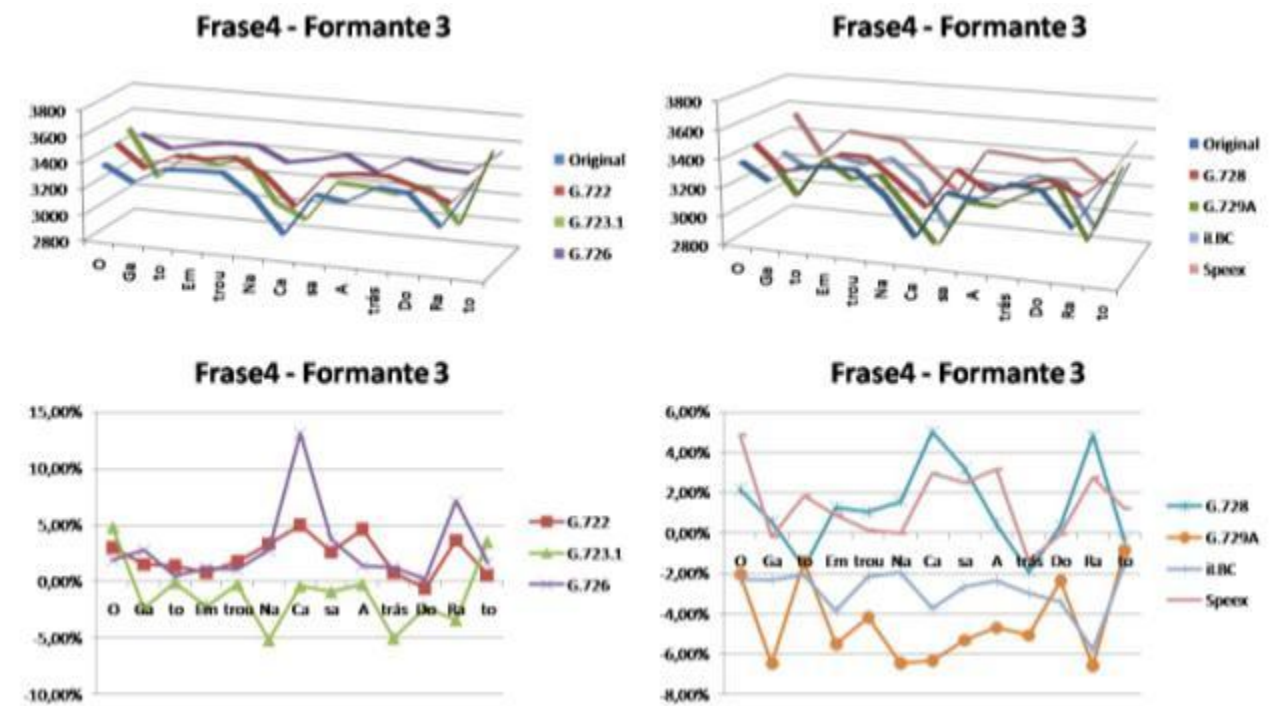


Figura 89 - Variações da frequência da terceira formante da frase 4 do indivíduo M8-17 do grupo MG1.

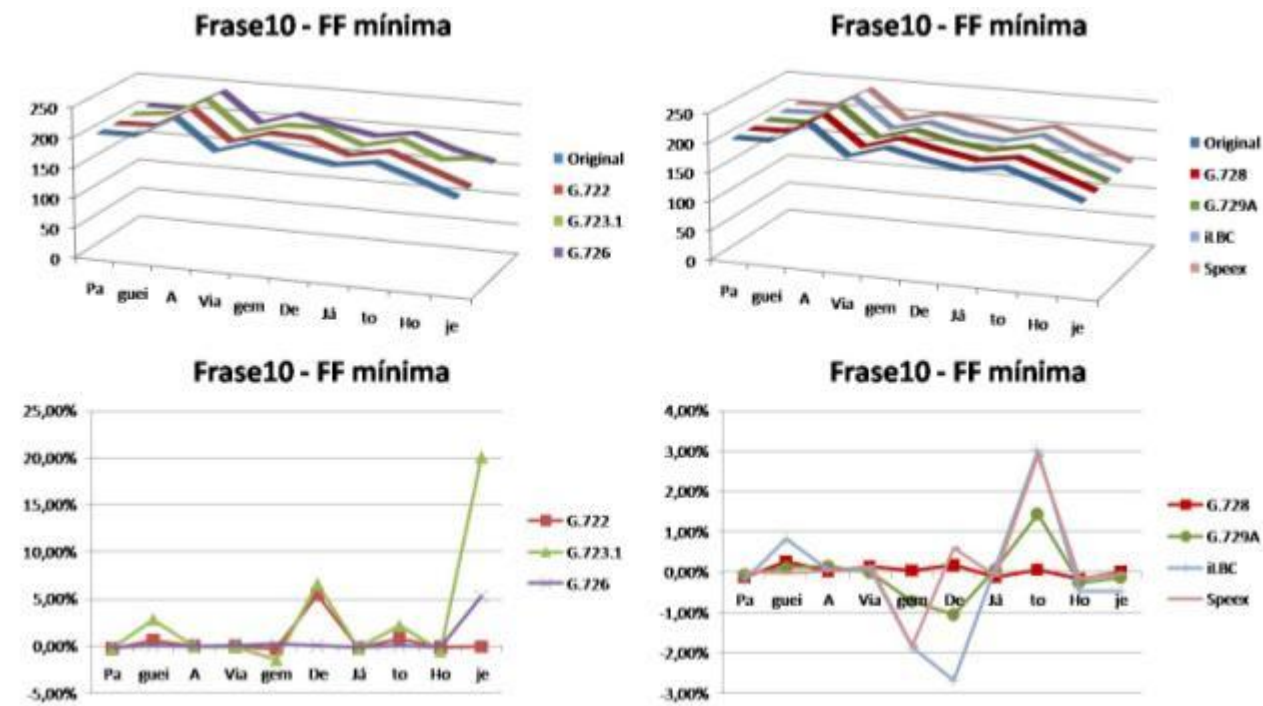


Figura 90 - Variações da frequência fundamental mínima da frase 10 do indivíduo M8-17 do grupo MG1.

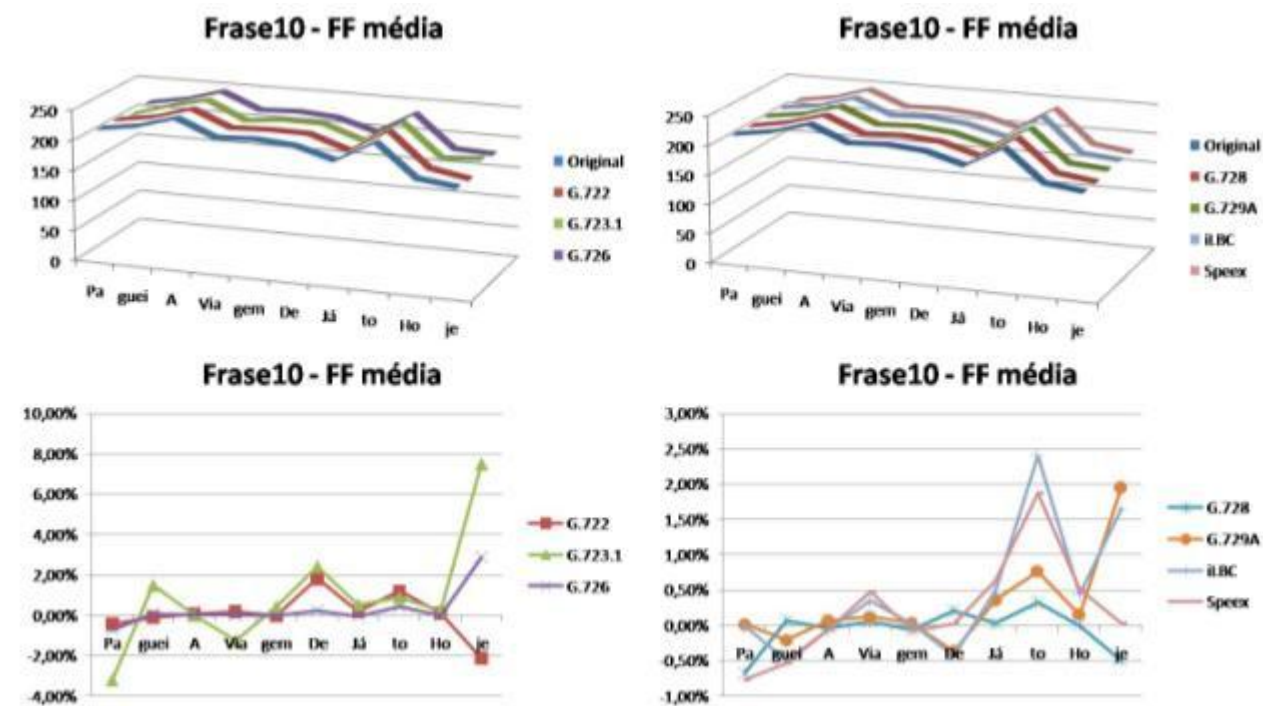


Figura 91 - Variações da frequência fundamental média da frase 10 do indivíduo M8-17 do grupo MG1.

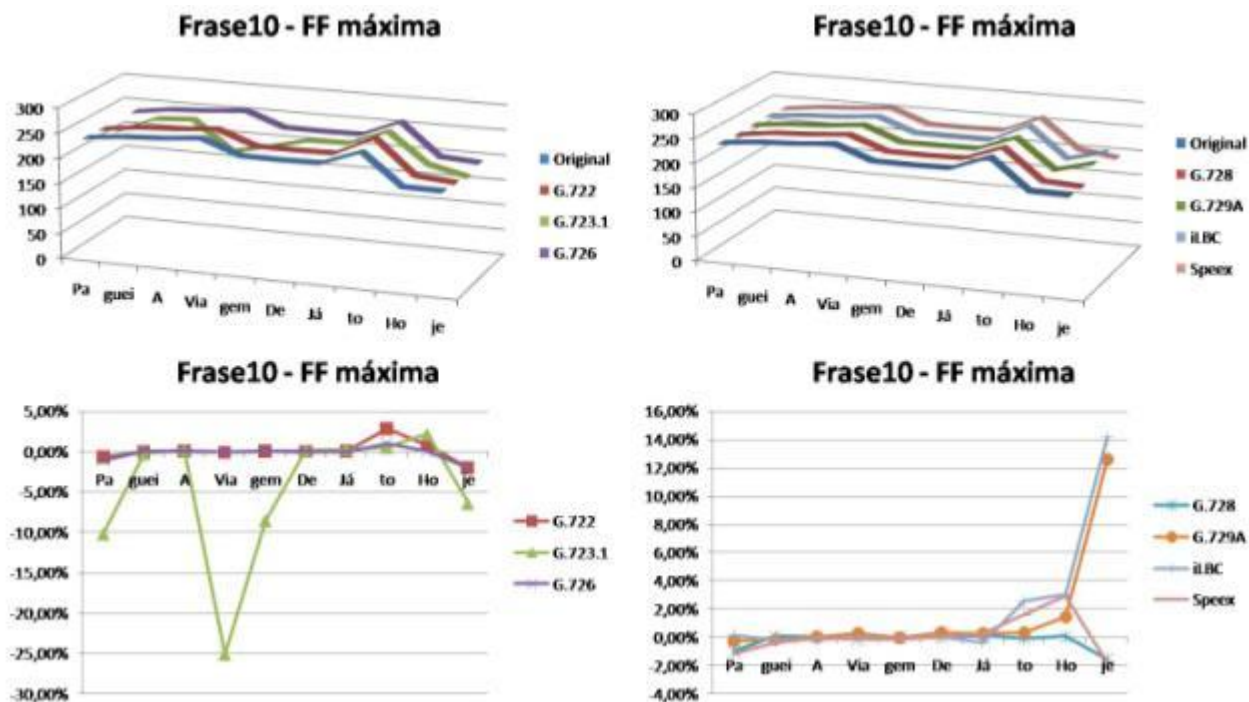


Figura 92 - Variações da frequência fundamental máxima da frase 10 do indivíduo M8-17 do grupo MG1.

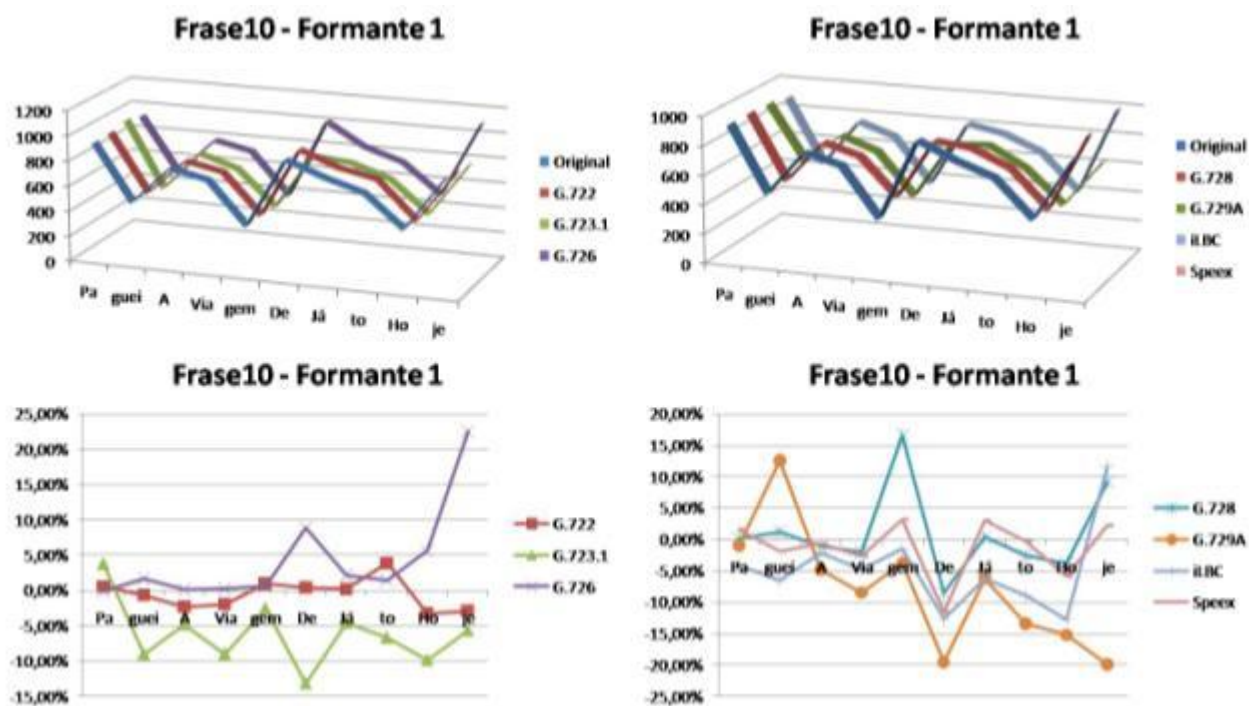


Figura 93 - Variações da frequência da primeira formante da frase 10 do indivíduo M8-17 do grupo MG1.

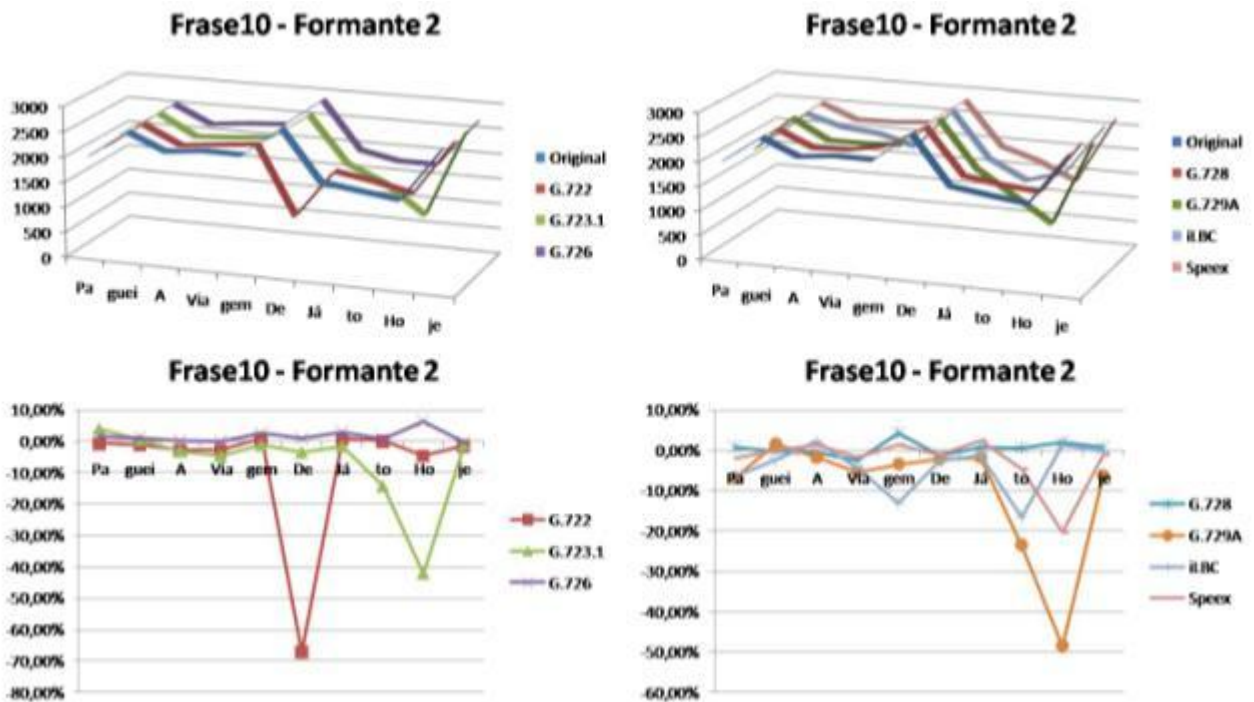


Figura 94 - Variações da frequência da segunda formante da frase 10 do indivíduo M8-17 do grupo MG1.

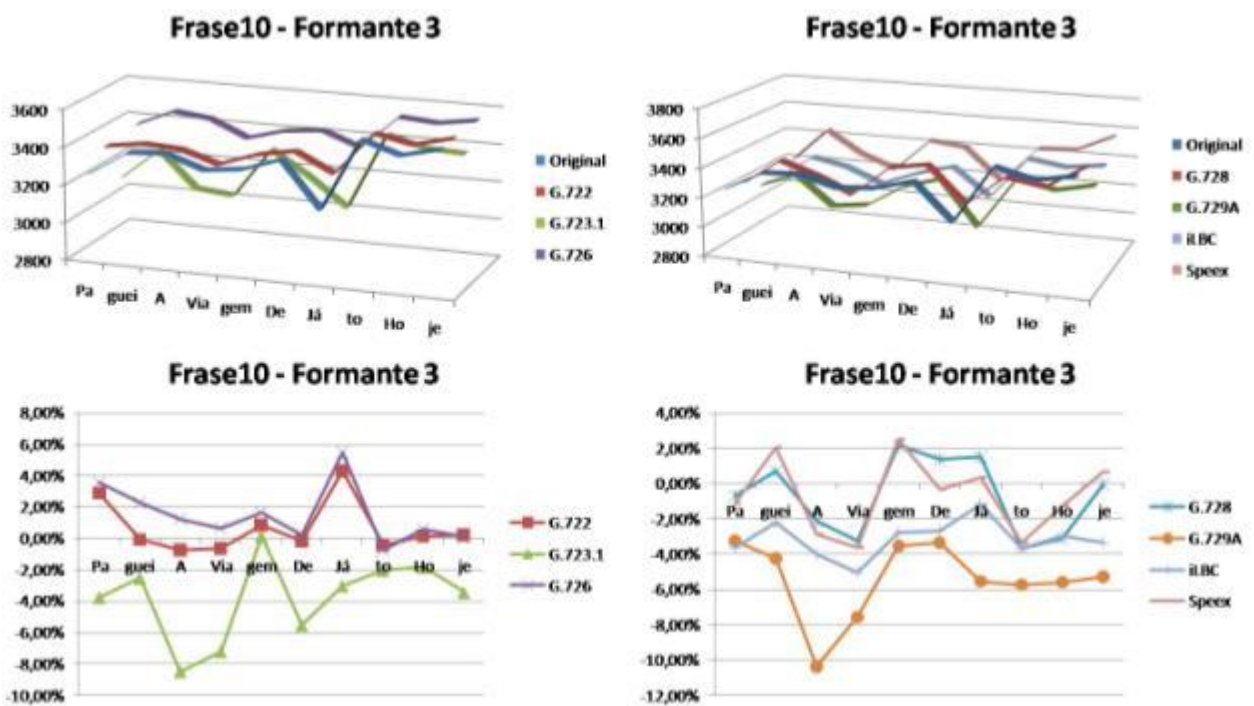


Figura 95 - Variações da frequência da terceira formante da frase 10 do indivíduo M8-17 do grupo MG1.

4.4.5 – MG2

Fazem parte do grupo os indivíduos M1-52 e M3-48. O indivíduo M3-48 foi escolhido para ter seus resultados individuais por ter apresentado erros de processamento.

Neste grupo, o indivíduo M3-48 apresentou erros de processamento com a inserção de ruído e forte atenuação para o codificador G.723.1 nas frases 3, 6, 7 e 10.

Na tabela 32 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo M3-48 e o grupo MG2. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento.

Tabela 32 - Duração dos intervalos médios e dos arquivos do grupo MG2 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento			
		Intervalo Médio (s)	Duração (s)	G.723.1	G.728	G.729A	iLBC
M3-48	Frase 1	0,344	4,133	19,36%	7,26%	2,42%	4,84%
	Frase 2	0,312	4,687	17,07%	8,53%	4,27%	4,27%
	Frase 3	0,250	2,253	22,19%	8,88%	4,44%	4,44%
	Frase 4	0,267	3,466	20,20%	8,66%	2,89%	5,77%
	Frase 5	0,278	2,499	20,01%	8,00%	4,00%	4,00%
	Frase 6	0,242	1,939	20,63%	10,31%	0,00%	5,16%
	Frase 7	0,296	3,256	18,43%	9,22%	3,07%	3,07%
	Frase 8	0,282	3,388	20,66%	8,85%	2,95%	5,90%
	Frase 9	0,320	3,842	18,22%	7,81%	2,60%	2,60%
	Frase 10	0,399	3,991	20,04%	7,52%	2,51%	5,01%
	Frase 11	0,312	3,117	19,25%	9,62%	3,21%	6,42%
	Frase 12	0,313	4,069	19,66%	7,37%	4,91%	4,91%
Grupo	Frase 1	0,282	3,381	21,09%	9,33%	3,11%	6,22%
	Frase 2	0,237	3,551	20,96%	10,48%	4,20%	6,27%
	Frase 3	0,211	1,895	20,85%	10,94%	2,22%	5,47%
	Frase 4	0,204	2,652	23,70%	9,77%	1,44%	5,61%
	Frase 5	0,227	2,040	19,49%	10,33%	2,00%	5,16%
	Frase 6	0,193	1,542	23,43%	9,53%	0,00%	6,95%
	Frase 7	0,218	2,398	18,96%	11,10%	1,54%	4,78%
	Frase 8	0,200	2,404	20,89%	11,47%	1,48%	6,47%
	Frase 9	0,228	2,737	18,30%	10,03%	1,30%	4,36%
	Frase 10	0,281	2,811	19,22%	9,89%	1,25%	5,57%
	Frase 11	0,220	2,205	21,23%	8,68%	1,60%	7,08%
	Frase 12	0,238	3,100	21,56%	8,38%	2,46%	4,80%

Nas tabelas 33 e 34 temos os erros de processamento das frequências fundamentais mínimas, médias e máximas para cada frase para o indivíduo M3-48 e médios do grupo MG2.

Tabela 33 - Erros de processamento da frequência fundamental dos codificadores G.722, G.723.1 e G726 para o indivíduo M3-48 e o grupo MG2.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M3-48	Frase 1	4,24%	-1,94%	-4,22%	0,28%	0,14%	-0,13%	0,31%	0,47%	0,17%
	Frase 2	1,14%	0,53%	-2,95%	6,65%	2,27%	-1,12%	0,42%	0,47%	0,11%
	Frase 3	15,75%	0,80%	-0,45%	26,68%	3,40%	-2,63%	-0,09%	0,17%	0,30%
	Frase 4	13,18%	0,23%	-6,19%	1,18%	-0,86%	-5,65%	1,47%	0,45%	-1,11%
	Frase 5	1,77%	-2,19%	-3,45%	-0,13%	1,89%	10,38%	-0,31%	1,65%	9,05%
	Frase 6	1,70%	0,14%	-0,50%	5,50%	-6,82%	-12,18%	-0,36%	0,37%	0,09%
	Frase 7	1,73%	0,24%	-1,99%	-4,43%	-0,42%	2,63%	0,05%	-0,24%	-1,51%
	Frase 8	1,89%	-2,01%	-9,12%	1,85%	5,59%	-0,33%	-0,15%	-0,21%	-2,34%
	Frase 9	0,19%	-4,34%	-6,13%	9,47%	3,07%	5,60%	0,27%	0,19%	-0,66%
	Frase 10	29,09%	2,03%	-2,73%	26,39%	13,53%	18,11%	10,56%	0,97%	2,50%
	Frase 11	1,66%	-7,93%	-13,45%	9,49%	0,68%	-10,12%	-0,07%	1,25%	5,86%
	Frase 12	12,48%	0,77%	-1,51%	16,83%	7,80%	4,28%	0,33%	-0,95%	-0,32%
MG2	Frase 1	1,46%	-1,24%	-3,85%	-0,35%	0,77%	0,06%	-0,43%	0,66%	0,14%
	Frase 2	1,36%	0,65%	-1,31%	3,60%	1,41%	-0,19%	0,18%	0,23%	0,07%
	Frase 3	7,91%	0,35%	-0,20%	13,58%	2,07%	-0,68%	-0,06%	0,20%	0,15%
	Frase 4	6,62%	-0,14%	-4,32%	0,73%	-0,70%	-3,86%	0,83%	0,35%	-0,71%
	Frase 5	1,26%	-1,09%	-1,98%	-0,16%	3,72%	10,75%	-0,10%	0,72%	4,07%
	Frase 6	1,38%	0,34%	-0,13%	3,18%	-2,91%	-5,64%	-0,23%	0,18%	0,06%
	Frase 7	0,97%	-0,13%	-1,96%	-2,06%	-0,02%	0,74%	0,06%	-0,19%	-1,45%
	Frase 8	1,48%	-0,89%	-4,51%	1,28%	2,90%	0,05%	-0,34%	-0,11%	-0,94%
	Frase 9	1,06%	-1,50%	-2,94%	4,98%	1,84%	2,63%	0,09%	0,42%	-0,33%
	Frase 10	14,94%	1,06%	-1,45%	13,70%	6,91%	9,12%	5,58%	0,48%	1,18%
	Frase 11	0,96%	-4,03%	-6,94%	4,81%	0,43%	-4,90%	-0,05%	0,59%	2,81%
	Frase 12	6,44%	-0,18%	-0,73%	8,49%	4,62%	2,03%	0,32%	-1,15%	-0,10%

Tabela 34 - Erros de processamento da frequência fundamental dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M3-48 e o grupo MG2.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M3-48	Frase 1	-0,08%	-0,12%	-0,13%	9,81%	4,95%	2,54%	9,76%	4,49%	2,02%	9,47%	4,14%	2,12%
	Frase 2	6,62%	2,40%	0,46%	-0,53%	0,86%	-0,22%	2,40%	1,67%	1,30%	5,99%	2,21%	-0,78%
	Frase 3	-0,19%	0,18%	0,51%	-0,23%	-0,01%	-0,15%	-6,16%	-1,19%	2,10%	0,15%	0,42%	2,18%
	Frase 4	1,27%	0,00%	-0,12%	0,01%	-0,94%	-5,09%	1,12%	-0,30%	-3,76%	0,28%	-0,47%	-4,48%
	Frase 5	0,05%	2,46%	10,39%	-0,80%	0,06%	4,94%	0,07%	-0,97%	-0,88%	-0,02%	-0,85%	-1,88%
	Frase 6	-1,91%	-0,07%	0,03%	-0,34%	0,04%	0,78%	0,32%	0,27%	0,47%	0,50%	-0,37%	-0,70%
	Frase 7	0,06%	-0,17%	-1,21%	-0,06%	-0,06%	-0,36%	-0,14%	0,10%	-0,68%	-0,05%	-0,04%	-0,51%
	Frase 8	-0,24%	-0,53%	-2,75%	-2,68%	-2,43%	-8,02%	-4,04%	-2,53%	-8,36%	-3,35%	-2,46%	-8,01%
	Frase 9	0,06%	-0,41%	0,05%	1,56%	1,79%	8,95%	0,00%	-0,32%	3,19%	-0,38%	-0,98%	0,12%
	Frase 10	0,08%	-0,04%	-0,23%	0,50%	-1,03%	-2,45%	10,39%	0,05%	0,14%	0,25%	0,10%	0,29%
	Frase 11	-0,03%	0,79%	4,19%	-0,19%	-1,60%	-4,78%	-0,03%	-2,13%	-5,27%	0,10%	-2,27%	-5,43%
	Frase 12	0,32%	-0,96%	-0,33%	1,12%	0,09%	0,10%	-3,10%	-0,49%	2,28%	0,02%	0,17%	-0,02%
MG2	Frase 1	-1,34%	0,54%	1,51%	4,35%	2,41%	3,17%	2,42%	1,97%	0,63%	3,47%	2,29%	2,44%
	Frase 2	3,22%	1,19%	0,23%	-0,36%	0,45%	0,01%	-0,03%	0,65%	1,10%	2,93%	1,11%	-0,36%
	Frase 3	-0,06%	0,14%	0,24%	-0,12%	0,04%	0,07%	-2,83%	-0,55%	1,16%	0,29%	0,42%	1,39%
	Frase 4	0,73%	-0,25%	-1,37%	-0,05%	-0,74%	-3,51%	0,67%	-0,30%	-2,68%	0,35%	-0,27%	-1,87%
	Frase 5	0,05%	1,14%	4,73%	-0,84%	-0,14%	2,38%	0,18%	-0,43%	-0,91%	-0,03%	-0,18%	-0,40%
	Frase 6	-0,53%	0,08%	0,05%	-0,04%	0,33%	0,82%	7,84%	2,03%	1,31%	0,25%	0,22%	0,34%
	Frase 7	0,07%	-0,02%	-1,06%	0,14%	-0,19%	-1,21%	-1,00%	0,43%	0,02%	0,23%	0,14%	-0,96%
	Frase 8	-0,04%	-0,24%	-1,34%	-1,12%	-1,17%	-3,79%	-2,24%	-1,26%	-4,10%	-1,73%	-1,14%	-3,87%
	Frase 9	0,04%	-0,05%	-0,14%	1,22%	1,33%	4,73%	0,36%	-0,02%	1,47%	0,03%	-0,13%	-0,14%
	Frase 10	0,42%	0,01%	-0,12%	0,34%	-0,38%	-1,01%	5,79%	0,23%	0,21%	0,67%	0,22%	0,18%
	Frase 11	-0,04%	0,37%	2,01%	-0,09%	-0,78%	-2,40%	0,18%	-0,82%	-2,45%	0,55%	-0,84%	-2,58%
	Frase 12	0,37%	-1,10%	-0,21%	0,48%	1,18%	0,18%	-1,47%	0,08%	1,44%	0,21%	0,86%	0,03%

Nas tabelas 35 e 36 temos os erros de processamento das formantes para cada frase para o indivíduo M3-48 e médios do grupo MG2.

Tabela 35 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M3-48 e o grupo MG2.

		G.722			G.723.1			G.726		
		Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M3-48	Frase 1	-0,35%	0,44%	2,23%	-3,37%	-2,18%	0,88%	-2,55%	1,03%	-0,04%
	Frase 2	1,52%	0,70%	1,37%	-8,48%	-0,19%	-0,76%	-1,92%	1,61%	1,96%
	Frase 3	3,70%	1,07%	2,90%	-1,98%	0,63%	0,48%	0,67%	1,90%	2,93%
	Frase 4	-0,62%	0,59%	1,71%	-3,42%	0,19%	-0,70%	-0,60%	1,90%	3,73%
	Frase 5	2,42%	-1,43%	1,59%	-3,65%	-0,53%	-0,96%	-6,17%	-1,48%	1,55%
	Frase 6	1,86%	0,39%	1,43%	-9,99%	-6,67%	-6,93%	0,74%	1,87%	1,74%
	Frase 7	2,02%	0,76%	2,07%	7,07%	-8,02%	-6,03%	0,42%	2,01%	2,55%
	Frase 8	-0,07%	-0,60%	1,98%	-12,70%	-19,92%	-5,67%	0,62%	0,76%	2,84%
	Frase 9	4,03%	1,00%	1,94%	-16,14%	-12,30%	-4,93%	8,69%	2,75%	2,65%
	Frase 10	1,60%	0,41%	1,69%	-15,82%	-10,40%	-7,91%	1,72%	1,47%	2,56%
	Frase 11	4,03%	-0,84%	0,75%	6,81%	-6,91%	-8,45%	-1,25%	0,57%	1,10%
	Frase 12	-1,46%	-0,58%	1,98%	-6,56%	-10,63%	-7,83%	-3,03%	1,26%	3,62%
MG2	Frase 1	0,95%	0,31%	2,60%	-4,77%	-1,94%	0,74%	1,28%	1,93%	1,85%
	Frase 2	1,33%	0,99%	1,38%	-5,35%	1,74%	-0,61%	-0,98%	2,92%	1,86%
	Frase 3	-0,35%	0,03%	2,95%	-2,55%	1,60%	0,24%	0,40%	1,94%	3,31%
	Frase 4	-1,23%	0,73%	2,39%	-6,62%	1,20%	0,17%	-0,81%	2,47%	3,95%
	Frase 5	2,90%	0,46%	2,44%	-4,43%	2,00%	0,46%	-0,22%	0,50%	2,15%
	Frase 6	3,73%	0,92%	2,34%	-11,68%	-3,93%	-3,44%	3,45%	2,25%	2,10%
	Frase 7	3,04%	0,60%	3,18%	4,77%	-4,41%	-3,41%	-1,08%	1,36%	3,24%
	Frase 8	-1,84%	-1,17%	2,11%	-15,97%	-12,95%	-2,41%	1,91%	3,43%	2,81%
	Frase 9	3,62%	1,35%	3,73%	-11,36%	-6,08%	-2,17%	8,45%	3,24%	3,96%
	Frase 10	1,96%	0,36%	2,66%	-8,87%	-5,28%	-4,00%	5,06%	2,40%	3,16%
	Frase 11	0,34%	-0,48%	1,33%	-0,83%	-2,35%	-3,50%	3,41%	1,87%	1,51%
	Frase 12	0,85%	-0,29%	2,85%	-6,59%	-6,46%	-4,00%	0,09%	2,47%	4,34%

Tabela 36 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M3-48 e o grupo MG2.

		G.728			G.729A			iLBC			Speex		
		Formantes			Formantes			Formantes			Formantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M3-48	Frase 1	-2,96%	1,63%	2,47%	-9,10%	-7,26%	-3,48%	-5,96%	-5,05%	-2,51%	-1,50%	-0,89%	1,65%
	Frase 2	-2,20%	1,22%	0,83%	-8,73%	-5,27%	-3,76%	-3,89%	-4,32%	-2,98%	-1,00%	-0,99%	1,16%
	Frase 3	2,03%	1,30%	1,40%	-6,05%	-7,81%	-5,21%	-2,47%	-5,48%	-3,50%	-1,48%	-0,93%	1,66%
	Frase 4	-0,60%	2,03%	1,87%	-10,13%	-5,57%	-4,57%	-8,88%	-5,22%	-2,78%	-2,63%	-0,45%	1,61%
	Frase 5	-3,28%	-0,66%	0,24%	-4,27%	-5,08%	-4,32%	-3,91%	-5,55%	-3,13%	0,06%	-1,60%	0,78%
	Frase 6	-0,84%	2,12%	1,40%	-5,06%	-5,66%	-3,71%	-5,08%	-3,76%	-3,47%	1,24%	0,26%	1,35%
	Frase 7	0,73%	1,57%	1,63%	-7,24%	-8,73%	-5,15%	-1,55%	-5,09%	-2,75%	0,77%	-0,27%	1,28%
	Frase 8	-1,28%	0,22%	1,75%	-7,04%	-10,01%	-4,26%	-4,60%	-5,93%	-3,05%	1,16%	-1,15%	2,18%
	Frase 9	-1,14%	0,83%	2,06%	-14,42%	-12,15%	-12,34%	-0,40%	-3,23%	-2,68%	1,09%	-0,16%	2,10%
	Frase 10	-1,60%	0,86%	1,22%	-9,51%	-0,94%	-3,52%	-4,19%	-1,66%	-2,36%	-0,09%	0,53%	1,06%
	Frase 11	0,97%	1,05%	0,72%	-4,59%	-9,10%	-4,54%	-1,19%	-6,31%	-2,76%	0,87%	-1,02%	0,83%
	Frase 12	-1,65%	0,22%	1,09%	-11,96%	-5,47%	-5,73%	-9,46%	-4,38%	-3,21%	-3,90%	-0,86%	0,96%
MG2	Frase 1	1,02%	2,09%	2,60%	-5,78%	-6,25%	-3,03%	-5,83%	-5,19%	-2,84%	0,20%	-0,14%	1,95%
	Frase 2	-1,62%	2,88%	0,93%	-8,89%	-4,88%	-3,13%	-5,35%	-4,09%	-2,61%	-1,17%	0,39%	1,23%
	Frase 3	-0,36%	1,88%	1,79%	-8,22%	-7,57%	-4,62%	-4,63%	-5,68%	-3,22%	-3,32%	-0,79%	1,83%
	Frase 4	-2,17%	3,64%	2,32%	-10,72%	-5,07%	-4,28%	-7,70%	-5,49%	-2,88%	-3,37%	1,45%	2,27%
	Frase 5	-0,42%	1,41%	1,00%	-6,69%	-4,00%	-3,30%	-5,64%	-4,96%	-2,49%	1,84%	0,81%	1,88%
	Frase 6	0,62%	2,08%	1,15%	-11,23%	-5,38%	-3,86%	-9,61%	-4,32%	-3,44%	-2,48%	-0,57%	1,42%
	Frase 7	0,28%	1,34%	2,25%	-5,72%	-5,54%	-4,72%	-2,19%	-3,65%	-2,79%	3,30%	0,37%	2,06%
	Frase 8	-3,02%	0,71%	2,04%	-13,64%	-9,12%	-3,49%	-9,57%	-6,48%	-2,62%	-4,88%	-2,39%	2,36%
	Frase 9	0,84%	1,24%	2,33%	-12,46%	-7,72%	-7,44%	-3,35%	-2,58%	-2,49%	2,19%	0,61%	2,90%
	Frase 10	-0,33%	0,89%	1,47%	-6,80%	-3,53%	-3,29%	-3,23%	-2,89%	-2,25%	-0,21%	0,54%	1,30%
	Frase 11	-2,22%	1,34%	0,93%	-5,36%	-5,94%	-3,37%	-2,16%	-5,20%	-2,69%	0,92%	-0,30%	1,27%
	Frase 12	-1,19%	0,74%	1,83%	-9,60%	-4,97%	-4,49%	-7,16%	-3,64%	-2,89%	-3,03%	-0,68%	1,53%

Nas figuras de 96 a 113 temos os resultados do indivíduo M3-48 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas das frequências

em Hertz por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

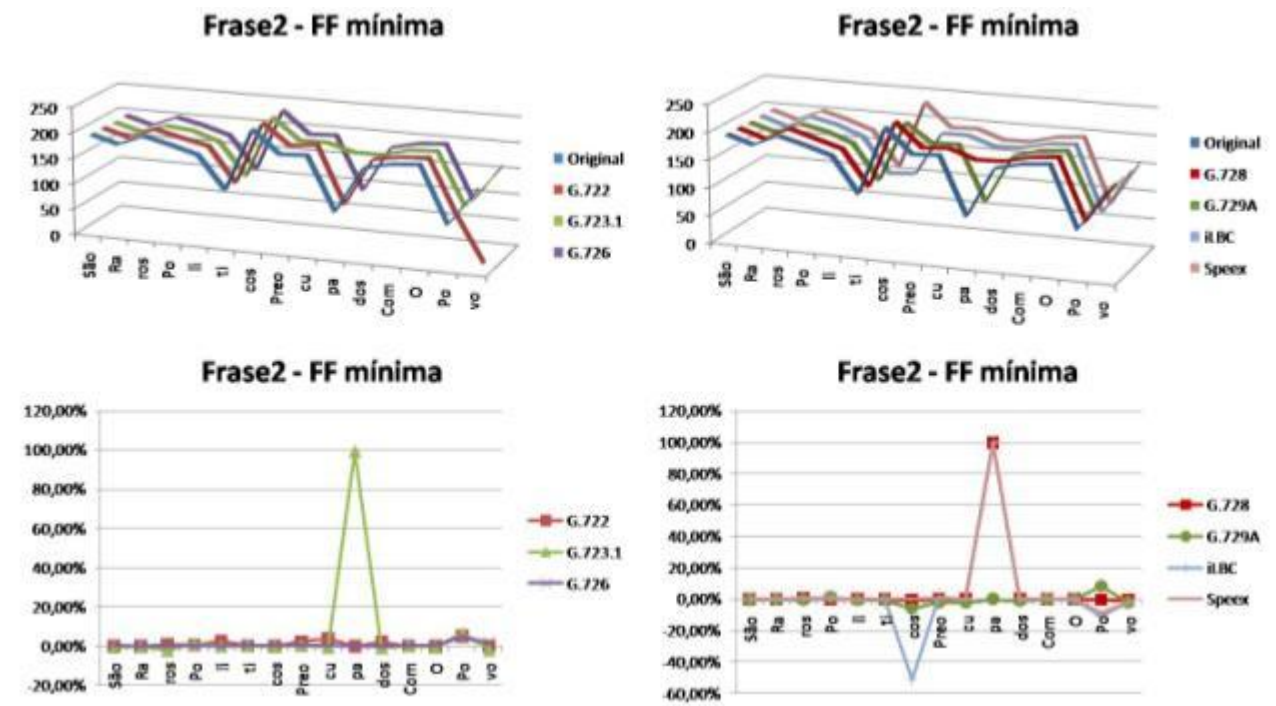


Figura 96 - Variações da frequência fundamental mínima da frase 2 do indivíduo M3-48 do grupo MG2.

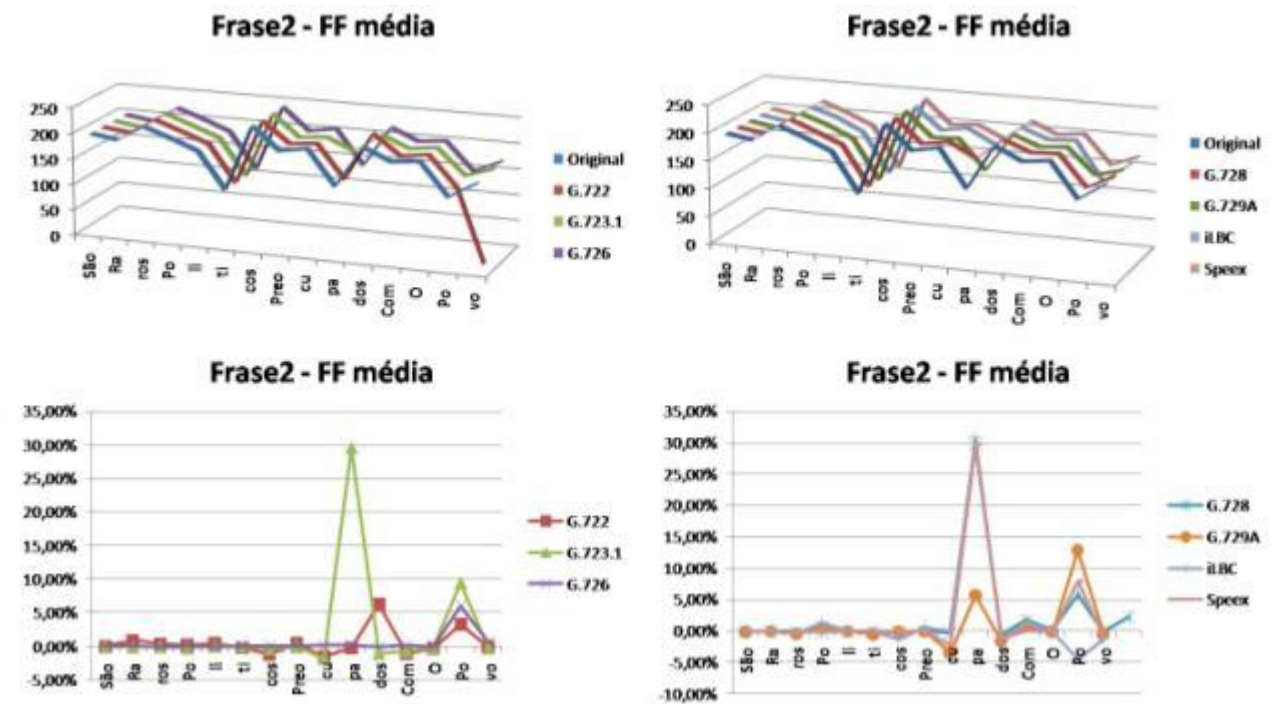


Figura 97 - Variações da frequência fundamental média da frase 2 do indivíduo M3-48 do grupo MG2.

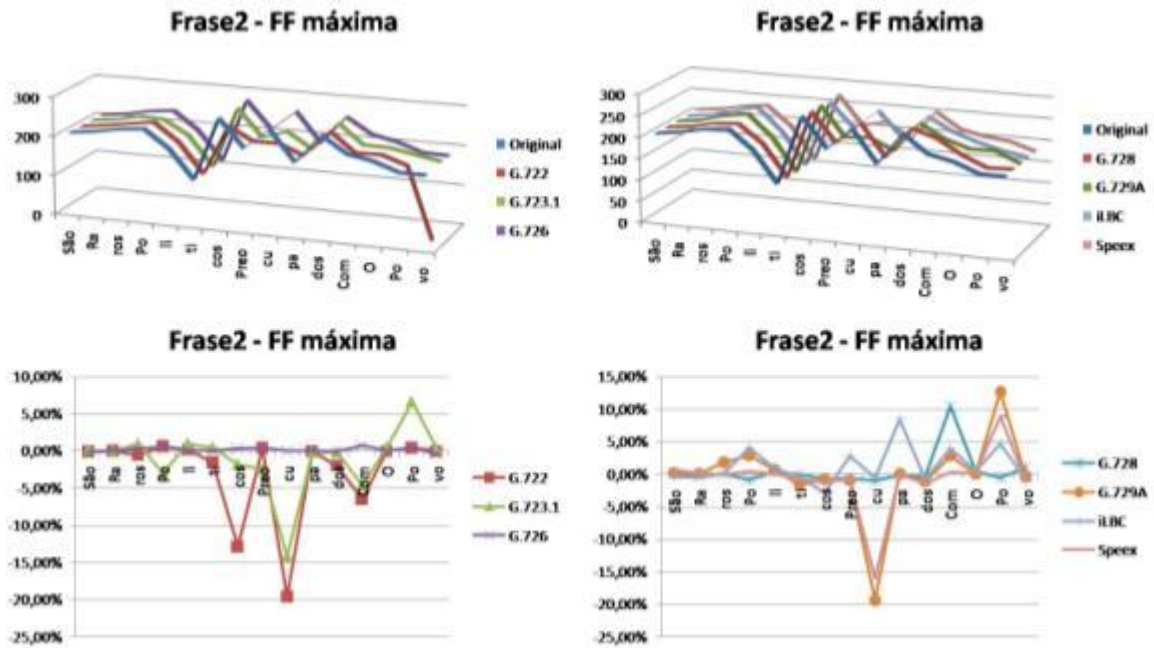


Figura 98 - Variações da frequência fundamental máxima da frase 2 do indivíduo M3-48 do grupo MG2.

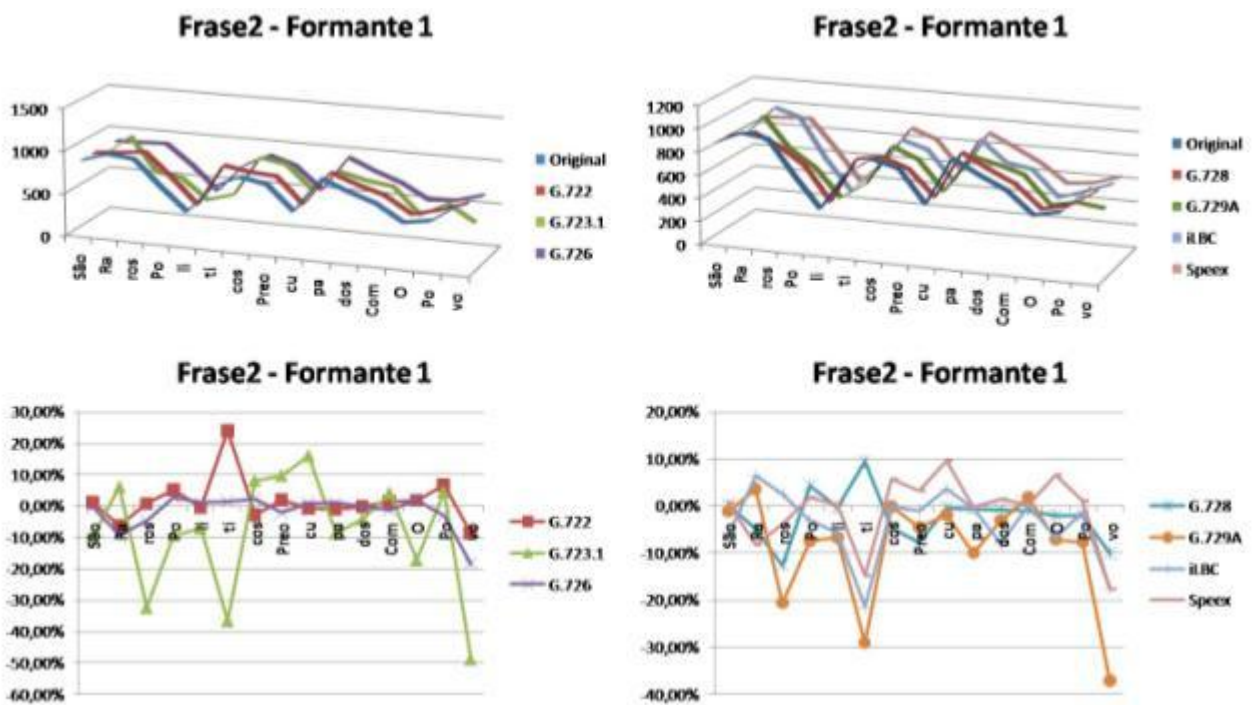


Figura 99 - Variações da frequência da primeira formante da frase 2 do indivíduo M3-48 do grupo MG2.

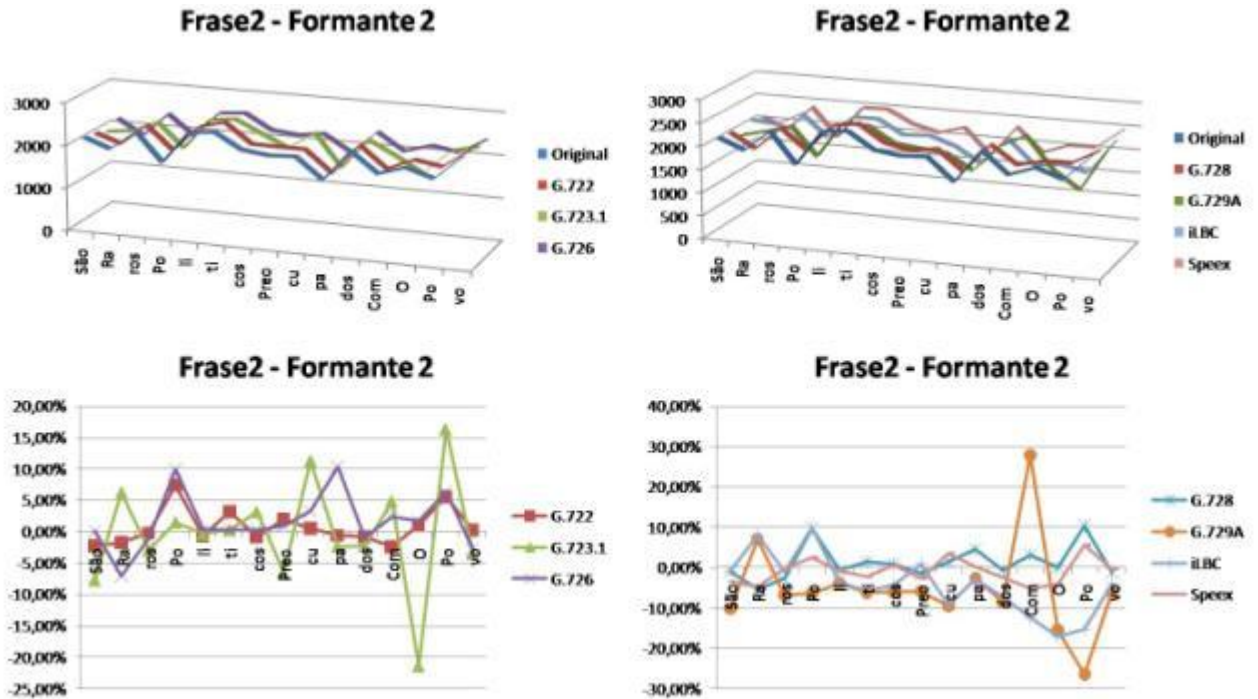


Figura 100 - Variações da frequência da segunda formante da frase 2 do indivíduo M3-48 do grupo MG2.

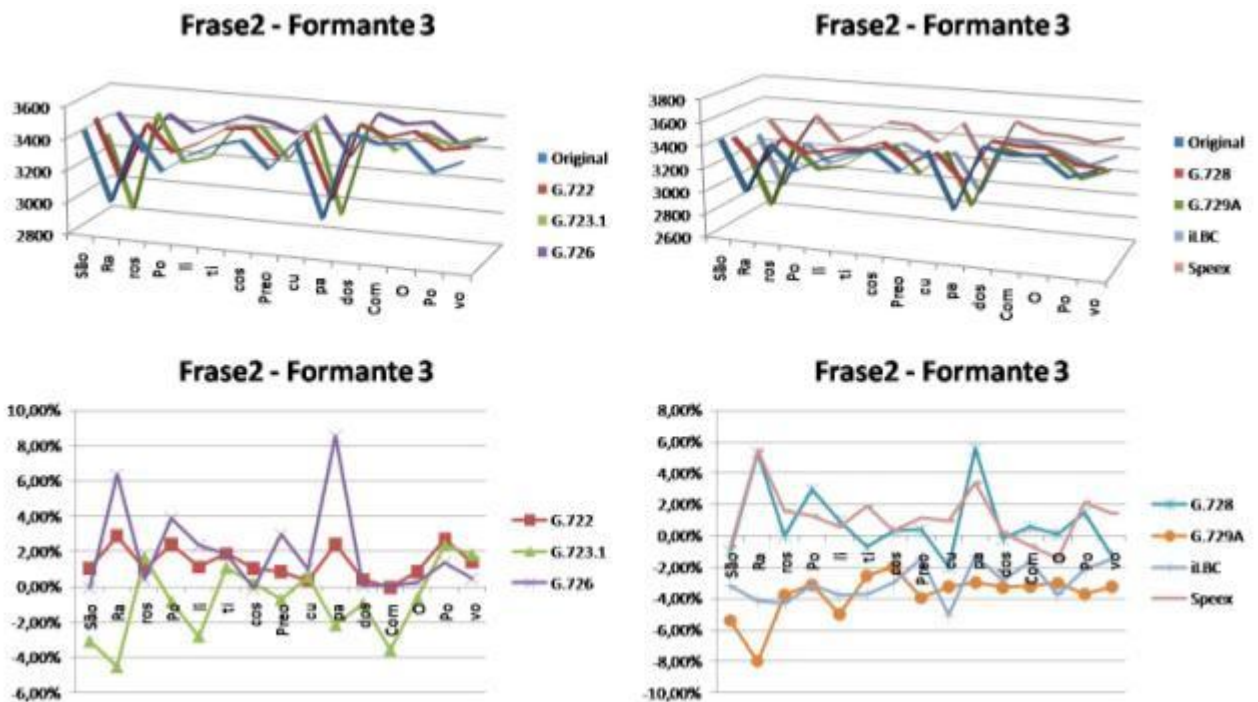


Figura 101 - Variações da frequência da terceira formante da frase 2 do indivíduo M3-48 do grupo MG2.

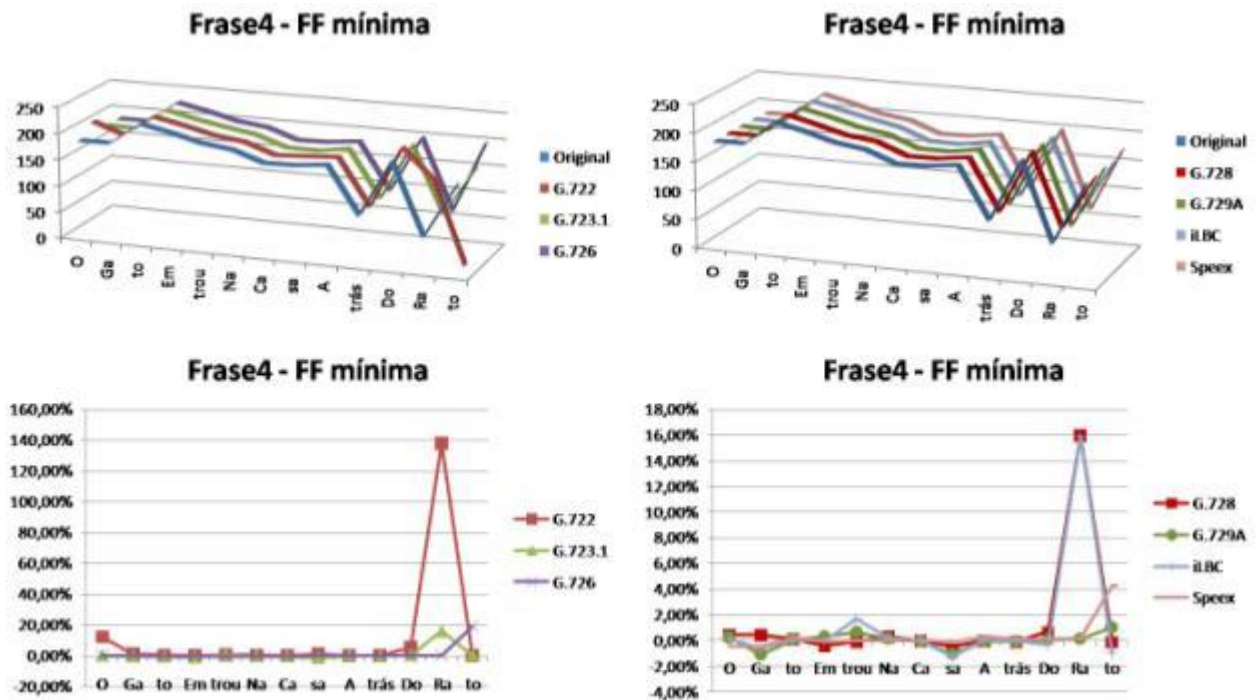


Figura 102 - Variações da frequência fundamental mínima da frase 4 do indivíduo M3-48 do grupo MG2.

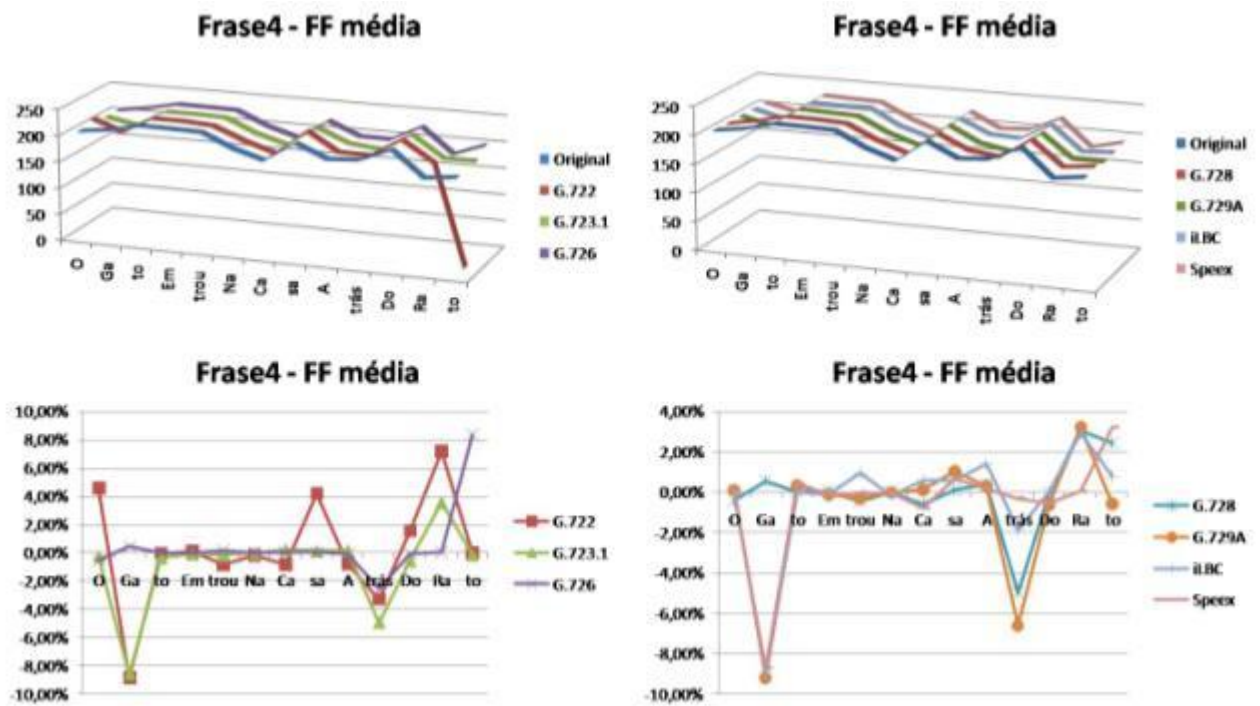


Figura 103 - Variações da frequência fundamental média da frase 4 do indivíduo M3-48 do grupo MG2.

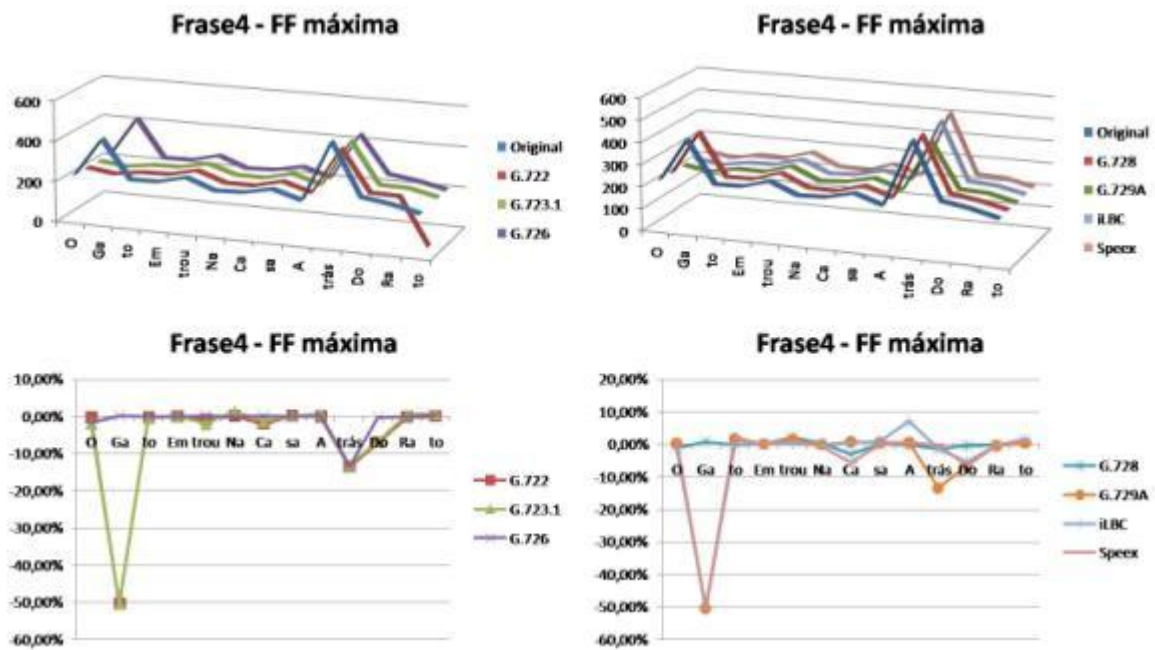


Figura 104 - Variações da frequência fundamental máxima da frase 4 do indivíduo M3-48 do grupo MG2.

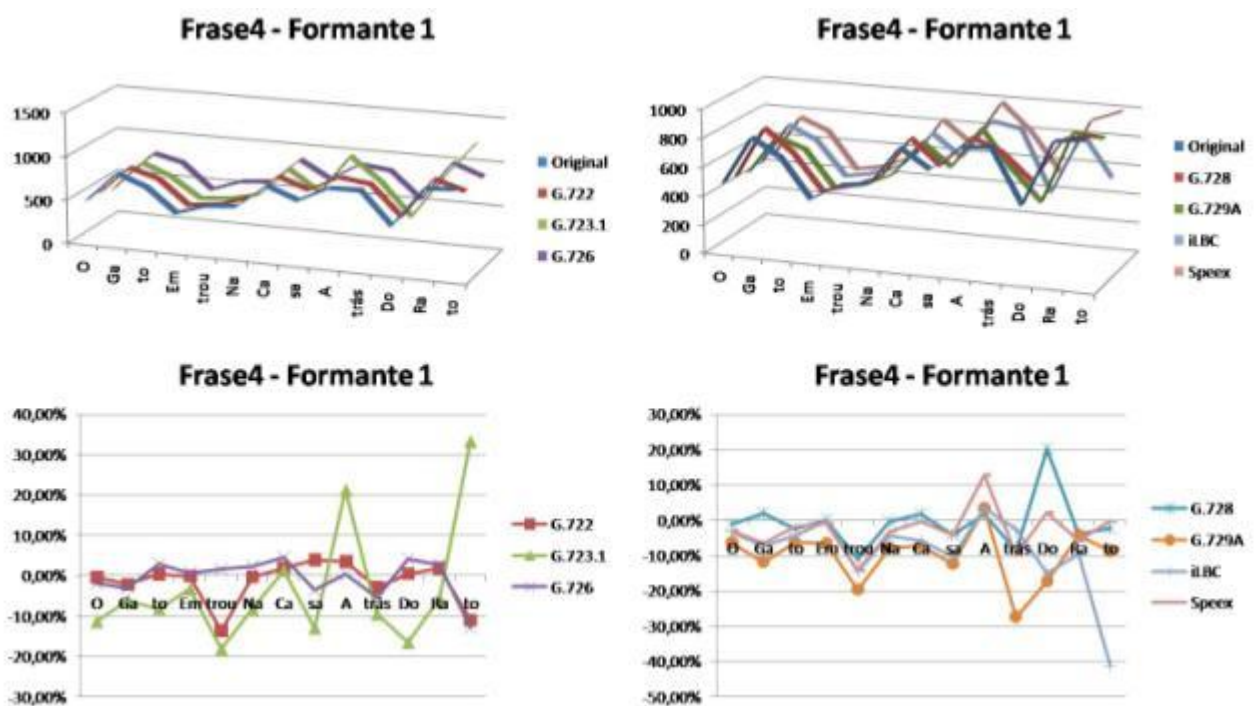


Figura 105 - Variações da frequência da primeira formante da frase 4 do indivíduo M3-48 do grupo MG2.

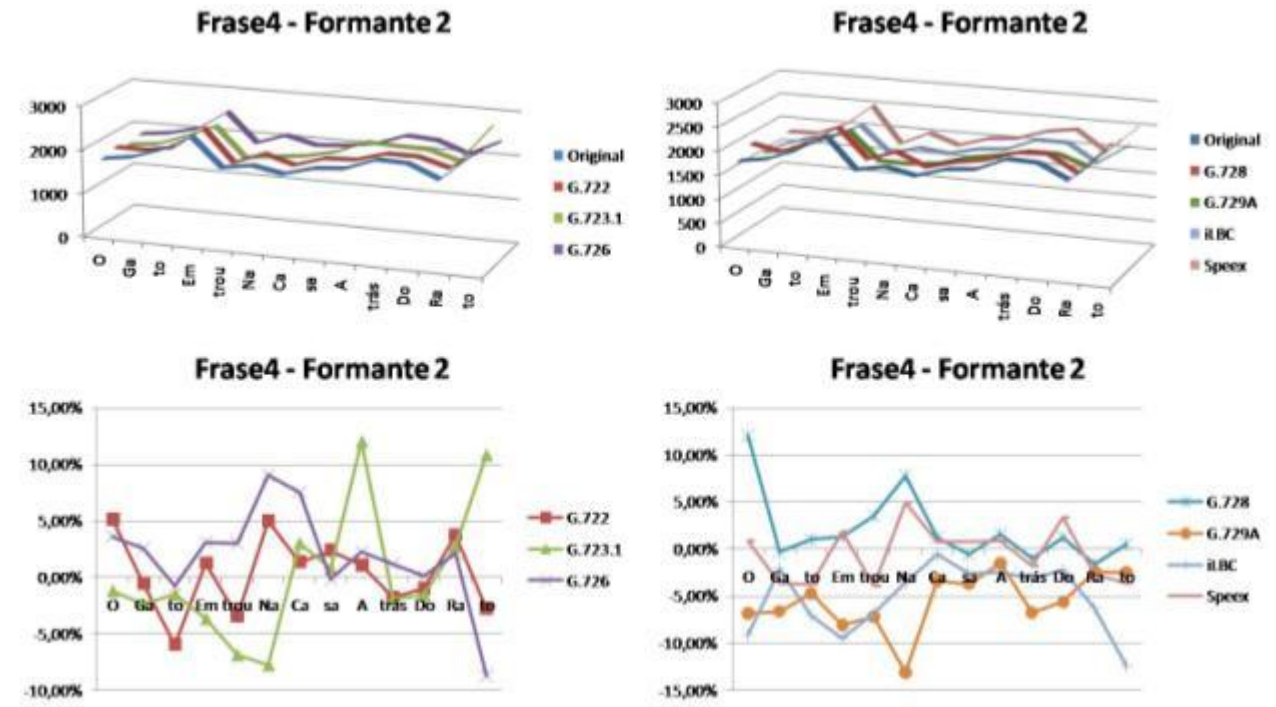


Figura 106 - Variações da frequência da segunda formante da frase 4 do indivíduo M3-48 do grupo MG2.

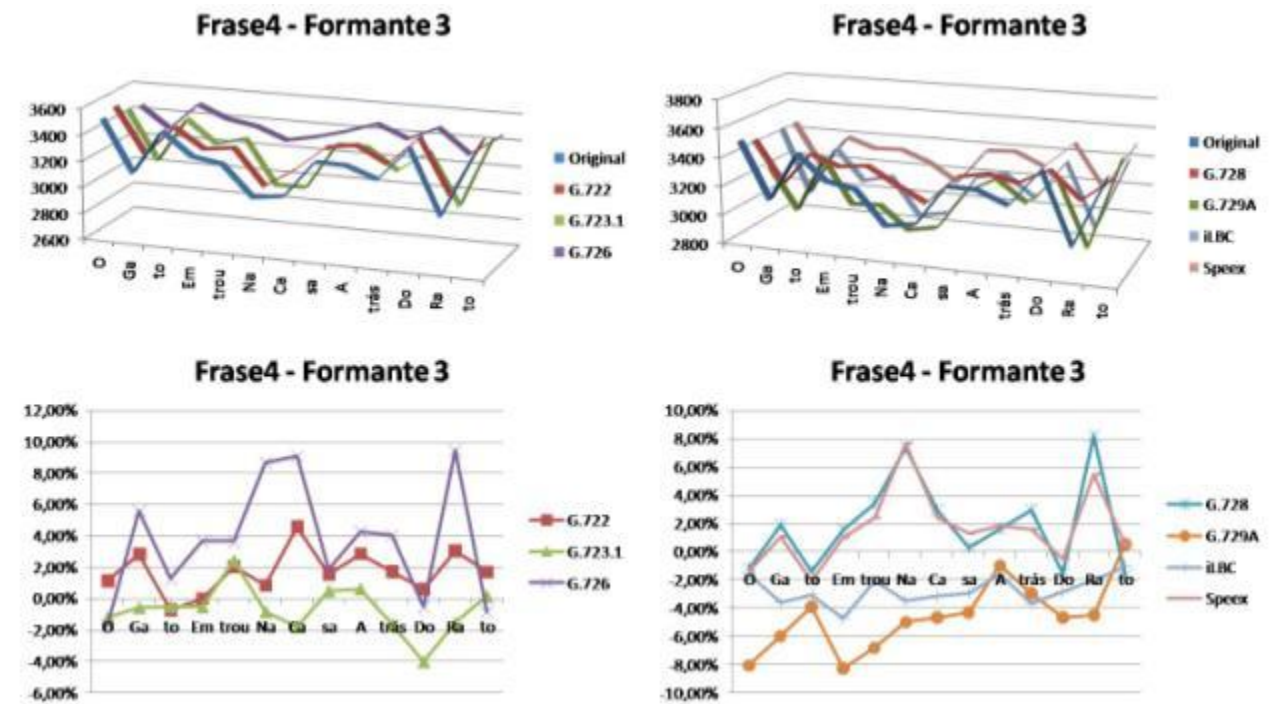


Figura 107 - Variações da frequência da terceira formante da frase 4 do indivíduo M3-48 do grupo MG2.

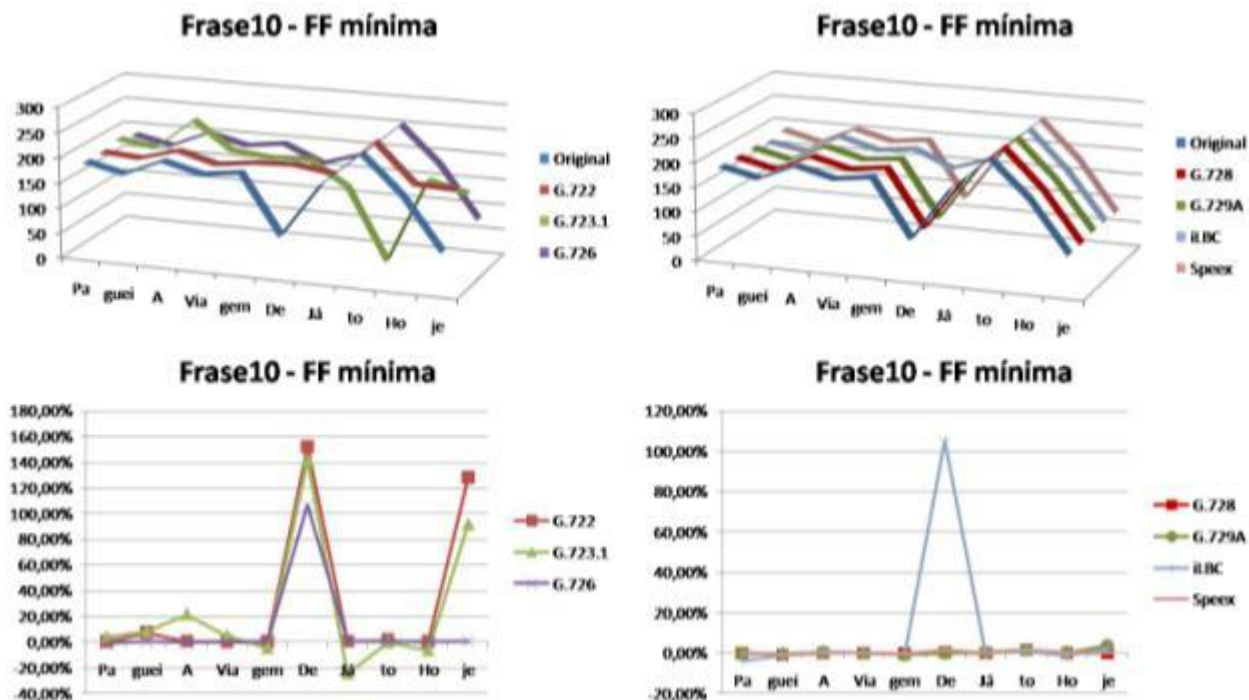


Figura 108 - Variações da frequência fundamental mínima da frase 10 do indivíduo M3-48 do grupo MG2.

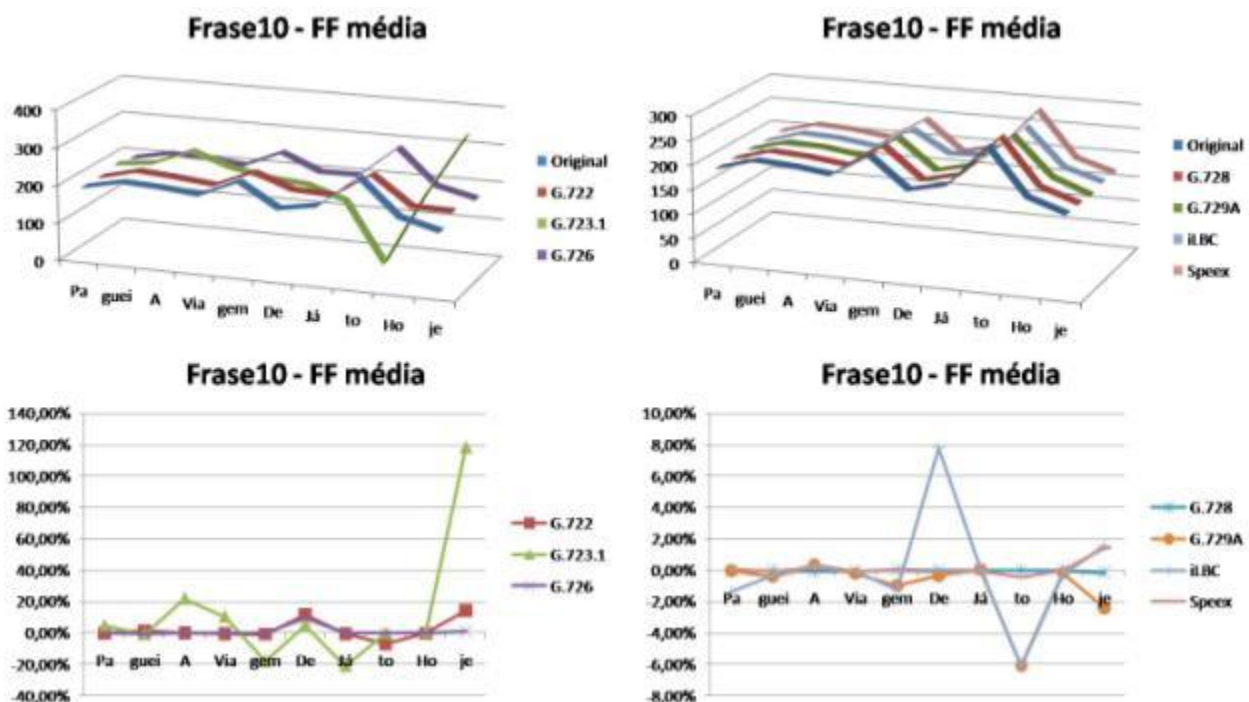


Figura 109 - Variações da frequência fundamental média da frase 10 do indivíduo M3-48 do grupo MG2.

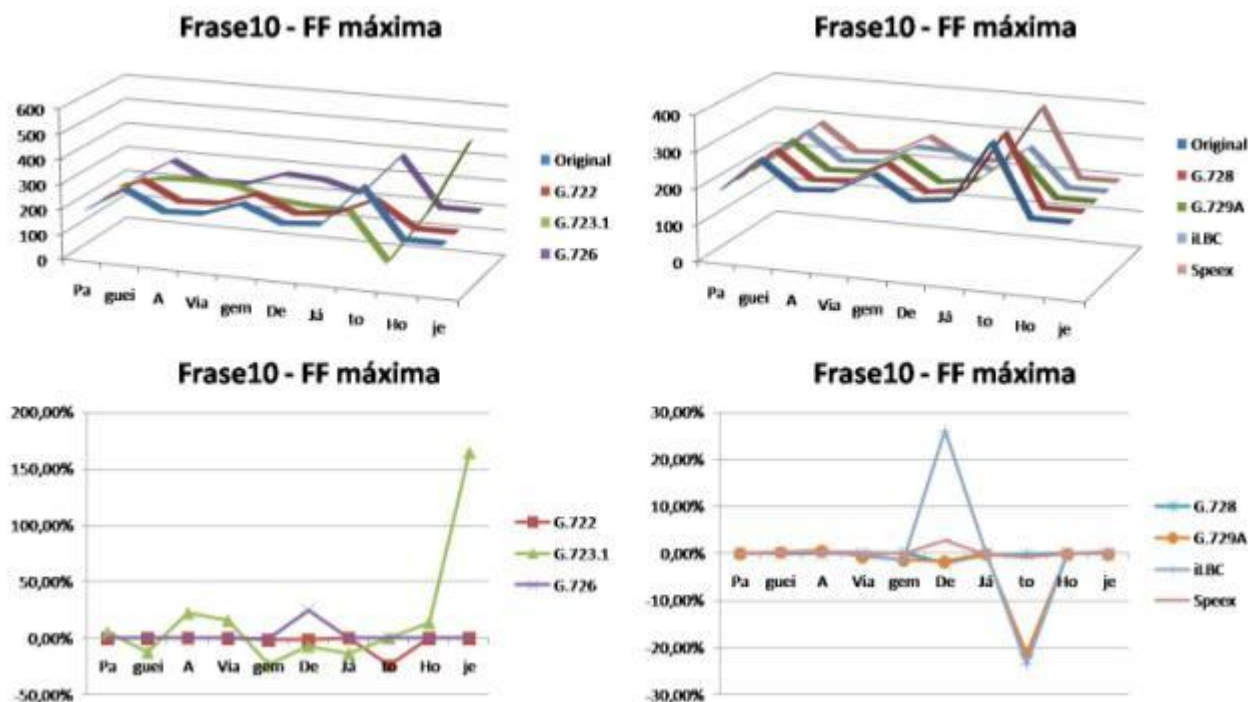


Figura 110 - Variações da frequência fundamental máxima da frase 10 do indivíduo M3-48 do grupo MG2.

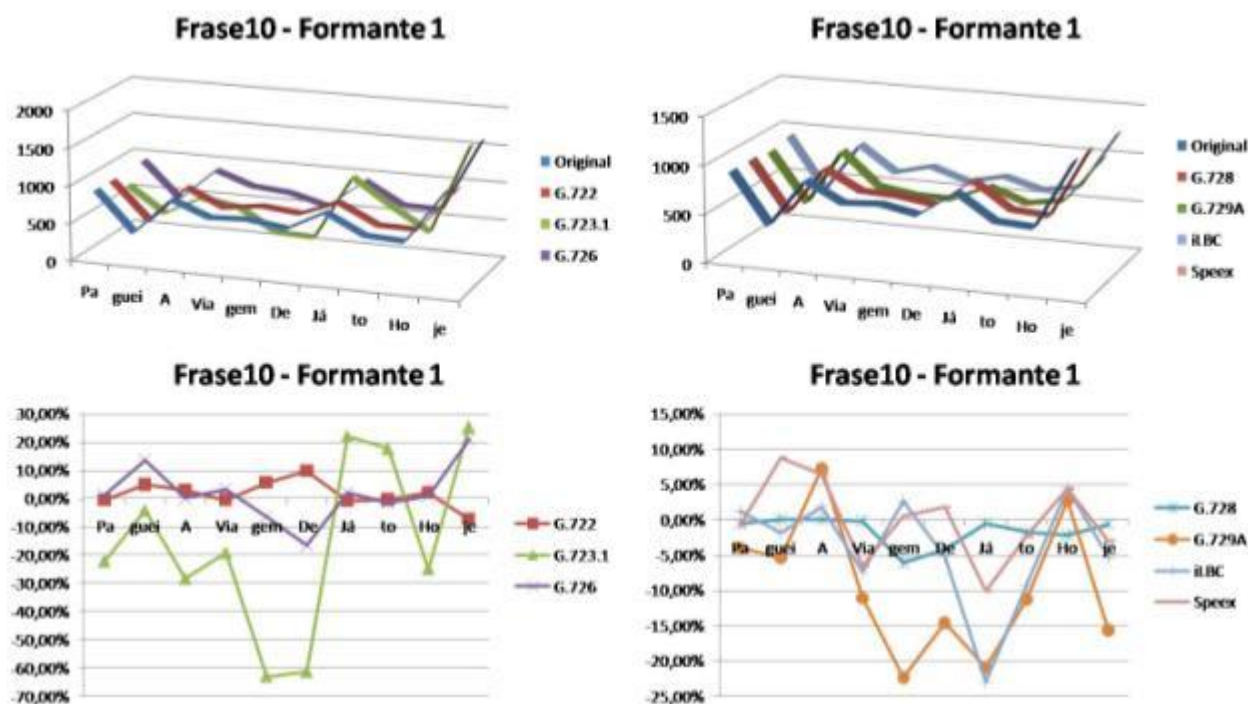


Figura 111 - Variações da frequência da primeira formante da frase 10 do indivíduo M3-48 do grupo MG2.

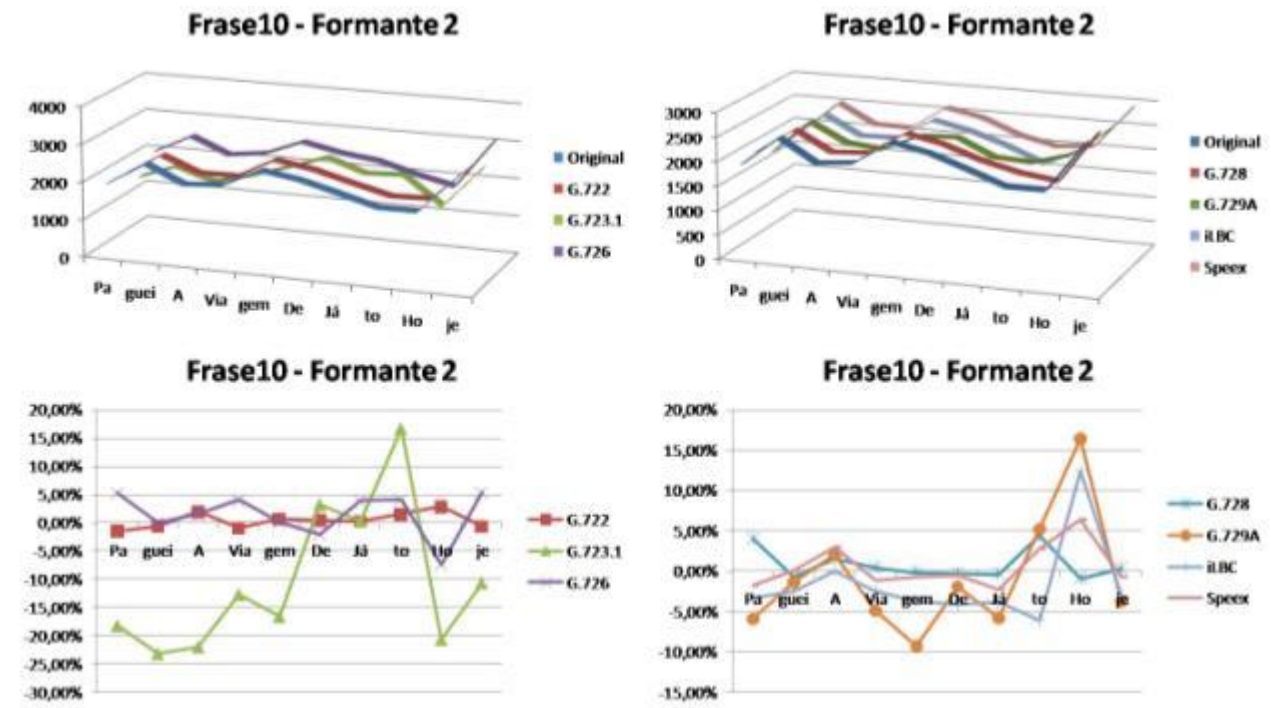


Figura 112 - Variações da freqüência da segunda formante da frase 10 do indivíduo M3-48 do grupo MG2.

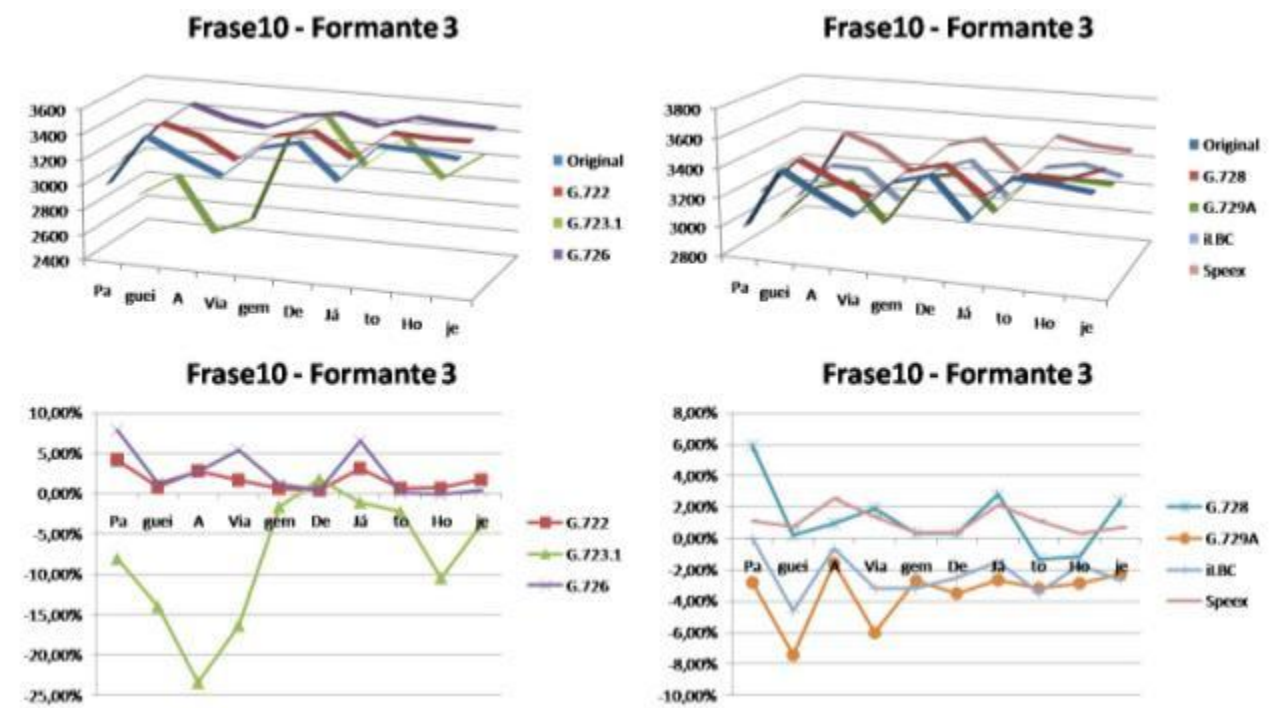


Figura 113 - Variações da freqüência da terceira formante da frase 10 do indivíduo M3-48 do grupo MG2.

4.4.6 – MG3

Fazem parte do grupo os indivíduos M2-78 e M7-48. O indivíduo M7-79 foi escolhido para ter seus resultados individuais por ter apresentado erros de processamento.

Neste grupo, o indivíduo M7-79 apresentou erros de processamento com a inserção de ruído e forte atenuação para o codificador G.723.1 nas frases 3 e 12. O indivíduo M2-78 também apresentou tais erros para esse codificador na frase 4.

Na tabela 37 temos a duração dos intervalos médios e arquivos originais analisados juntamente com as suas taxas de processamento para o indivíduo M7-79 e o grupo MG3. Pode-se observar que o codificador G.723.1 possui as piores taxas de tempo de processamento.

Tabela 37 - Duração dos intervalos médios e dos arquivos do grupo MG3 com suas taxas de tempo de processamento.

		Arquivos Originais		Taxa de tempo de processamento			
		Intervalo Médio (s)	Duração (s)	G.723.1	G.728	G.729A	iLBC
M7-79	Frase 1	0,251	3,016	19,89%	9,95%	3,32%	3,32%
	Frase 2	0,210	3,157	19,01%	9,50%	3,17%	6,34%
	Frase 3	0,232	2,085	23,98%	9,59%	4,80%	4,80%
	Frase 4	0,179	2,327	21,49%	12,89%	4,30%	4,30%
	Frase 5	0,283	2,544	23,58%	11,79%	3,93%	3,93%
	Frase 6	0,286	2,287	21,86%	8,74%	4,37%	8,74%
	Frase 7	0,308	3,387	20,67%	8,86%	2,95%	5,91%
	Frase 8	0,212	2,759	21,75%	7,25%	3,62%	3,62%
	Frase 9	0,231	2,774	21,63%	10,81%	3,60%	3,60%
	Frase 10	0,263	2,633	22,78%	11,39%	3,80%	3,80%
	Frase 11	0,293	2,931	20,47%	10,24%	3,41%	3,41%
	Frase 12	0,256	3,334	21,00%	9,00%	3,00%	3,00%
MG3	Frase 1	0,241	2,891	20,79%	10,39%	3,46%	5,27%
	Frase 2	0,223	3,338	19,45%	9,01%	3,00%	4,59%
	Frase 3	0,221	1,991	22,53%	10,07%	2,40%	5,03%
	Frase 4	0,198	2,569	21,41%	11,78%	3,93%	5,71%
	Frase 5	0,258	2,322	23,70%	10,66%	4,35%	4,35%
	Frase 6	0,243	1,947	20,26%	10,59%	2,19%	7,48%
	Frase 7	0,263	2,898	20,71%	8,58%	3,55%	7,10%
	Frase 8	0,201	2,518	21,85%	8,02%	4,01%	4,01%
	Frase 9	0,222	2,663	22,57%	11,29%	3,76%	3,76%
	Frase 10	0,235	2,349	21,08%	10,54%	4,32%	4,32%
	Frase 11	0,252	2,522	19,70%	9,85%	4,07%	4,07%
	Frase 12	0,230	2,992	21,82%	10,16%	3,39%	3,39%

Nas tabelas 38 e 39 temos os erros de processamento das formantes para cada frase para o indivíduo M7-79 e médios do grupo MG3. Observa-se a presença de uma tendência no grupo a maiores erros na frase 12.

Tabela 38 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M7-79 e o grupo MG3.

		G.722			G.723.1			G.726		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M7-79	Frase 1	1,11%	-4,67%	-9,38%	0,31%	-3,10%	-7,32%	0,29%	0,12%	0,16%
	Frase 2	0,94%	-4,58%	-9,41%	-0,58%	-3,24%	-7,19%	-1,02%	-0,60%	-0,15%
	Frase 3	2,31%	-0,97%	-2,11%	2,86%	-0,57%	-3,30%	-2,64%	0,20%	1,67%
	Frase 4	-3,38%	-3,41%	-3,75%	-0,19%	0,13%	0,19%	-3,59%	-3,51%	-3,65%
	Frase 5	1,01%	-4,95%	-8,86%	-3,10%	-2,14%	0,61%	-3,10%	-2,14%	0,61%
	Frase 6	0,71%	-0,85%	-6,62%	12,65%	11,25%	1,87%	12,65%	11,25%	1,87%
	Frase 7	2,56%	-1,42%	-5,68%	-0,27%	-0,70%	1,67%	-0,27%	-0,70%	1,67%
	Frase 8	2,61%	0,56%	0,07%	-0,62%	4,02%	6,94%	1,17%	0,34%	1,02%
	Frase 9	0,44%	-1,27%	-3,03%	-0,64%	-0,41%	-1,16%	-0,64%	-0,41%	-1,16%
	Frase 10	39,35%	40,07%	52,67%	46,84%	51,13%	71,17%	46,84%	51,13%	71,17%
	Frase 11	10,71%	19,70%	11,27%	8,78%	20,70%	25,12%	8,78%	20,70%	25,12%
	Frase 12	25,97%	29,88%	28,75%	29,05%	32,22%	28,39%	22,37%	34,03%	54,60%
MG3	Frase 1	1,99%	-2,25%	-5,53%	-1,00%	-1,59%	-3,65%	-0,94%	0,03%	0,11%
	Frase 2	4,51%	-2,47%	-7,03%	-0,31%	-2,48%	-3,61%	-0,88%	-0,05%	-0,91%
	Frase 3	8,72%	0,96%	-4,02%	8,53%	3,16%	-0,39%	5,54%	3,11%	0,85%
	Frase 4	0,51%	-1,73%	-3,36%	5,33%	0,50%	-2,28%	-1,92%	-1,73%	-1,88%
	Frase 5	1,43%	-2,26%	-4,97%	-0,87%	-1,01%	0,70%	-1,51%	-1,01%	0,41%
	Frase 6	1,15%	-0,18%	-3,37%	6,48%	5,70%	1,12%	6,61%	5,64%	1,01%
	Frase 7	3,16%	0,09%	-4,12%	-0,17%	-0,21%	1,01%	0,12%	-0,23%	1,01%
	Frase 8	3,38%	-0,04%	-0,32%	-0,31%	2,18%	3,44%	2,39%	0,35%	0,52%
	Frase 9	1,46%	-2,60%	-4,26%	-1,13%	-2,39%	-3,77%	0,01%	0,02%	-0,55%
	Frase 10	20,40%	20,67%	26,48%	23,39%	25,35%	35,83%	23,53%	25,44%	36,92%
	Frase 11	5,40%	9,93%	4,11%	4,10%	11,32%	13,19%	4,36%	11,05%	12,55%
	Frase 12	12,77%	14,09%	10,74%	14,63%	15,59%	11,94%	11,22%	16,36%	25,11%

Tabela 39 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M7-79 e o grupo MG3.

		G.728			G.729A			iLBC			Speex		
		Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental			Frequencia Fundamental		
		Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima	Mínima	Média	Máxima
M7-79	Frase 1	0,28%	-1,04%	0,19%	-0,14%	-3,15%	-5,39%	-0,31%	-3,71%	-5,68%	-0,02%	-3,04%	-5,53%
	Frase 2	-0,32%	-3,09%	-7,31%	0,11%	-3,82%	-6,68%	-0,52%	-2,57%	-7,29%	-0,17%	-1,97%	-5,57%
	Frase 3	-2,57%	0,24%	3,01%	-1,96%	-0,66%	-1,14%	-2,90%	-0,48%	-0,59%	-2,64%	-0,70%	-1,97%
	Frase 4	0,14%	-0,03%	-0,21%	-4,76%	-1,00%	0,28%	-2,99%	0,16%	1,94%	-7,27%	-4,67%	-3,76%
	Frase 5	-3,72%	-0,94%	-0,08%	10,21%	-4,17%	-12,86%	4,93%	-5,70%	-13,20%	-3,70%	-6,80%	-12,43%
	Frase 6	0,66%	-0,18%	-2,75%	13,14%	11,21%	-3,82%	13,05%	10,66%	-4,19%	12,04%	10,65%	-4,34%
	Frase 7	-0,33%	0,33%	3,86%	-3,73%	-3,00%	-9,65%	-5,78%	-3,02%	-8,75%	-0,32%	-1,93%	-5,82%
	Frase 8	0,09%	0,04%	-3,78%	0,56%	0,18%	0,60%	0,47%	0,10%	0,49%	0,02%	0,16%	-0,07%
	Frase 9	-0,56%	0,03%	-0,64%	-0,33%	-0,08%	-2,94%	0,01%	1,77%	3,96%	0,04%	0,11%	-0,01%
	Frase 10	37,66%	40,78%	55,51%	46,90%	43,87%	53,45%	37,26%	38,81%	55,97%	46,01%	48,02%	56,54%
	Frase 11	0,78%	27,10%	34,10%	7,34%	19,44%	11,90%	13,99%	20,58%	16,07%	8,99%	19,89%	13,61%
	Frase 12	22,43%	32,72%	42,95%	25,75%	29,83%	28,45%	24,33%	31,71%	45,44%	24,09%	29,29%	29,07%
MG3	Frase 1	-0,83%	-0,48%	0,14%	-1,14%	-1,41%	-1,74%	-11,45%	-13,39%	-14,18%	-11,27%	-13,00%	-14,20%
	Frase 2	-0,59%	-1,51%	-4,33%	0,39%	-4,45%	-7,59%	-2,12%	-6,64%	-10,96%	-1,83%	-5,93%	-8,60%
	Frase 3	5,66%	2,97%	1,08%	6,00%	1,43%	-0,41%	9,01%	1,43%	-5,17%	9,46%	1,24%	-6,53%
	Frase 4	0,02%	0,03%	-0,14%	-4,67%	-2,54%	-1,31%	-4,28%	-2,83%	-1,05%	-6,78%	-5,39%	-3,98%
	Frase 5	-2,15%	-0,58%	-0,16%	0,53%	-4,98%	-4,15%	-4,15%	-6,74%	-4,56%	-10,21%	-7,14%	-3,40%
	Frase 6	0,40%	0,06%	0,21%	6,13%	7,16%	2,79%	6,06%	6,67%	2,00%	5,27%	6,70%	1,83%
	Frase 7	-0,15%	0,23%	1,89%	-4,95%	-0,43%	-0,84%	-6,69%	-0,43%	-0,02%	-1,69%	0,69%	2,53%
	Frase 8	1,84%	0,20%	-1,87%	-0,38%	-1,19%	0,77%	-0,45%	-1,28%	0,63%	-0,85%	-1,19%	0,08%
	Frase 9	-0,24%	0,91%	0,83%	1,96%	-3,08%	-5,86%	2,30%	-1,24%	0,40%	2,22%	-2,93%	-4,15%
	Frase 10	18,88%	20,36%	28,85%	22,47%	18,64%	25,34%	14,24%	14,08%	26,97%	22,09%	20,86%	27,12%
	Frase 11	0,88%	14,07%	16,96%	-2,93%	5,97%	1,37%	3,44%	7,10%	5,34%	-1,01%	6,48%	2,98%
	Frase 12	10,01%	15,27%	24,32%	3,96%	7,19%	6,25%	2,80%	8,92%	19,83%	2,63%	6,77%	6,40%

Nas tabelas 40 e 41 temos os erros de processamento das formantes para cada frase para o indivíduo M7-79 e médios do grupo MG3. Observa-se, também, a presença de uma tendência no grupo a maiores erros na frase 12.

Tabela 40 - Erros de processamento das formantes dos codificadores G.722, G.723.1 e G726 para o indivíduo M7-79 e o grupo MG3.

		G.722			G.723.1			G.726		
		Fomantes			Fomantes			Fomantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M7-79	Frase 1	-3,17%	-0,24%	2,52%	-1,50%	-0,56%	3,31%	-0,18%	1,50%	4,10%
	Frase 2	4,20%	1,63%	1,84%	7,26%	10,11%	2,60%	3,67%	1,81%	3,58%
	Frase 3	-0,35%	0,79%	2,91%	-8,21%	5,66%	3,15%	-2,14%	2,87%	5,08%
	Frase 4	-3,43%	-2,07%	3,03%	-7,99%	2,01%	2,09%	3,85%	3,17%	5,00%
	Frase 5	0,53%	0,15%	1,48%	-1,59%	3,53%	0,32%	-1,59%	3,53%	0,32%
	Frase 6	3,84%	0,19%	2,20%	-7,32%	4,05%	0,93%	-7,32%	4,05%	0,93%
	Frase 7	-1,39%	-0,50%	2,55%	-4,62%	0,19%	0,76%	-4,62%	0,19%	0,76%
	Frase 8	-1,96%	-1,14%	2,57%	-2,45%	2,89%	2,71%	-5,81%	-0,90%	1,98%
	Frase 9	2,37%	1,71%	3,04%	-4,24%	2,86%	1,41%	-4,24%	2,86%	1,41%
	Frase 10	15,94%	-3,68%	-4,60%	13,77%	-0,10%	-7,59%	13,77%	-0,10%	-7,59%
	Frase 11	-1,04%	-1,85%	-1,66%	-6,93%	1,89%	-2,99%	-6,93%	1,89%	-2,99%
	Frase 12	27,18%	-3,50%	-4,72%	18,36%	-4,60%	-8,04%	28,44%	-2,47%	-2,74%
MG3	Frase 1	1,04%	2,98%	0,68%	-2,72%	1,07%	2,78%	-1,69%	1,64%	4,42%
	Frase 2	2,18%	1,19%	1,90%	1,10%	7,22%	2,64%	2,07%	1,60%	3,56%
	Frase 3	0,43%	1,11%	3,89%	-8,64%	3,01%	2,25%	-0,39%	2,39%	5,85%
	Frase 4	-1,38%	-0,23%	2,92%	-5,28%	2,95%	1,66%	3,26%	2,73%	4,75%
	Frase 5	0,60%	0,55%	1,49%	-5,25%	2,39%	0,52%	-1,47%	2,29%	1,10%
	Frase 6	0,73%	0,55%	2,94%	-7,67%	1,77%	0,75%	-4,57%	2,99%	2,87%
	Frase 7	0,77%	0,74%	3,24%	-0,92%	0,15%	0,63%	-2,58%	1,04%	3,22%
	Frase 8	-0,45%	0,65%	3,33%	-2,66%	1,46%	3,10%	-2,21%	3,05%	3,82%
	Frase 9	3,11%	1,45%	3,11%	-2,85%	2,01%	2,00%	0,85%	3,19%	1,47%
	Frase 10	8,52%	-1,67%	0,42%	0,80%	-0,11%	-2,75%	11,42%	1,20%	0,36%
	Frase 11	0,47%	0,14%	1,36%	-5,29%	2,80%	0,08%	-3,99%	2,54%	0,99%
	Frase 12	13,74%	-1,14%	-0,76%	6,77%	-1,81%	-3,04%	14,17%	-0,18%	1,60%

Tabela 41 - Erros de processamento das formantes dos codificadores G.728, G.729A, iLBC e Speex para o indivíduo M7-79 e o grupo MG3.

		G.728			G.729A			iLBC			Speex		
		Fomantes			Fomantes			Fomantes			Fomantes		
		Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira	Primeira	Segunda	Terceira
M7-79	Frase 1	-0,21%	2,25%	3,86%	-10,73%	-5,63%	-3,79%	-8,42%	-6,41%	-4,17%	1,94%	0,67%	4,03%
	Frase 2	6,54%	6,60%	2,60%	-6,47%	-0,49%	-2,36%	-4,43%	-5,08%	-3,08%	-2,82%	5,14%	3,11%
	Frase 3	0,87%	4,90%	4,28%	-7,73%	-4,53%	-2,70%	-7,38%	-4,38%	-2,79%	-2,11%	2,73%	4,23%
	Frase 4	0,58%	2,91%	3,91%	-9,90%	-5,71%	-3,96%	-4,39%	-8,29%	-3,93%	-2,97%	0,65%	3,44%
	Frase 5	-3,31%	0,39%	1,46%	-8,42%	-0,13%	-1,83%	-6,92%	-5,68%	-2,26%	-0,30%	1,88%	2,48%
	Frase 6	-1,20%	2,13%	2,11%	-11,38%	-0,84%	-4,10%	-5,74%	-3,80%	-3,22%	-3,29%	0,00%	2,68%
	Frase 7	-1,05%	0,69%	3,35%	-11,86%	-6,47%	-5,30%	-8,62%	-4,53%	-3,13%	-3,99%	-0,95%	3,10%
	Frase 8	1,99%	2,37%	4,41%	-11,21%	-5,25%	-3,60%	-2,60%	-4,93%	-2,51%	0,49%	1,38%	4,56%
	Frase 9	-3,10%	1,13%	4,06%	-10,23%	-3,08%	-3,40%	-4,58%	-3,57%	-2,32%	-0,41%	0,84%	4,14%
	Frase 10	8,34%	-2,02%	-4,50%	-2,07%	-7,26%	-11,61%	4,76%	-9,46%	-9,88%	4,63%	-5,18%	-4,19%
	Frase 11	-7,10%	-0,22%	-1,66%	-7,19%	-4,06%	-7,10%	-6,64%	-7,21%	-5,98%	-3,33%	0,13%	-1,05%
	Frase 12	24,06%	-3,65%	-3,96%	18,69%	-7,89%	-12,33%	17,97%	-8,72%	-11,85%	24,34%	-4,02%	-3,75%
MG3	Frase 1	-1,52%	2,40%	4,25%	-9,25%	-5,99%	-4,59%	-10,37%	-5,27%	-4,09%	-0,36%	2,02%	4,17%
	Frase 2	3,81%	4,98%	3,08%	-4,80%	-0,11%	-2,36%	-8,76%	-6,65%	-3,80%	-6,96%	-2,10%	2,33%
	Frase 3	0,66%	4,24%	5,15%	-7,16%	-5,44%	-3,69%	-7,66%	-5,27%	-2,83%	-2,43%	1,68%	4,12%
	Frase 4	0,78%	3,17%	3,86%	-14,61%	-8,86%	-5,50%	-10,46%	-11,08%	-5,16%	-8,41%	-2,36%	-1,30%
	Frase 5	-1,96%	1,18%	1,66%	-12,63%	-2,14%	-2,43%	-10,81%	-7,12%	-2,86%	-5,18%	-0,08%	1,83%
	Frase 6	-3,51%	1,92%	3,21%	-12,08%	1,40%	-2,52%	-6,88%	-1,58%	-1,56%	-4,26%	2,17%	4,39%
	Frase 7	-2,88%	1,57%	4,57%	-8,95%	-4,62%	-4,46%	-5,67%	-2,51%	-2,24%	-0,78%	1,84%	4,08%
	Frase 8	1,21%	5,23%	5,00%	-8,84%	-2,09%	-3,63%	0,29%	-2,01%	-2,51%	3,40%	4,94%	4,44%
	Frase 9	-2,06%	1,97%	4,14%	-11,70%	-0,41%	-3,79%	-9,04%	-0,95%	-2,70%	-2,24%	3,70%	2,59%
	Frase 10	3,48%	0,61%	1,55%	-8,67%	-5,78%	-6,78%	-2,14%	-7,39%	-4,63%	-2,13%	-3,25%	0,98%
	Frase 11	-3,40%	1,95%	1,94%	-9,09%	-4,39%	-4,00%	-8,39%	-7,61%	-2,84%	-5,51%	-0,10%	2,30%
	Frase 12	15,40%	1,32%	0,73%	5,25%	-3,65%	-9,32%	4,91%	-4,48%	-8,84%	10,38%	0,43%	-0,44%

Nas figuras de 114 a 131 temos os resultados do indivíduo M7-79 para as frases 2, 4 e 10. Na parte superior da figura temos as variações absolutas das frequências

em Hertz por intervalo analisado de cada frase, enquanto na parte inferior as variações percentuais relativas ao arquivo original.

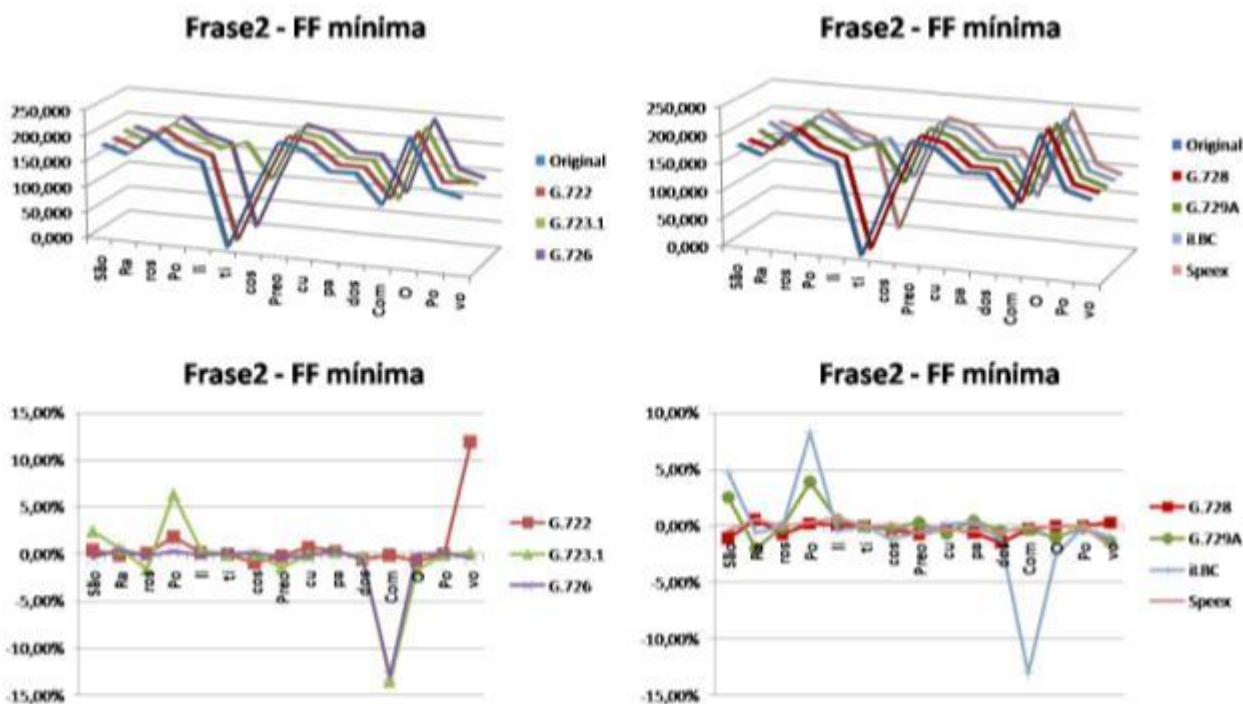


Figura 114 - Variações da frequência fundamental mínima da frase 2 do indivíduo M7-79 do grupo MG3.

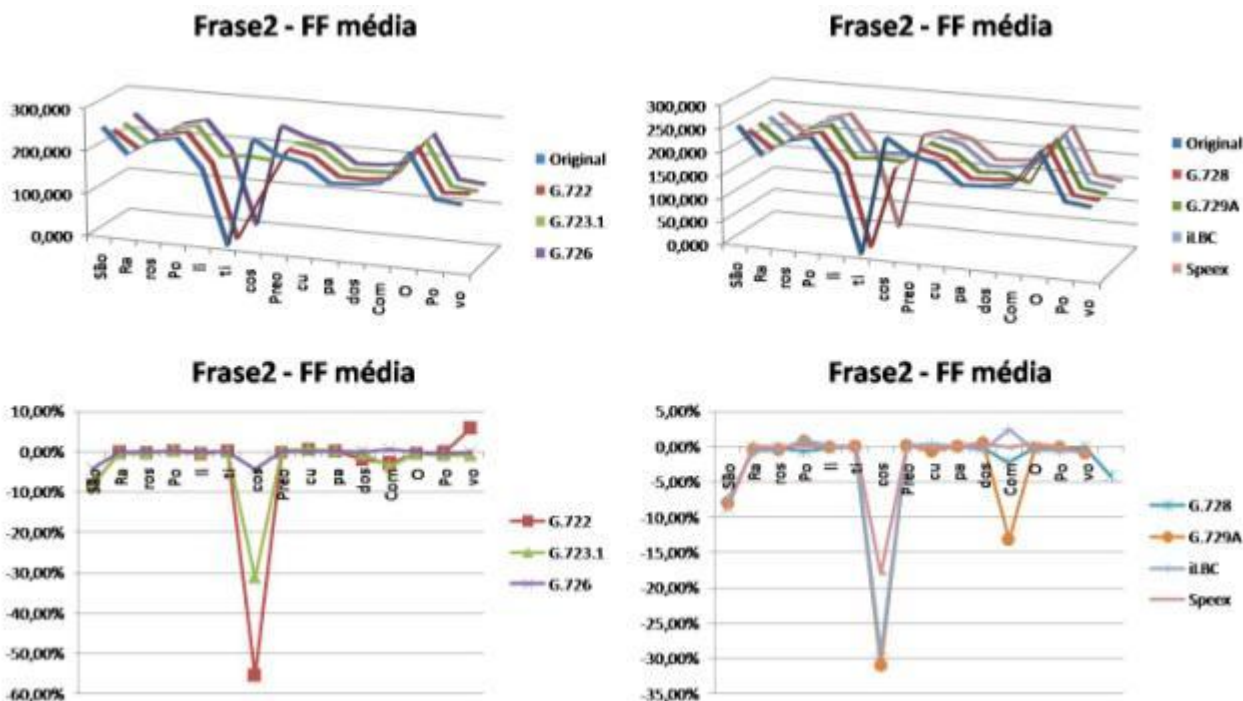


Figura 115 - Variações da frequência fundamental média da frase 2 do indivíduo M7-79 do grupo MG3.

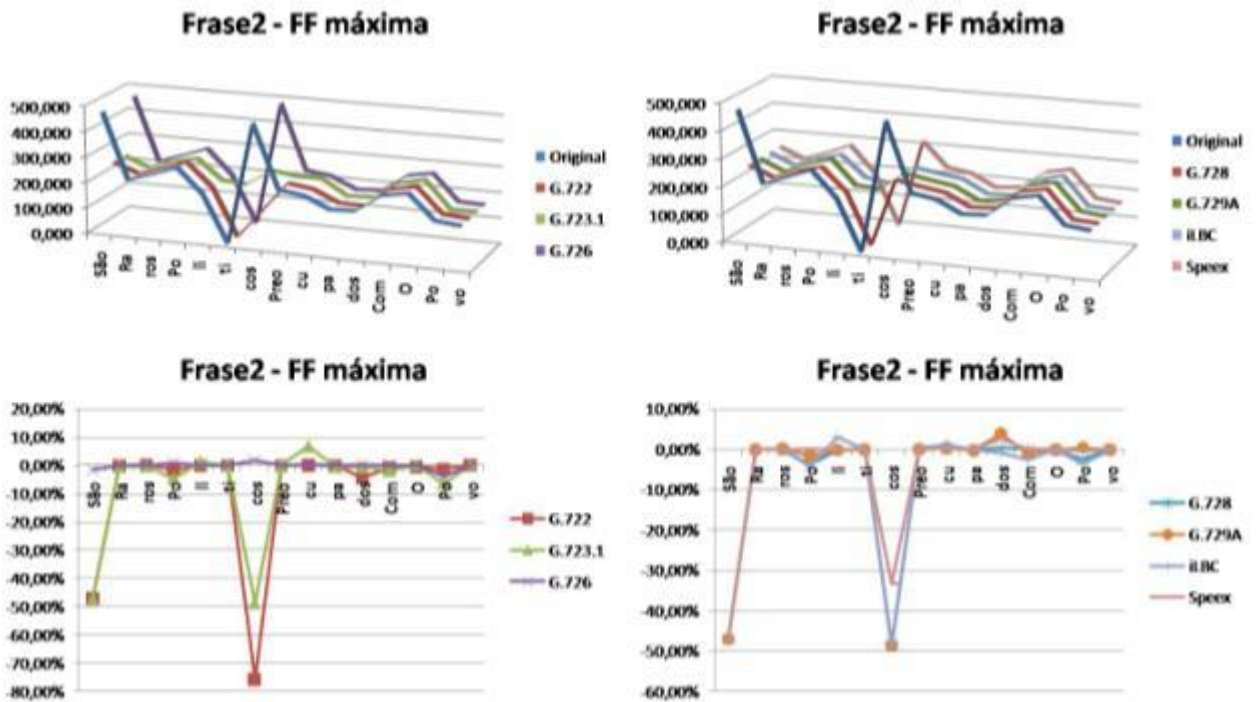


Figura 116 - Variações da frequência fundamental máxima da frase 2 do indivíduo M7-79 do grupo MG3.

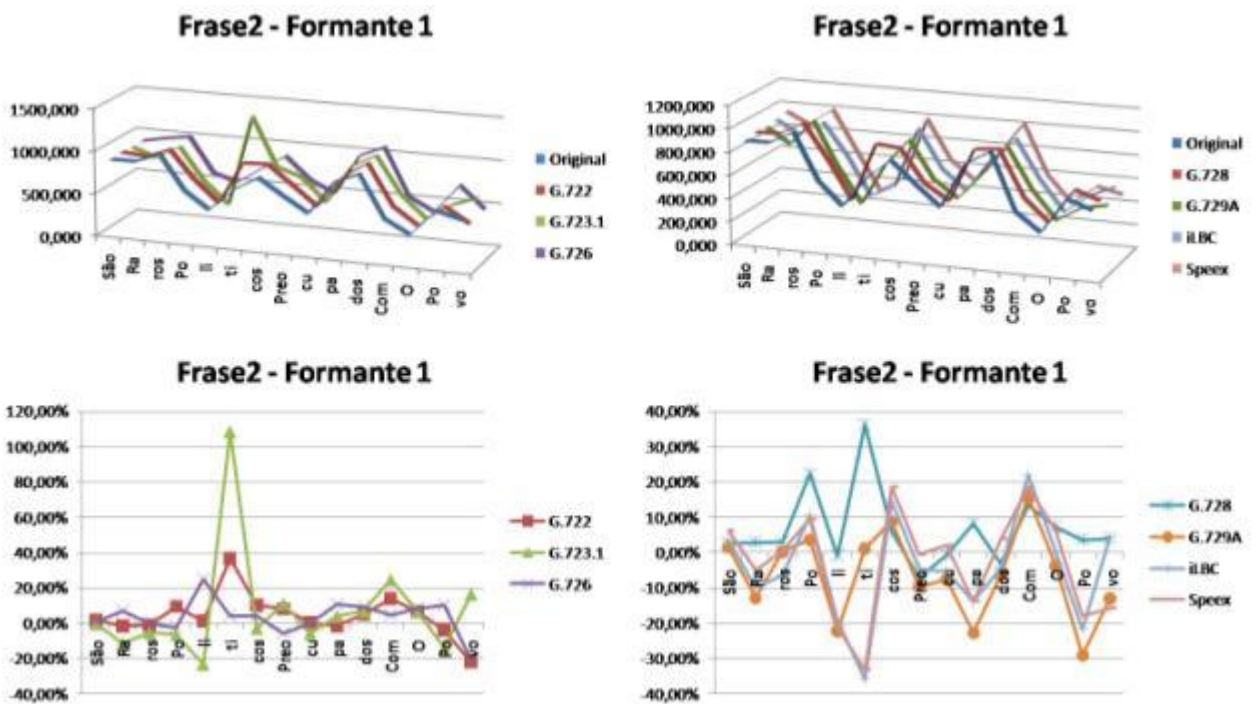


Figura 117 - Variações da frequência da primeira formante da frase 2 do indivíduo M7-79 do grupo MG3.

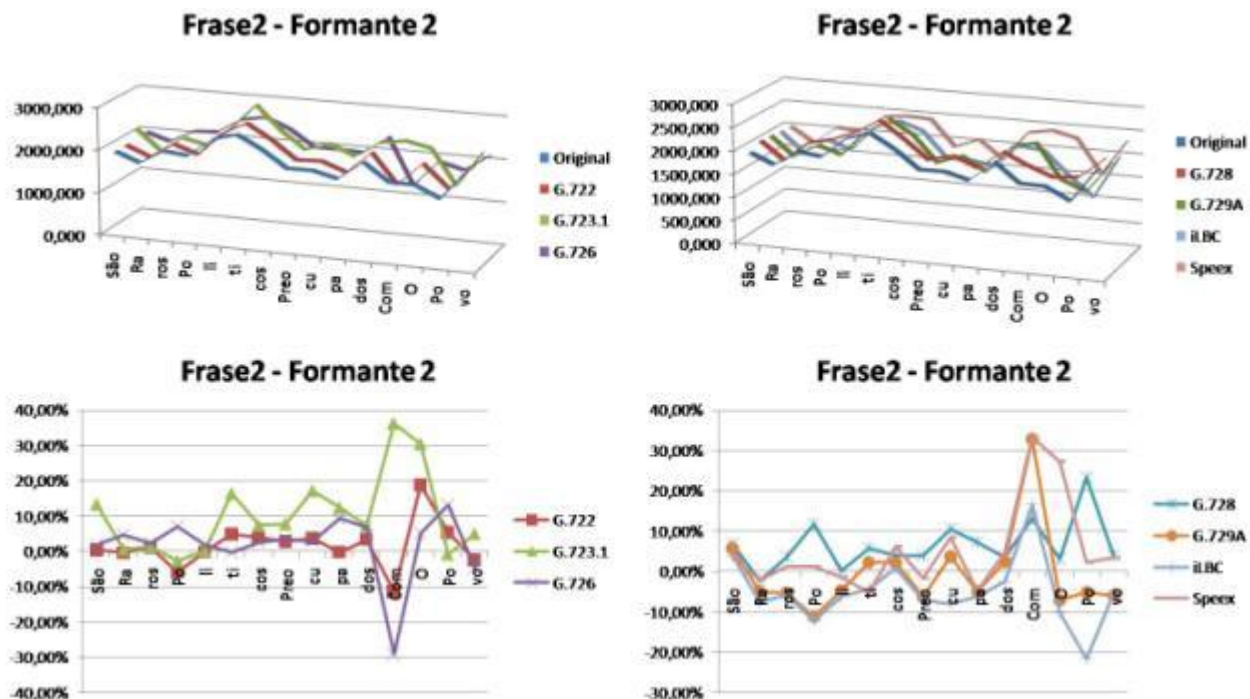


Figura 118 - Variações da frequência da segunda formante da frase 2 do indivíduo M7-79 do grupo MG3.

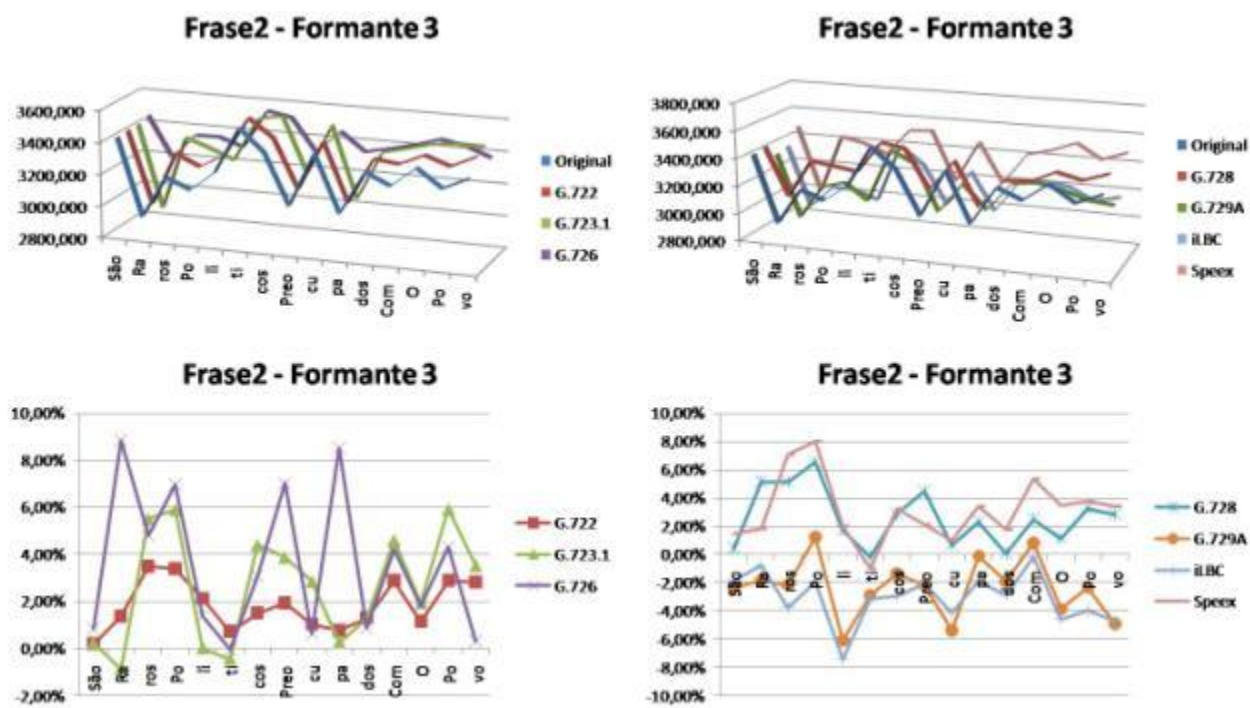


Figura 119 - Variações da frequência da terceira formante da frase 2 do indivíduo M7-79 do grupo MG3.

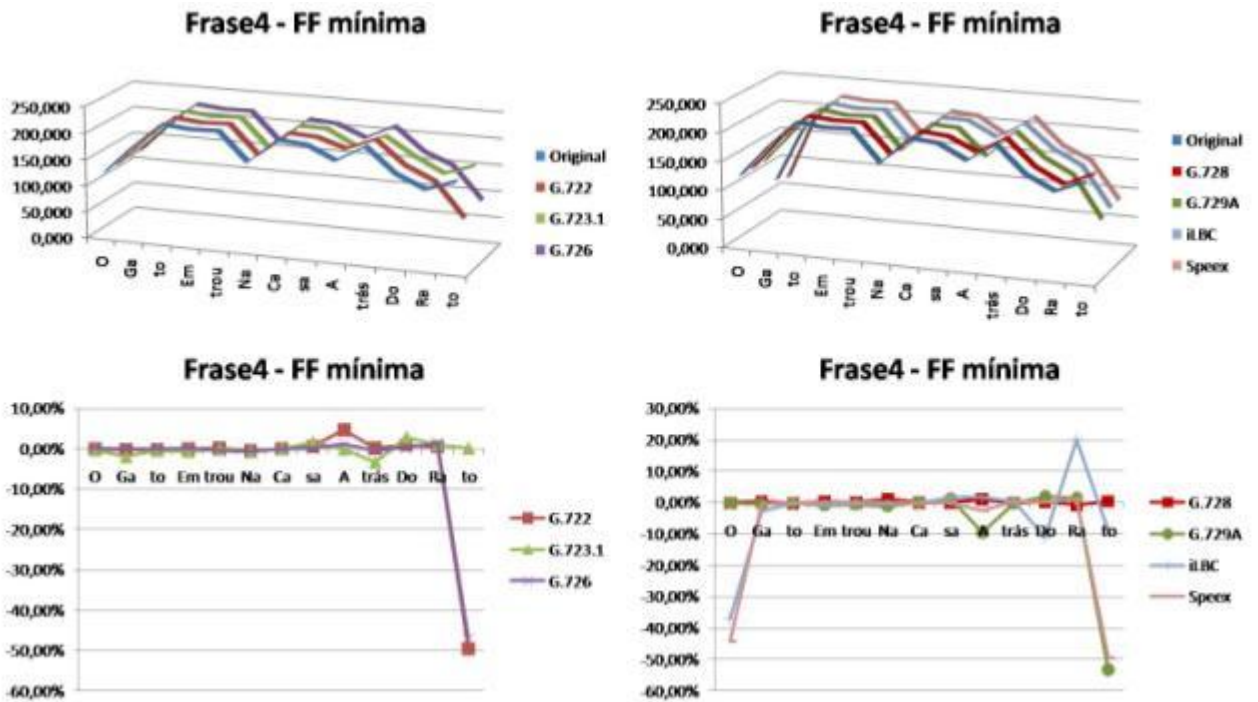


Figura 120 - Variações da freqüência fundamental mínima da frase 4 do indivíduo M7-79 do grupo MG3.

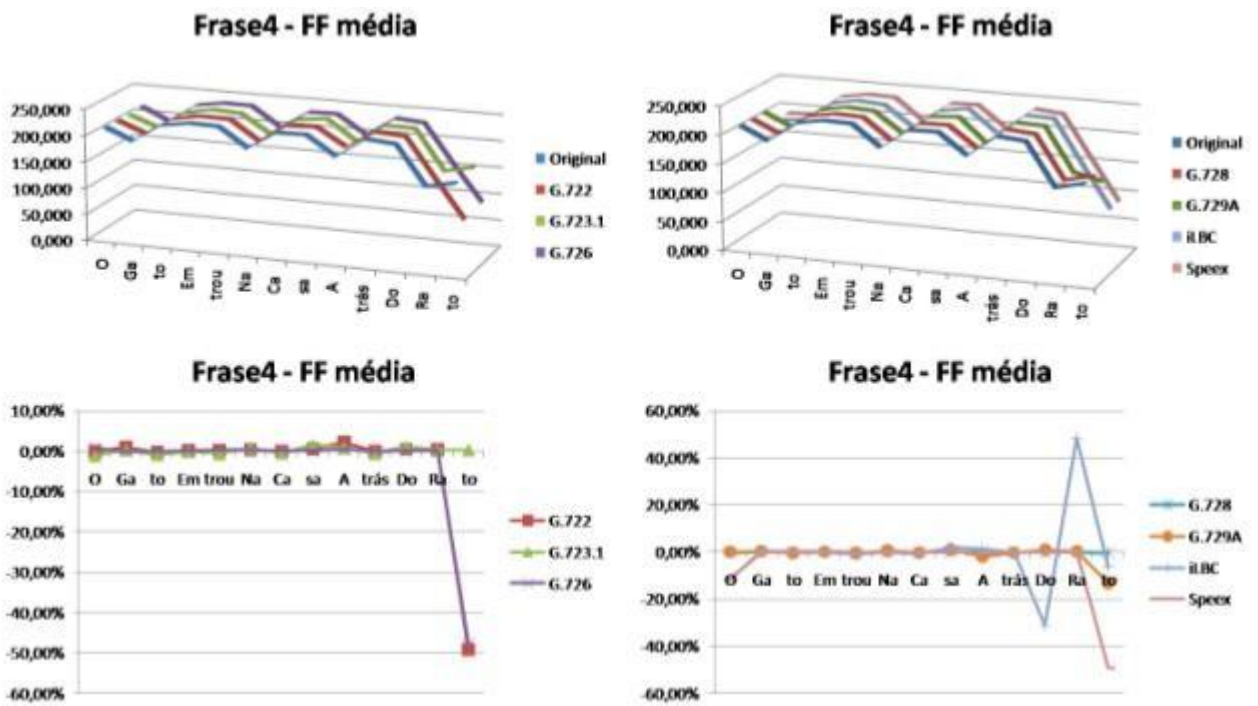


Figura 121 - Variações da freqüência fundamental média da frase 4 do indivíduo M7-79 do grupo MG3.

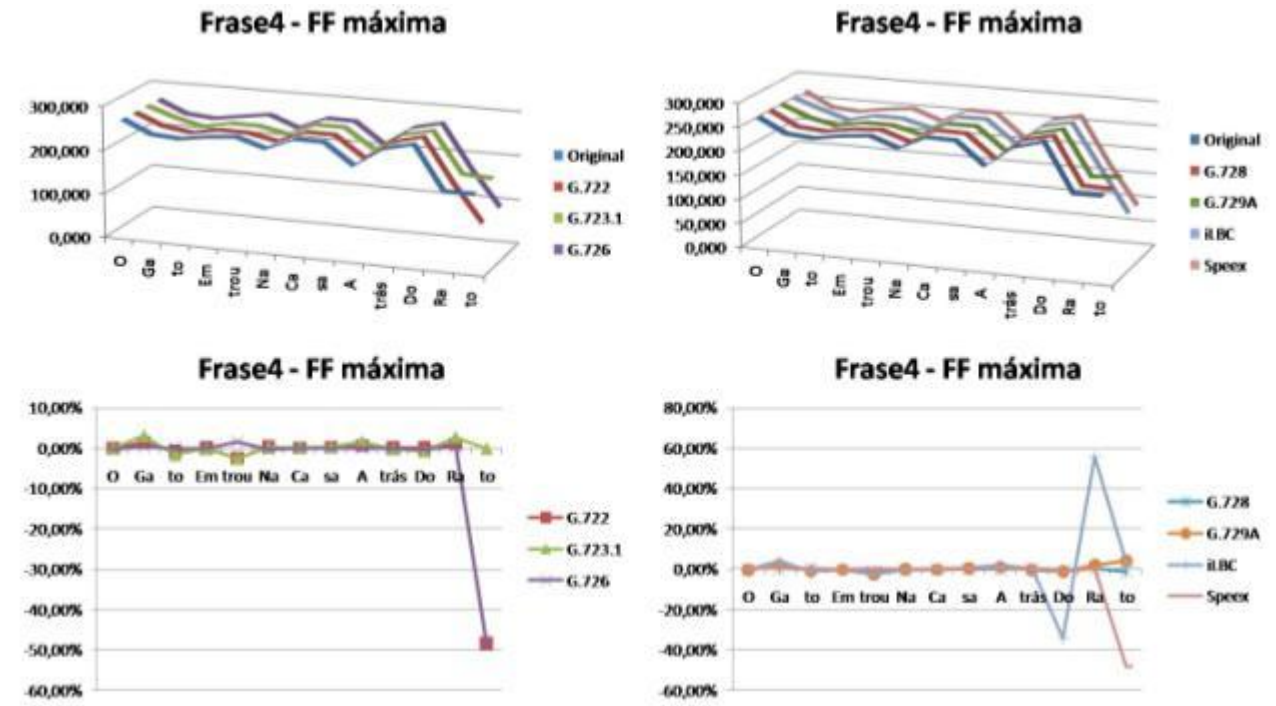


Figura 122 - Variações da freqüência fundamental máxima da frase 4 do indivíduo M7-79 do grupo MG3.

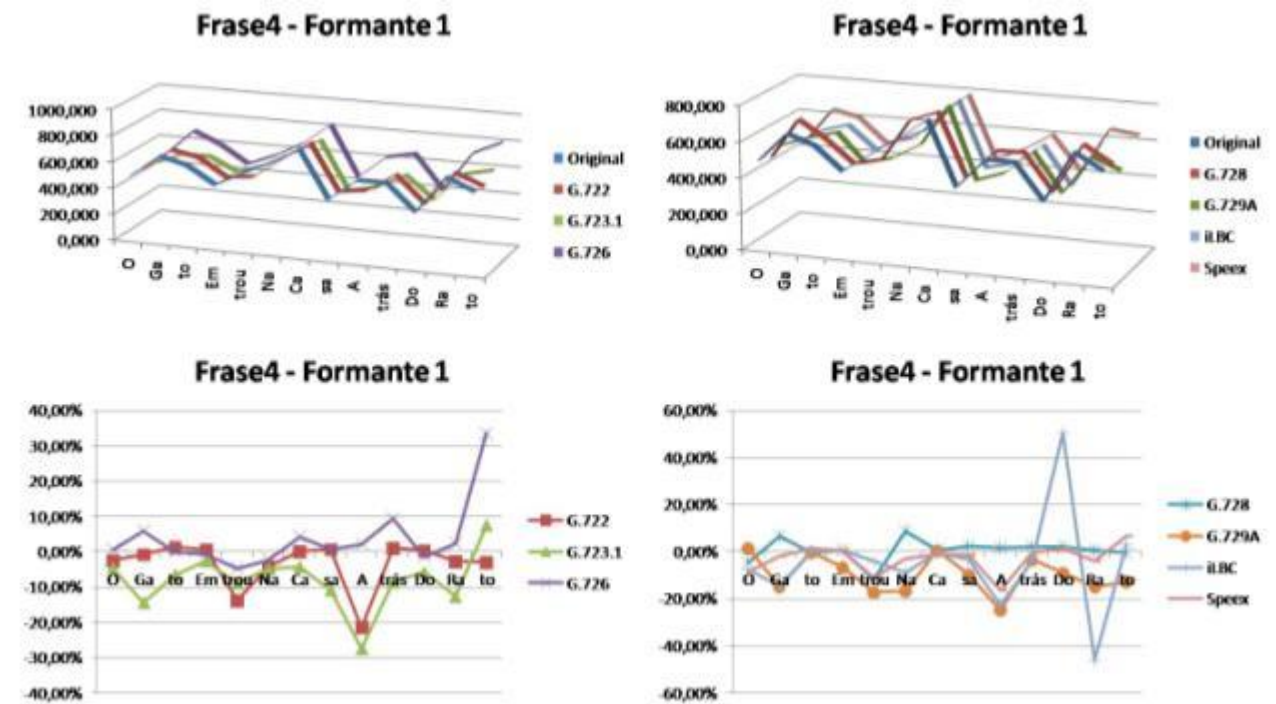


Figura 123 - Variações da freqüência da primeira formante da frase 4 do indivíduo M7-79 do grupo MG3.

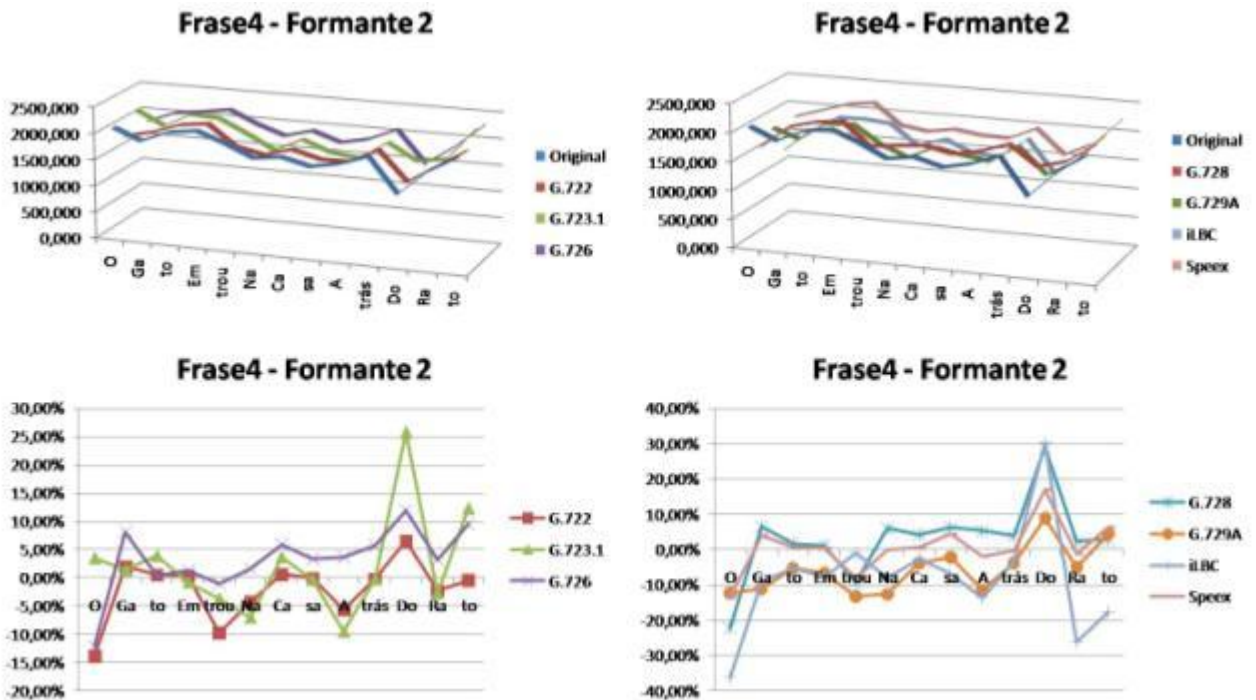


Figura 124 - Variações da frequência da segunda formante da frase 4 do indivíduo M7-79 do grupo MG3.

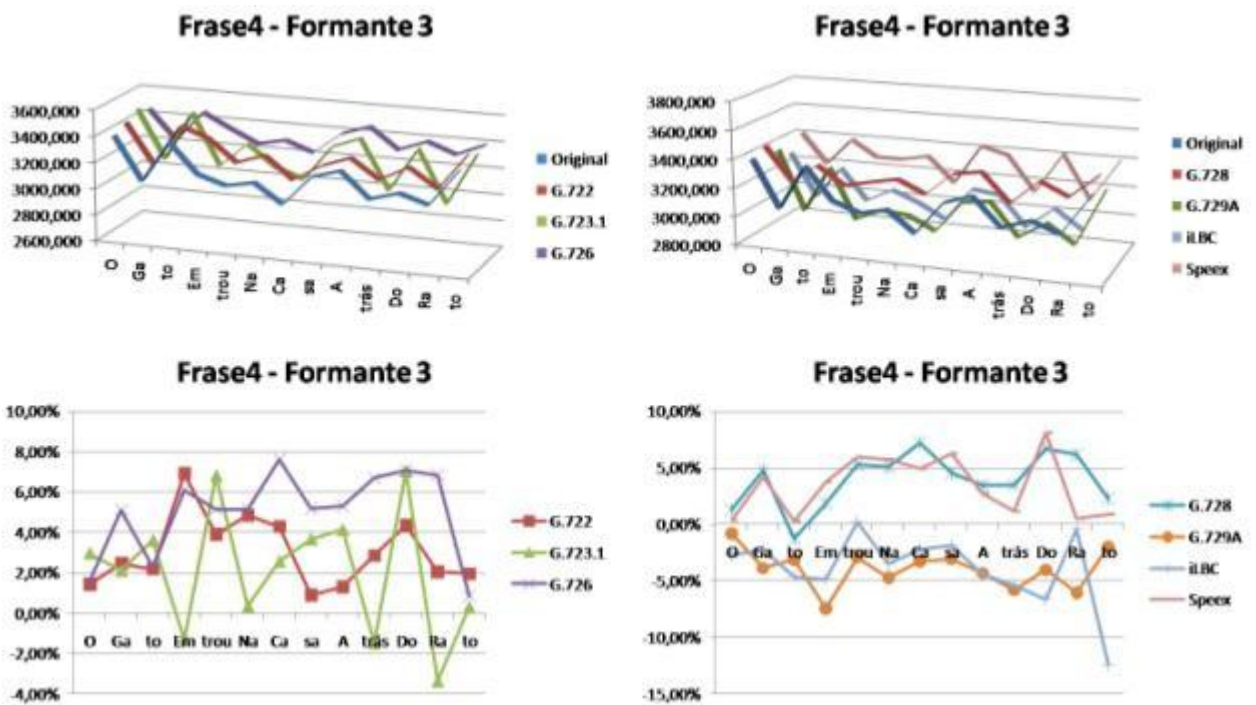


Figura 125 - Variações da frequência da terceira formante da frase 4 do indivíduo M7-79 do grupo MG3.

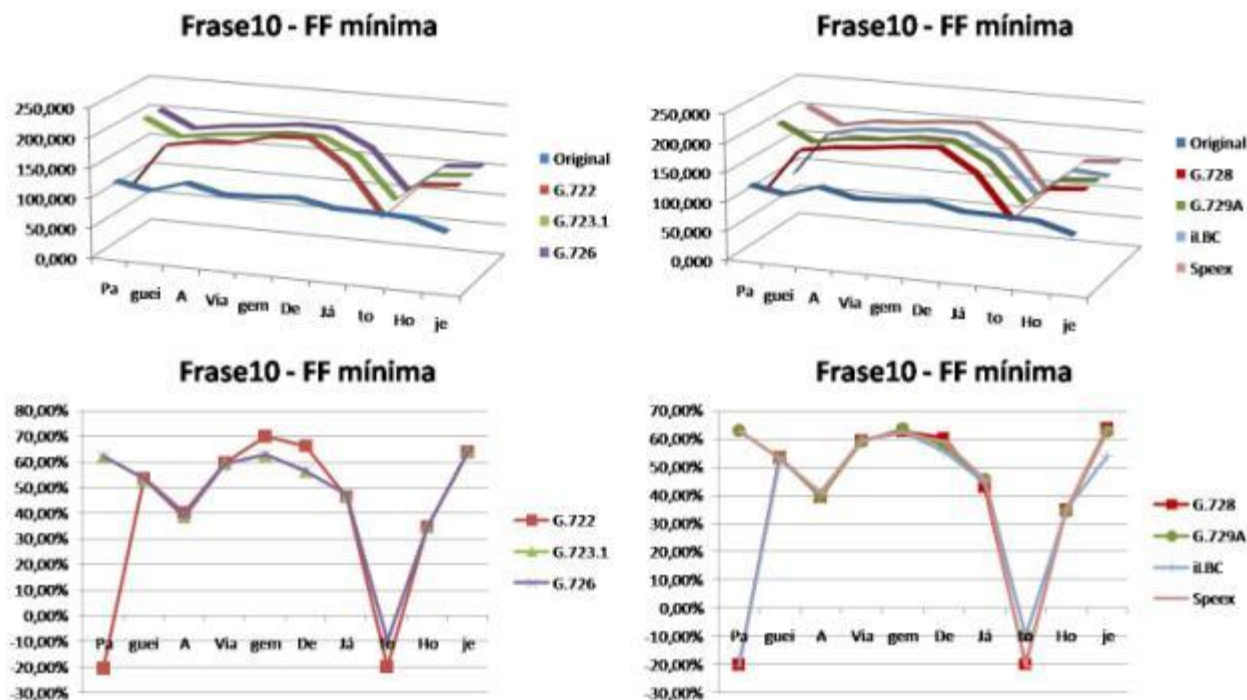


Figura 126 - Variações da frequência fundamental mínima da frase 10 do indivíduo M7-79 do grupo MG3.

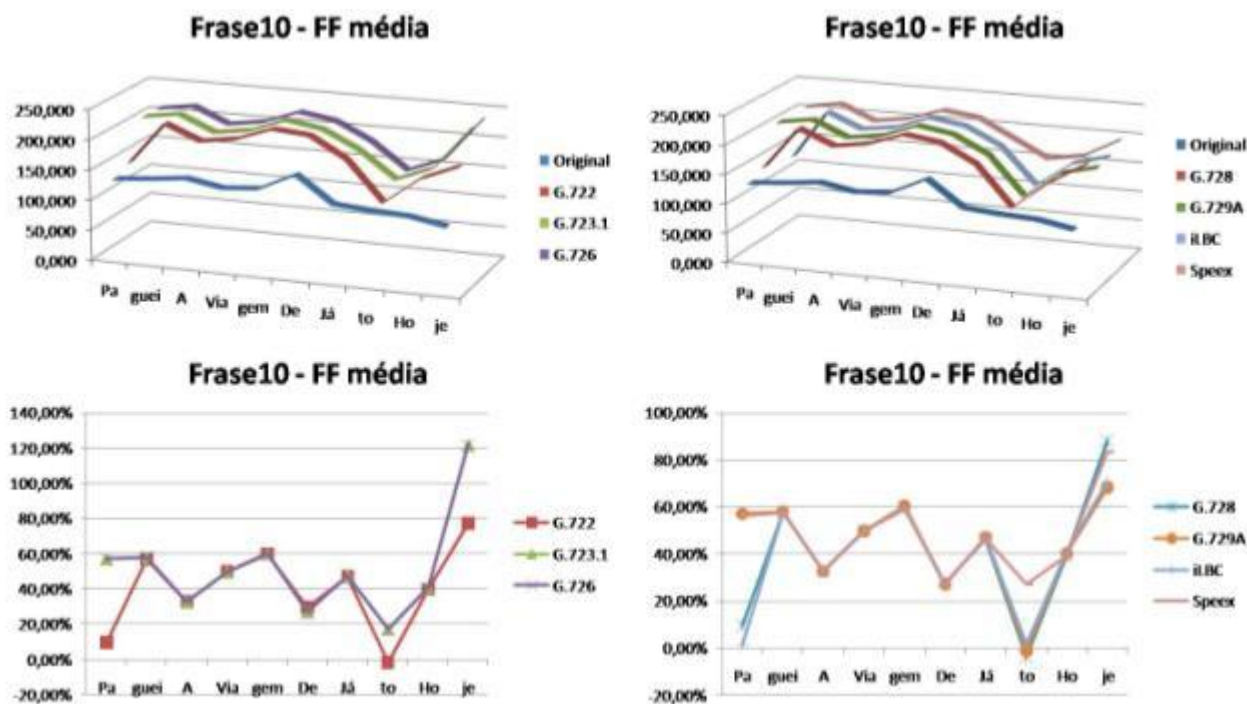


Figura 127 - Variações da frequência fundamental média da frase 10 do indivíduo M7-79 do grupo MG3.

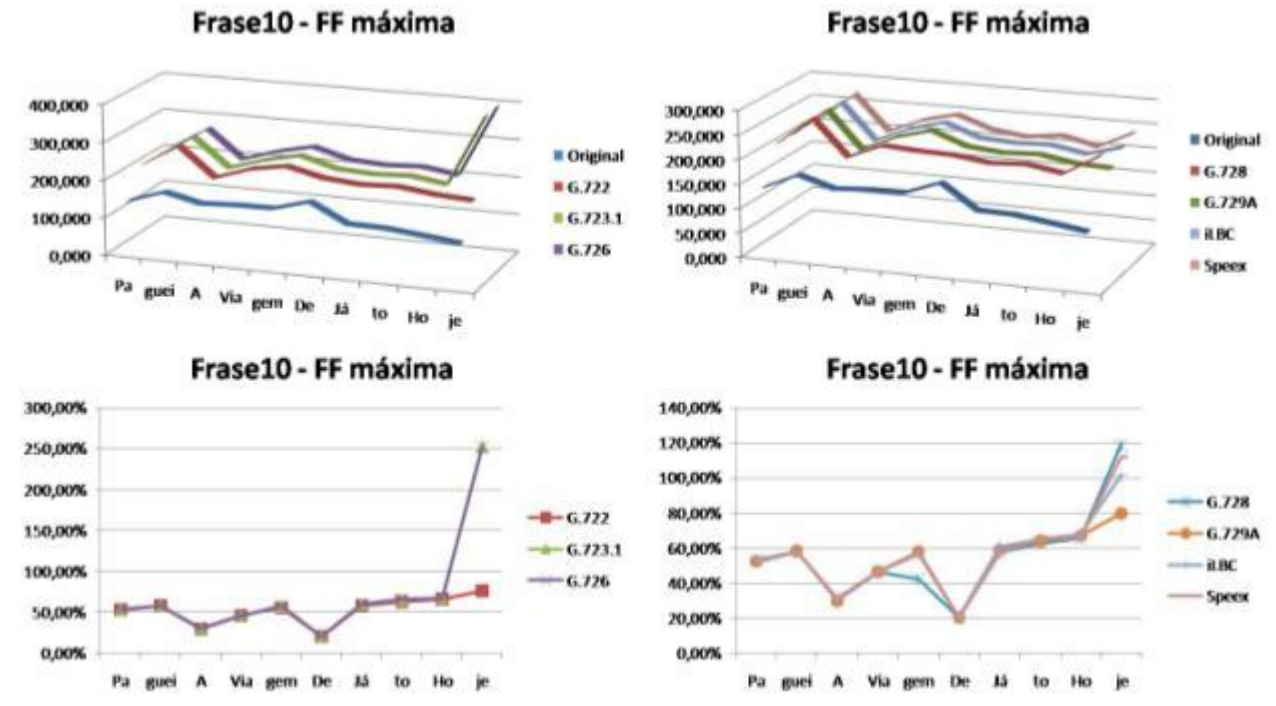


Figura 128 - Variações da frequência fundamental máxima da frase 10 do indivíduo M7-79 do grupo MG3.

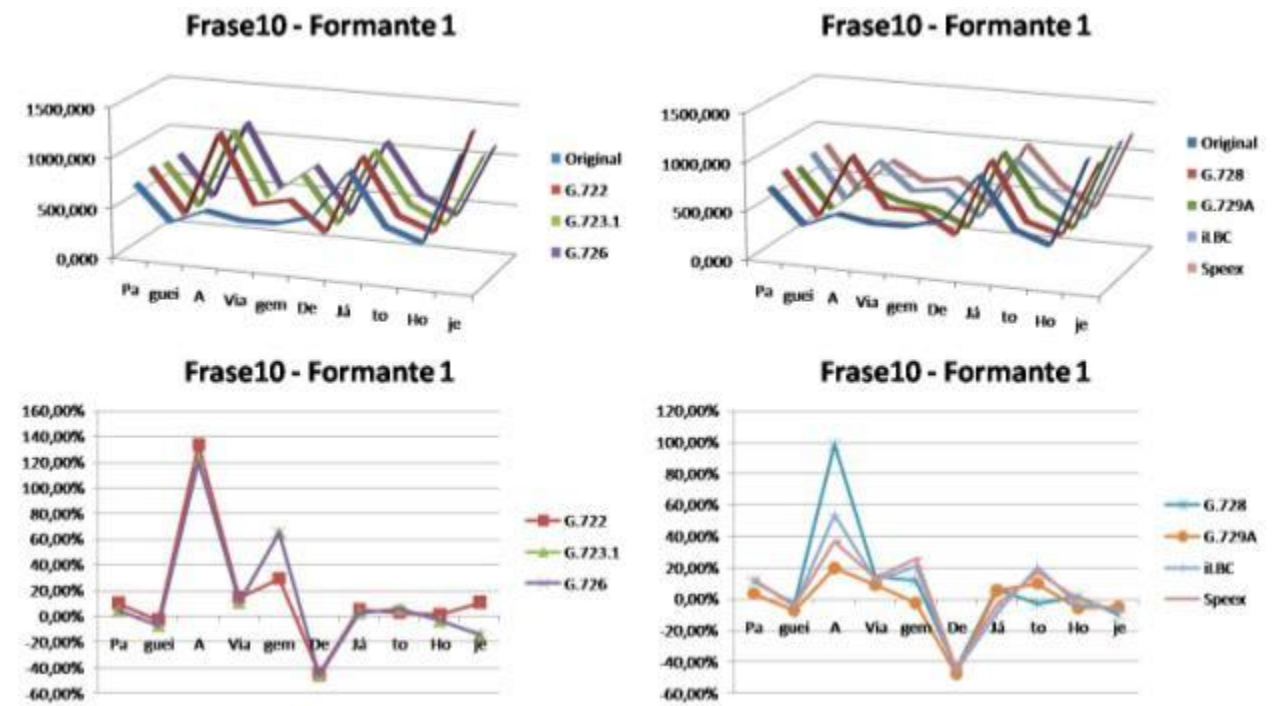


Figura 129 - Variações da frequência da primeira formante da frase 10 do indivíduo M7-79 do grupo MG3.

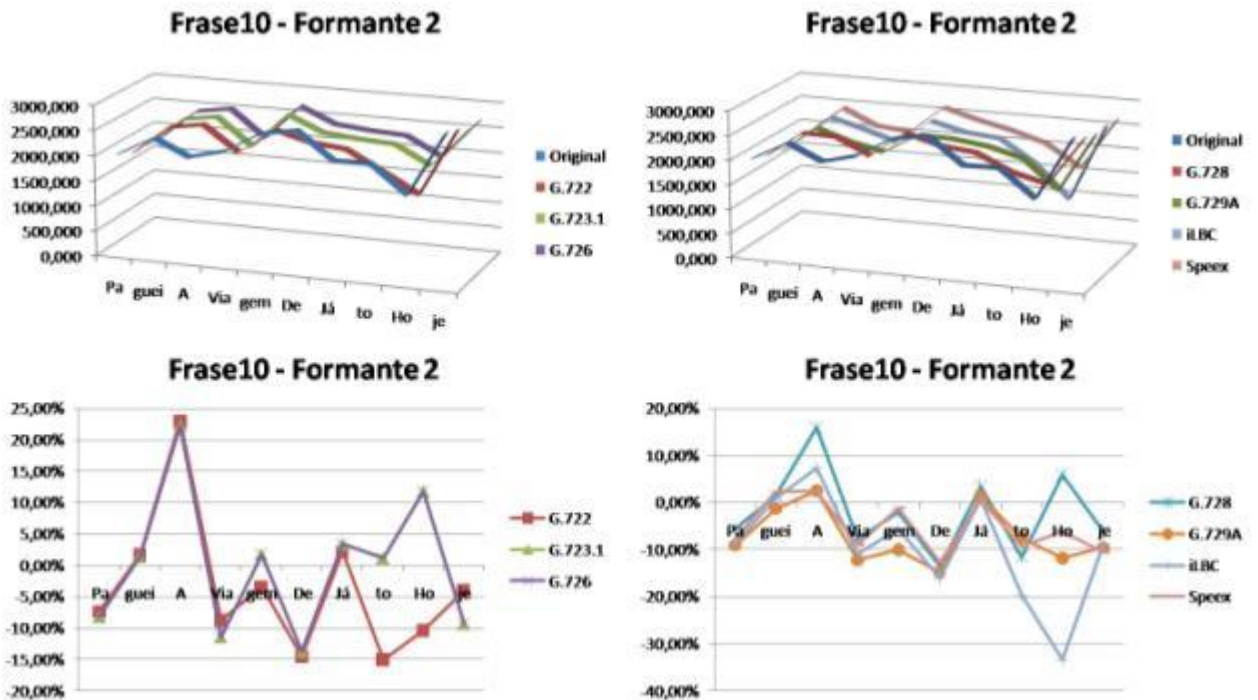


Figura 130 - Variações da frequência da segunda formante da frase 10 do indivíduo M7-79 do grupo MG3.

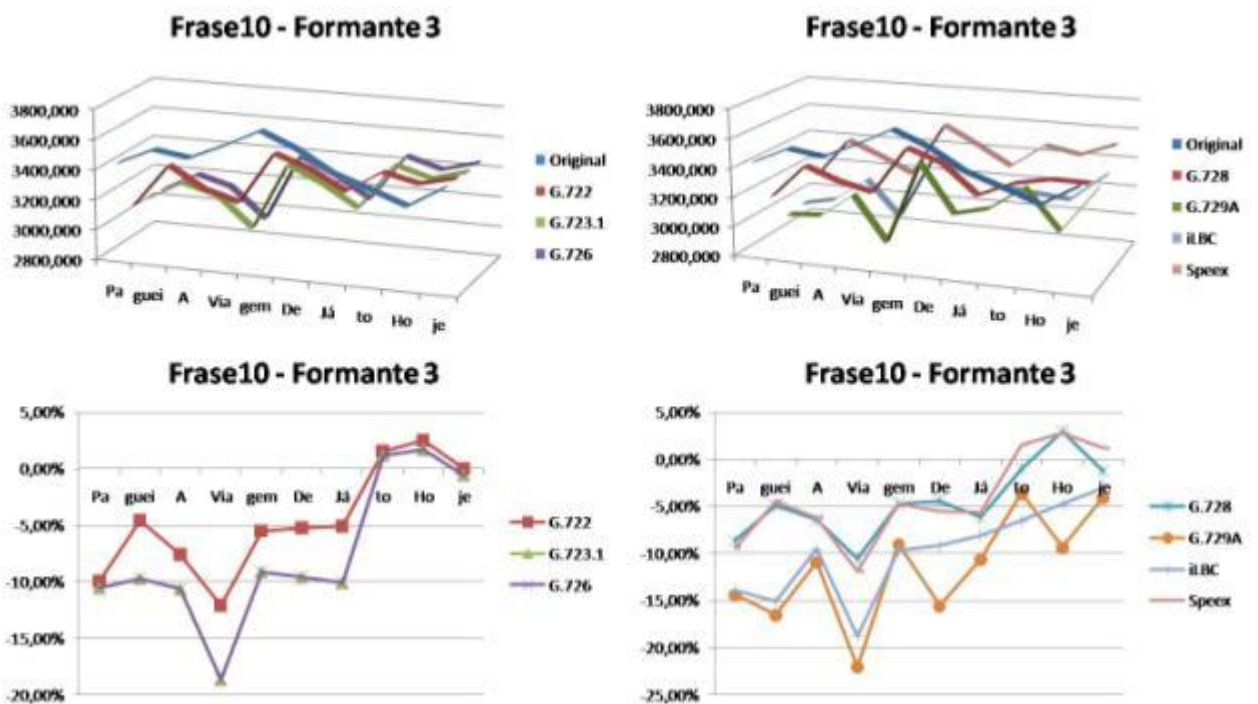


Figura 131 - Variações da frequência da terceira formante da frase 10 do indivíduo M7-79 do grupo MG3.

4.5 – Análise dos resultados

Os resultados médios de cada grupo são representativos da tendência do erro de processamento do sinal, seja através de um ganho nos valores das frequências ou de uma atenuação. Os resultados são analisados de acordo com três aspectos: taxa de tempo de processamento, erro de processamento das frequências fundamental e das formantes e inserção de ruídos.

4.5.1 – Taxa de tempo de processamento

Neste aspecto, o codificador G.723.1 apresentou as maiores taxas de tempo de processamento, geralmente acima de 20%. A segunda maior taxa de tempo de processamento pertence ao codificador G.728, com taxas médias de 10%. Na figura 132 temos um gráfico comparativo dos resultados médios obtidos para cada um dos grupos analisados.

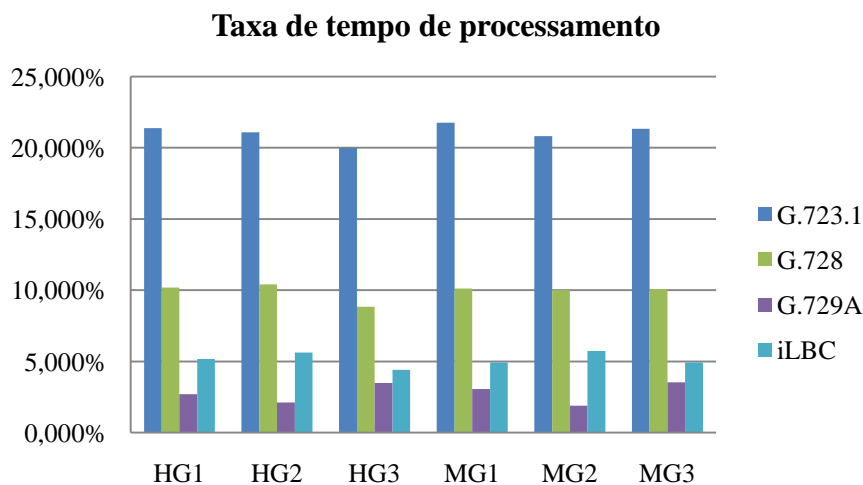


Figura 132 - Gráfico com os resultados médios da Taxa de Tempo de processamento de todos os grupos.

Entre os grupos dos homens, pode-se perceber que não há diferenças significativas entre os grupos HG1 e HG2. No entanto, pode-se perceber uma diminuição na taxa com relação ao grupo HG3 para os codificadores G.728, G.729A, iLBC e Speex, mas um aumento para o codificador G.726, único grupo a possuir dados deste codificador dentro da faixa de resolução do tempo de processamento utilizada.

Entre o grupo das mulheres, não se observa uma variação na taxa de processamento significativa com a variação da faixa etária.

Não foram observadas grandes variações na taxa de processamento nem com a variação da idade, nem com o sexo.

4.5.2 – Erro de processamento da frequência fundamental

A obtenção dos dados das frequências fundamentais mínima, média e máxima visou uma caracterização mais eficaz da vibração das cordas vocais, dado que nos intervalos existem transições entre sons vocálicos e sons consonantais. Nas figuras 133 a 135 temos os gráficos com os resultados médios do erro de processamento da frequência Fundamental de todos os codificadores analisados para cada um dos grupos.

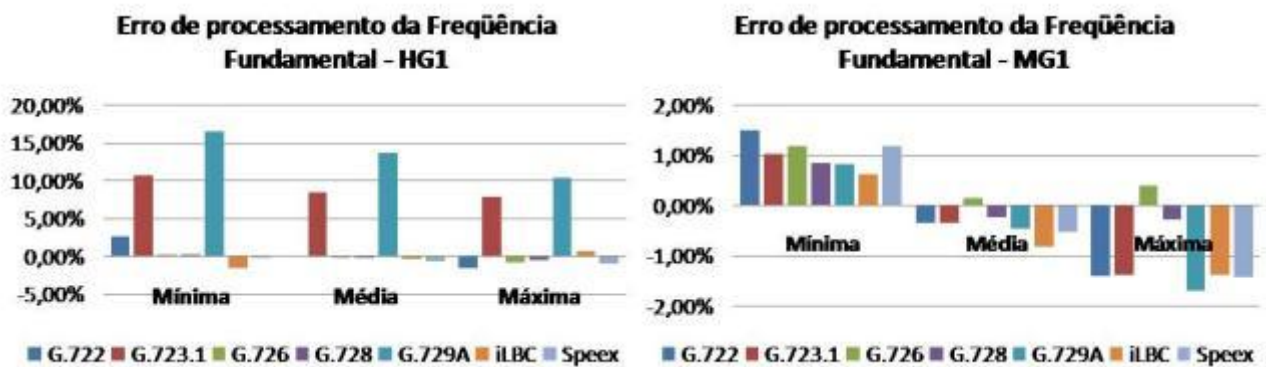


Figura 133 - Gráficos com o erro médio de processamento da frequência Fundamental dos codificadores analisados para os grupos HG1 e MG1.

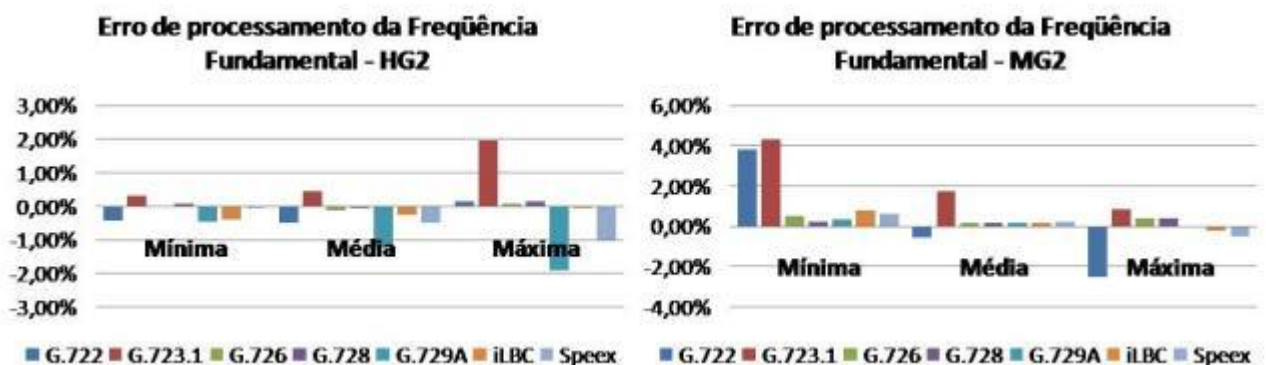


Figura 134 - Gráficos com o erro médio de processamento da frequência Fundamental dos codificadores analisados para os grupos HG2 e MG2.

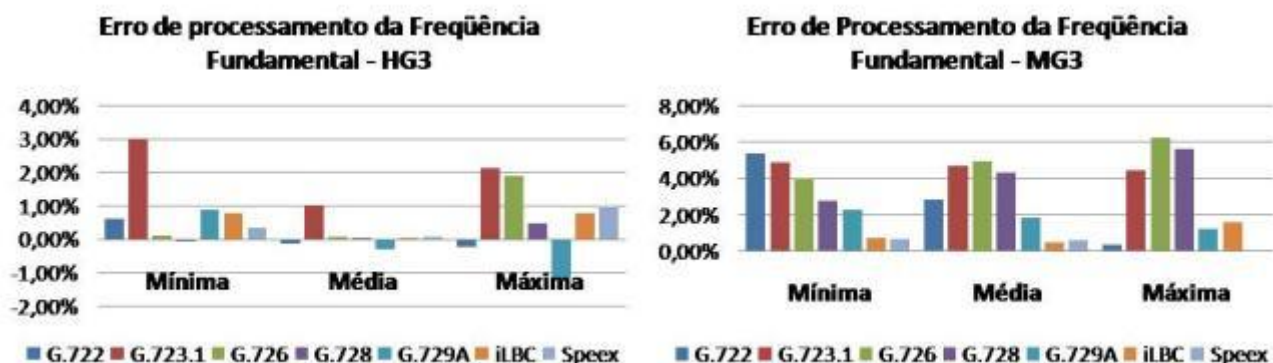


Figura 135 - Gráficos com o erro médio de processamento da frequência Fundamental dos codificadores analisados para os grupos HG3 e MG3.

Nas tabelas 42 a 47 temos o desvio padrão do erro médio de processamento para cada um dos gráficos. As células sombreadas representam variações no sinal do erro, ou seja, atenuações ou ganhos no processamento da frequência Fundamental.

Tabela 42 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo HG1.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	1,80%	17,36%	0,48%	0,44%	7,19%	1,94%	2,75%
Média	1,25%	14,00%	0,56%	0,74%	7,35%	0,98%	1,29%
Máxima	2,49%	11,59%	1,34%	1,27%	8,54%	2,75%	2,74%

Tabela 43 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo MG1.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	3,10%	0,67%	1,42%	3,11%	0,74%	1,63%	2,59%
Média	0,67%	1,42%	1,94%	0,74%	1,63%	2,59%	1,20%
Máxima	1,42%	1,94%	0,75%	1,63%	2,59%	1,20%	2,69%

Tabela 44 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo HG2.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	1,28%	0,92%	1,43%	0,48%	0,25%	0,42%	1,21%
Média	0,92%	1,43%	1,66%	0,25%	0,42%	1,21%	1,79%
Máxima	1,43%	1,66%	1,10%	0,42%	1,21%	1,79%	3,46%

Tabela 45 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo MG2.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	4,35%	1,35%	2,05%	1,07%	0,62%	1,70%	1,40%
Média	1,35%	2,05%	5,21%	0,62%	1,70%	1,40%	1,02%
Máxima	2,05%	5,21%	2,59%	1,70%	1,40%	1,02%	2,59%

Tabela 46 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo HG3.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	1,04%	0,49%	1,64%	1,71%	0,36%	1,93%	1,30%
Média	0,49%	1,64%	2,53%	0,36%	1,93%	1,30%	0,76%
Máxima	1,64%	2,53%	0,93%	1,93%	1,30%	0,76%	1,85%

Tabela 47 - Desvio padrão do erro médio de processamento da frequência Fundamental para o grupo MG3.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Mínima	5,92%	7,71%	9,59%	6,06%	7,61%	11,10%	7,34%
Média	7,71%	9,59%	7,61%	7,61%	11,10%	7,34%	6,81%
Máxima	9,59%	7,61%	8,60%	11,10%	7,34%	6,81%	8,47%

Observa-se a presença de erros de processamento da frequência fundamental acima de 10% em todos os grupos, seja como uma atenuação ou ganho. Entre o grupo dos homens, sob tal aspecto, o codificador G.723.1 aparece em todos os grupos. O codificador G.729A apresenta tais erros nos grupos HG1 e HG2. Os codificadores iLBC e Speex, por sua vez, nos grupos HG2 e HG3.

Entre o grupo das mulheres, o codificador G.722 apresenta erros desta ordem em todos os grupos. O codificador G.723.1, nos grupos MG2 e MG3. O codificador G.729A, apenas no grupo MG3, e os demais nos grupos MG1 e MG2.

Considerando-se todos os grupos, pode-se observar que os codificadores apresentam uma maior dificuldade no processamento da frequência Fundamental das mulheres, especialmente as do grupo MG3, que apresentam erros médios de processamento superiores a homens de mesma faixa etária. A única exceção é o grupo

HG1, onde os codificadores G.723.1 e G.729A apresentam erros médios superiores aos do grupo MG1.

O fenômeno de correção de *pitch* pode ser observado em todos os grupos. Os intervalos contendo os fonemas /s/, /p/, /t/ e /k/, , contidos, por exemplo, nas palavras “Políticos”, “Arte” e “Está”, concentraram as aparições de sons puramente consonantais, ou seja, sem a vibração de cordas vocais. Também foi observada a presença de vogais de baixa intensidade, ou seja, que não possuíram vibração detectável das cordas vocais. Os codificadores iLBC, G.723.1, G.729A e Speex apresentaram uma maior incidência desse fenômeno. Nos resultados individuais apresentados, o fenômeno pode ser observado nas figuras 65 a 67, 83 a 85, 89 a 91, 125 a 127 e 155 a 157. Perceptualmente, o efeito da correção de *pitch* é um aumento significativo na intensidade e clareza do som.

4.5.3 – Erro de processamento das freqüências das formantes

Devido a taxa de amostragem de 8 kHz, optou-se pela análise apenas das três primeiras formantes. Através do Praat foi possível coletar a freqüência média das formantes para cada intervalo. Nas figuras 136 a 138 temos os gráficos com o erro médio de processamento das formantes para cada um dos grupos.

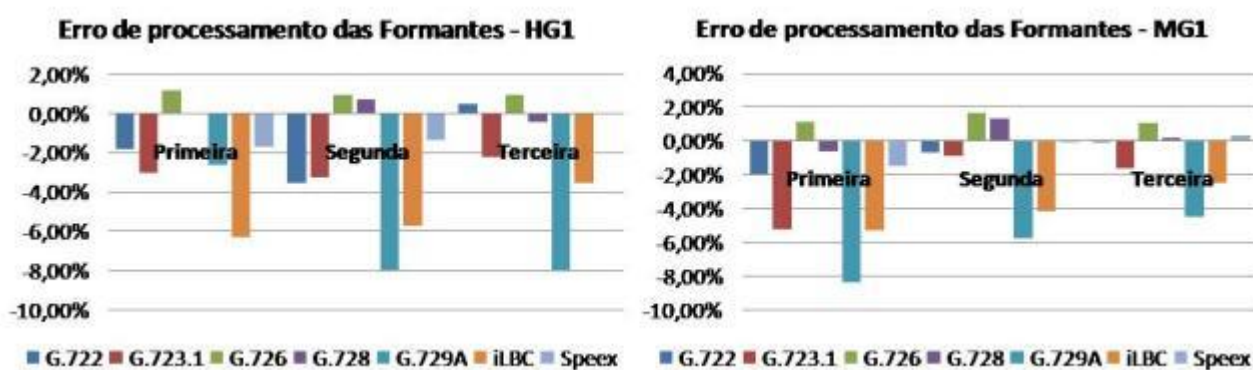


Figura 136 - Gráficos com o erro médio de processamento das formantes dos codificadores analisados para os grupos HG1 e MG1.

Nas tabelas de 48 a 53 temos o desvio padrão do erro médio de processamento das formantes para cada um dos grupos. As células sombreadas representam variações no sinal do erro, ou seja, atenuações ou ganhos no processamento das formantes.

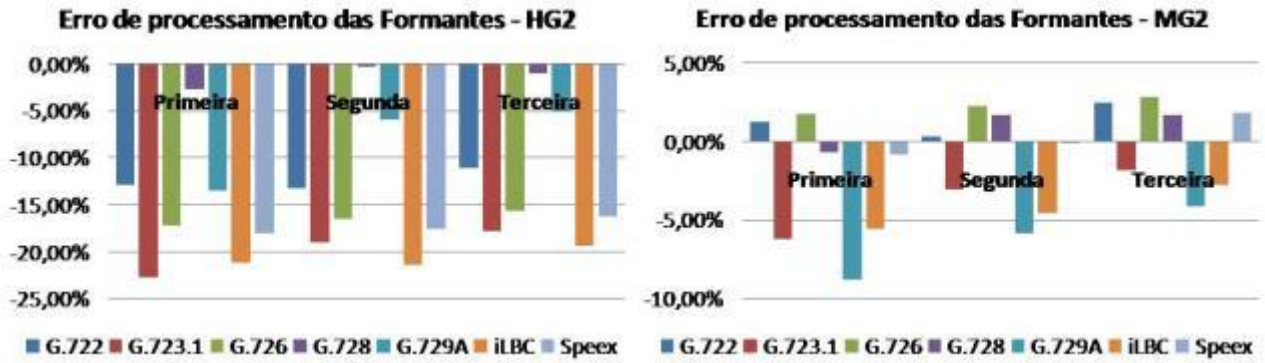


Figura 137 - Gráficos com o erro médio de processamento das formantes dos codificadores analisados para os grupos HG2 e MG2.

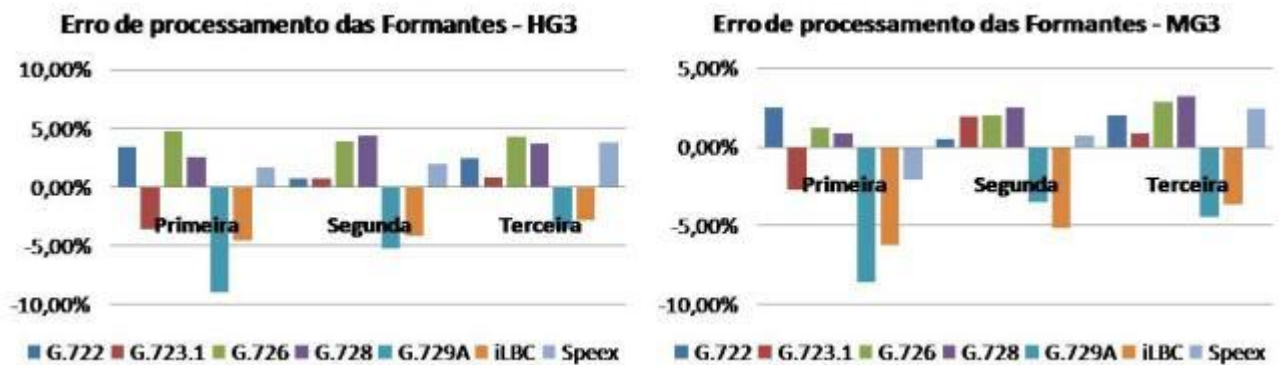


Figura 138 - Gráficos com o erro médio de processamento das formantes dos codificadores analisados para os grupos HG3 e MG3

Tabela 48 - Desvio padrão do erro médio de processamento das formantes para o grupo HG1.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	3,48%	6,91%	1,97%	2,08%	7,60%	3,75%	2,82%
Segunda	1,52%	2,77%	0,65%	1,37%	2,36%	1,26%	0,95%
Terceira	0,97%	1,21%	0,33%	0,52%	2,14%	0,37%	0,50%

Tabela 49 - Desvio padrão do erro médio de processamento das formantes para o grupo MG1.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	2,20%	2,43%	0,65%	1,24%	1,66%	0,42%	1,75%
Segunda	2,43%	0,65%	2,89%	1,66%	0,42%	1,75%	2,13%
Terceira	0,65%	2,89%	2,74%	0,88%	0,86%	0,80%	0,62%

Tabela 50 - Desvio padrão do erro médio de processamento das formantes para o grupo HG2.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	8,71%	8,24%	8,06%	5,57%	4,79%	4,85%	3,46%
Segunda	8,24%	8,06%	4,84%	4,79%	4,85%	3,46%	2,60%
Terceira	8,06%	4,84%	1,60%	4,85%	3,46%	2,60%	2,16%

Tabela 51 - Desvio padrão do erro médio de processamento das formantes para o grupo MG2.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	1,85%	0,70%	0,68%	1,33%	0,88%	0,61%	2,80%
Segunda	0,70%	0,68%	5,46%	0,88%	0,61%	2,80%	1,62%
Terceira	0,68%	5,46%	4,43%	0,61%	2,80%	1,62%	1,22%

Tabela 52 - Desvio padrão do erro médio de processamento das formantes para o grupo HG3.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	2,77%	0,74%	0,36%	2,89%	1,17%	1,18%	2,61%
Segunda	0,74%	0,36%	4,01%	1,17%	1,18%	2,61%	1,93%
Terceira	0,36%	4,01%	1,87%	1,18%	2,61%	1,93%	1,16%

Tabela 53 - Desvio padrão do erro médio de processamento das formantes para o grupo MG3.

	G.722	G.723.1	G.726	G.728	G.729A	iLBC	Speex
Primeira	4,34%	1,21%	1,43%	5,23%	1,52%	1,48%	5,07%
Segunda	1,21%	1,43%	4,26%	1,52%	1,48%	5,07%	2,94%
Terceira	1,43%	4,26%	2,21%	1,48%	5,07%	2,94%	2,01%

Observa-se a presença de erros de processamento das frequências das formantes acima de 10%, mantendo-se o mesmo foco utilizado para a frequência fundamental. Sob tal aspecto, o codificador G.723.1 apresenta tais erros na primeira formante de todos os grupos dos homens. O codificador G.729A, nos grupos HG1 e HG2, enquanto iLBC e Speex, nos grupos HG2 e HG3 e G.722 e G.723.1, nos grupos HG1 e HG3. Os codificadores G.728 e G.726 apresentaram tais erros apenas no grupo HG2. Na segunda e terceira formante, tais erros só se apresentam nos grupos HG1 e HG2, nos quais o codificador G.729A apresenta tais erros. Os codificadores iLBC e Speex apresentam tais erros apenas para a segunda formante no grupo HG2.

Pode-se observar que há uma dificuldade no processamento da primeira formante no grupo dos homens. Apenas o grupo HG2 apresentou erros iguais ou superiores a 10% na segunda e terceira formante. Não foi possível caracterizar um aumento do erro de processamento com o aumento da faixa etária para o grupo dos homens. No entanto, o grupo HG2 apresentou os piores resultados, revelando-se possivelmente uma dificuldade de processamento localizado nesta faixa etária, visto que todos os codificadores analisados apresentaram tais erros em todas as formantes.

Entre o grupo das mulheres, o codificador G.729A apresenta tais erros na primeira formante para todos os grupos. O codificador G.723.1 apresentou apenas nos grupos MG2 e MG3, enquanto os codificadores G.722, G.726, G.728, iLBC e Speex, apenas no grupo MG3. Quanto a segunda formante, apenas os codificadores G.723.1 e iLBC apresentaram tais erros nos grupos MG2 e MG3 respectivamente. Todos os codificadores analisados apresentaram tais erros na terceira formante no grupo MG3.

Considerando-se todos os grupos, é possível notar uma leve diferenciação no processamento do sinal por sexo, dado que os homens apresentaram uma maior incidência de erros com módulo maior ou igual a 10% para a segunda e terceira formante. É possível observar uma tendência de atenuação das formantes para os grupos HG1, MG1, HG2 e MG3, e de ganho para os grupos HG3 e MG3.

4.5.4 – Inserção de ruídos

Apenas os codificadores G.723.1 e G.722 apresentaram a inserção de ruídos nos arquivos processados.

O codificador G.723.1 apresentou também uma forte atenuação do arquivo processado. Um exemplo de efeito no espectrograma pode ser visualizado na figura 139, onde quase não é possível perceber as regiões de fala. Percebe-se apenas a vibração de cordas vocais e as formantes só podem ser vistas na área de ruído, selecionada em vermelho nessa figura.

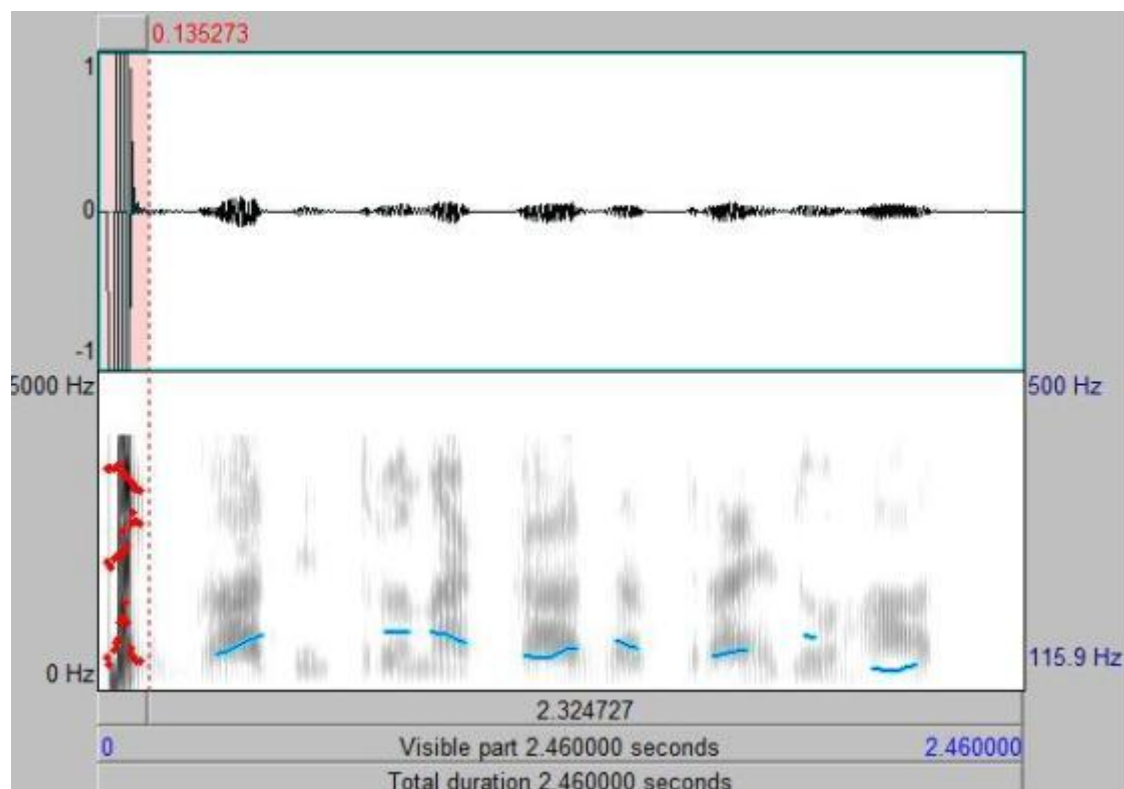


Figura 139 - Exemplo de inserção de ruído e atenuação gerada pelo codificador G.723.1.

O codificador G.722 inseriu ruído inicial, mas não houve degradação perceptual no restante do arquivo.

5. CONCLUSÕES E TRABALHOS FUTUROS

A obtenção de amostras com o controle dos fatores de degradação ambientais mostrou-se um desafio. O arranjo experimental utilizado, muito próximo do comumente utilizado em chamadas VoIP, mostrou ser muito sensível a ruídos de fundo, impossibilitando-se o processo de gravação de alguns indivíduos, mesmo com a diminuição da sensibilidade do microfone. A edição dos arquivos gravados para obtenção dos arquivos originais mostrou-se satisfatória com o uso do Praat. A utilização do espectrograma aliado à análise perceptual mostrou-se eficiente na separação dos intervalos fonéticos, exceto para intervalos contendo hiatos.

O processo de coleta de dados foi exaustivo, dado que cada indivíduo possui uma grade distinta para cada frase, associada à cadência ou ritmo de sua pronúncia. Devido à grande quantidade de informação a ser coletada, facultou-se a repetibilidade de gravação, o que permitiria uma análise mais detalhada, incluindo-se as variações fonéticas. No entanto, a análise efetuada possibilitou a detecção de variações no processamento de alguns grupos fonéticos levando-se em conta apenas aspectos de qualidade, fidelidade e tempo de processamento. Aspectos como taxa de transmissão e atraso não foram considerados.

Foi possível analisar a tendência dos codificadores em processar as variações do trato vocal durante a pronúncia das frases, representadas pelas frequências fundamental e das três primeiras formantes. Há uma dificuldade inerente a este tipo de análise: a necessidade de um grupo estatisticamente representativo de interlocutores para os fonemas da Língua Portuguesa. Entretanto, a utilização de frases, representando a concatenação de grupos fonéticos, mostrou-se próxima da situação de uma conversação em

tempo real, foco dos codificadores utilizados. Para uma melhor caracterização de tal situação, os arquivos originais foram processados uma única vez, visto que, em uma situação real, os codificadores possuem apenas uma chance de codificar corretamente uma amostra. Neste aspecto, o codificador G.723.1 apresentou os piores resultados com a inserção de ruído e forte atenuação de alguns arquivos processados. O codificador G.722 também inseriu ruído no início de todos os arquivos processados, mas não houve atenuação similar à observada no G.723.1.

Com a utilização de todos os fonemas básicos da Língua Portuguesa, é possível notar a tendência dos codificadores analisados em estimar as variações do trato vocal do interlocutor durante as transições dos grupos fonéticos analisados. Com a separação de grupos por sexo e faixas etárias distintas, é possível ter uma estimativa representativa dos erros de predição e considerar a utilização de um ou outro codificador sob tais aspectos. Esse tipo de análise para a Língua Portuguesa, levando-se em conta apenas o processamento do sinal, é inédita na literatura.

Foi possível observar variações no erro de processamento por sexo apenas para as formantes, tipicamente a segunda e terceira formantes, onde os homens apresentaram uma maior incidência de erros de processamento com módulos superiores a 10%. Apenas as mulheres com maior faixa etária apresentaram erros com tal magnitude na terceira formante. Quanto à variação da faixa etária, observou-se uma tendência de aumento no erro apenas no grupo das mulheres. Foi possível, também, observar uma dificuldade de processamento concentrada nos grupos HG2 e MG3. No entanto, apenas um aumento no número de indivíduos analisados tornaria possível confirmar tais fatos.

É possível verificar que os codificadores G.726 e G.728 apresentaram os melhores resultados gerais para todas as frequências em todos os grupos. Os codificadores G.723.1 e G.729A, por sua vez, apresentaram uma maior incidência de erros de processamento para todas as frequências analisadas acima do patamar de comparação escolhido. O codificador G.722, apesar de não possuir etapas de extração de informações sobre o *pitch*, apresentou bons resultados quanto à frequência fundamental, principalmente para os homens.

É possível perceber nos resultados individuais um aumento na dificuldade de predição com o aumento da ordem das frequências. Provavelmente

associada à ordem dos filtros de análise LPC utilizados nos preditores. Logo, um aumento na ordem dos filtros de análise tenderia a diminuir esse efeito.

O fenômeno de correção do *pitch* foi observado tanto em homens quanto em mulheres. Do ponto de vista da fidelidade ao som original, este fenômeno representa um erro de codificação. Ao mesmo tempo, representa uma estimativa exata do tipo de som emitido, não representando, então, um erro de processamento. Tal variação, considerando-se a frase como um todo, não é perceptível, dado que o cérebro pode completar o intervalo de forma a torná-lo compreensível. Nos intervalos de ocorrência do fenômeno, perceptualmente observa-se apenas um aumento na sua intensidade.

Outro ponto analisado foi a taxa de tempo de processamento, apesar de não ser o foco desse trabalho. A taxa de tempo de processamento indica uma relação entre o tempo necessário para codificar e decodificar o sinal de áudio e a duração deste sinal. Levando-se em conta o comprimento em segundos dos quadros de amostras, é possível ter uma noção do tempo real necessário para processar tal quadro. Por exemplo, um codificador que possua uma taxa de tempo de processamento igual a 20% e trabalhe com quadros de 10 ms, consome 2 ms para codificar e decodificar um quadro. A taxa de processamento, portanto, dá uma indicação se a resposta em tempo real, considerando-se fatores de degradação de rede como atrasos e latência, de tal codificador pode manter um nível satisfatório de qualidade de chamada. Neste aspecto, não foi possível detectar uma variação com o sexo e faixa etária. Os codificadores G.723.1 e G.728 apresentaram as piores taxas de tempo de processamento.

Este trabalho atingiu seu objetivo mostrando as variações na capacidade de processamento do sinal pelos codificadores utilizados em comunicações digitais, especificamente em chamadas VoIP.

5.1 – Trabalhos Futuros

A coleta de amostras de voz de outros indivíduos de várias regiões do Brasil, juntamente com o tipo de análise desenvolvida neste trabalho, seria de vital importância para a criação de um mapa do processamento de sinal para todos os fonemas da Língua Portuguesa falada no Brasil. Com uma amostra de indivíduos e grupos fonéticos

estatisticamente representativos, seria possível analisar as variações nas características do trato vocal por fonema indicando possíveis falhas de processamento por fonema e tipo de voz. Caso seja detectada uma variação estatisticamente comprovada por idade e sexo, poderia ser proposta uma ferramenta de análise capaz de identificar o codificador mais adequado para o tipo de usuário. Principalmente com a adição de outros codificadores à lista de análise, como os utilizados na telefonia celular.

A ferramenta utilizada possibilita a extração da banda passante dos filtros representativos das formantes, detalhando-se ainda mais as características da variação do trato vocal para essas frequências. Assim, torna-se possível uma melhor análise do modelo fonte-filtro e o impacto do processamento de sinal nestes parâmetros.

Seria possível analisar o impacto em sistemas de reconhecimento de voz do processamento de sinais e do fenômeno de correção de *pitch*. Como também ter uma melhor estimativa do patamar de erro de processamento tolerável, visto que nesse trabalho foi utilizado um patamar baseado apenas na observação dos erros médios por grupo.

REFERÊNCIAS

- [1] *Internet*, disponível em <http://en.wikipedia.org/wiki/Internet>. Acesso em: maio 2008.
- [2] *Digital Subscriber Line*, disponível em [http://en.wikipedia.org/wiki/Digital Subscriber Line](http://en.wikipedia.org/wiki/Digital_Subscriber_Line). Acesso em: maio 2008.
- [3] *Voice over Internet Protocol*, disponível em <http://en.wikipedia.org/wiki/VoIP>. Acesso em: maio 2008
- [4] *Network Voice Protocol*, disponível em http://en.wikipedia.org/wiki/Network_Voice_Protocol. Acesso em: maio 2008.
- [5] *H.323*, disponível em <http://en.wikipedia.org/wiki/H.323>. Acesso em: maio 2008.
- [6] CAVALCANTE, F. Estudo completo sobre VoIP no Brasil. **iMASTERS**, 2005. Disponível em http://imasters.uol.com.br/artigo/3520/tecnologia/estudo_completo_sobre_voip_no_brasil. Acesso em maio 2008.
- [7] GOODE, B. *Voice over Internet Protocol (VoIP)*. *Proceedings of the IEEE*, v.90, n.9, p. 1495-1517, Sept. 2002.
- [8] CRUZ, A. G. da. Voz sobre IP. Redes de Computadores I, UFRJ. Disponível em: http://www.gta.ufrj.br/grad/00_2/alexandre/VoIP.html#protocolos. Acesso em: maio 2008.
- [9] SIP, disponível em <http://pt.wikipedia.org/wiki/SIP>. Acesso em: junho 2008.
- [10] SIP Server Technical Overview, RADVISION, April 2004. Disponível em: http://www.sipforum.org/component/option,com_docman/task,cat_view/gid,13/Itemid,75/. Acesso em junho 2008.
- [11] PERJONS, M. *Measuring Voice Quality*, Global IP Sound, 2006. Disponível em: <http://developer.gipscorp.com>. Acesso em: maio 2008.

-
- [12] *Fonema*, disponível em <http://pt.wikipedia.org/wiki/Fonema>. Acesso em junho 2008.
- [13] HASEGAWA-JOHNSON, M. Lectures Notes in Speech Production, Speech Coding and Speech Recognition. University of Illinois, Feb. 17, 2000. Disponível em: <http://www.ifp.uiuc.edu/~hasegawa/notes/index.html>. Acesso em: maio 2008.
- [14] MANOSSO, R. Aspectos de Gramática. Disponível em: <http://www.radames.manosso.nom.br/gramatica/index.htm>. Acesso em: Junho 2008.
- [15] SCHÜTZ, R. Diferenças de pronúncia entre Inglês e Português. Disponível em <http://www.sk.com.br/sk-pron.html>. Acesso em: Junho 2008.
- [16] CHBANE, Dimas Trevizan. *Desenvolvimento de Sistema para Conversão de Textos em Fonemas no Idioma Português*. São Paulo, 1994. Dissertação (Mestrado em Engenharia Elétrica), Escola Politécnica, Universidade de São Paulo.
- [17] FLANAGAN, J. L. *Speech Analysis Synthesis and Perception*. 2 ed. New Jersey, Springer-Verlag, 1972.
- [18] VoIP – *Voice over IP*. INFREDE, 2006. Disponível em: <http://www.inforede.net/voip.htm>. Acesso em: julho, 2008.
- [19] DE ASSIS, T. J. M. Fundamentos do teorema de Nyquist e sua aplicação na transmissão digital. Disponível em: http://br.geocities.com/tvdigitalbr/tom_artigos/nyquist/nyquist.html. Acesso em: julho 2008.
- [20] Sampling Theorem and Nyquist Rate, eFunda 2008. Disponível em: http://www.efunda.com/designstandards/sensors/methods/dsp_nyquist.cfm. Acesso em: julho 2008.
- [21] *Linear Predictive Coding*, disponível em: http://en.wikipedia.org/wiki/Linear_predictive_coding. Acesso: julho 2008.
- [22] *Pulse Code Modulation (PCM) of Voice Frequencies*, ITU-T Recommendation G.711, 1993.
- [23] *G.711*, disponível em: <http://en.wikipedia.org/wiki/G.711>. Acesso em: agosto 2008.

-
- [24] *Vocabulary of Digital Transmission and Multiplexing, and Pulse Code Modulation (PCM) Terms*, ITU-T Recommendation G.701, 1993.
- [25] *7kHz Audio-coding within 64kbps*, ITU-T Recommendation G.722, 1988.
- [26] *Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*, ITU-T Recommendation G.723.1, 2006.
- [27] *ARMA (AutoRegressive Moving Average) Spectrum*, disponível em: [http://www.clecom.co.uk/science/autosignal/help/ARMA_\(AutoRegressive_Moving.htm\)](http://www.clecom.co.uk/science/autosignal/help/ARMA_(AutoRegressive_Moving.htm). Acesso em: outubro2008.
- [28] *40, 32, 24, 16kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)*, ITU-T Recommendation G.726, 1990.
- [29] RABINER, L., JUANG, B. H., *Fundamentals of Speech Recognition*, Ed. Prentice Hall, 1993.
- [30] *Window function*, disponível em: http://en.wikipedia.org/wiki/Window_function#Rectangular_window. Acesso: outubro 2008.
- [31] *5-, 4-, 3- and 2-bits sample embedded Adaptive Differential Pulse Modulation (ADPCM)*, ITU-T Recommendation G.727, 1990.
- [32] *Voice Packetization – Packetized Voice Protocols*, ITU-T Recommendation G.764, 1990.
- [33] *Coding of Speech at 16kbps using Low-Delay Code Excited Linear Prediction*, ITU-T Recommendation G.728, 1992.
- [34] *Coding of Speech at 8kbps using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)*, ITU-T Recommendation G.729, 2007.
- [35] *Finite Impulse Response*, disponível em: http://en.wikipedia.org/wiki/Finite_impulse_response. Acesso: dezembro 2008.
- [36] *Internet Low Bit rate Codec (iLBC)*, RFC3951, The Internet Society, 2004.

-
- [37] *internet Low Bit rate Codec*, disponível em: <http://en.wikipedia.org/wiki/ILBC>. Acesso: fevereiro 2009.
- [38] VALIN, J., *Speex Codec Manual, Version 1.2 Beta2*, 2007.
- [39] *Voice Quality (VQ) in converging Telephony and Internet Protocols (IP) Networks*, disponível em: http://www.iec.org/online/tutorials/voice_qual/index.asp. Acesso: março 2009.
- [40] MPLS, disponível em: <http://pt.wikipedia.org/wiki/MPLS>. Acesso: março, 2009.
- [41] NÓBREGA, O., KELNER, J, SADOK, D., MESQUITA, T. M. S., *Um algoritmo adaptativo de transmissão para serviços de voz sobre redes IP*, 19º SBRC 2001.
- [42] *One-way transmission time*, Recommendation ITU-T G.114, 2003.
- [43] *Communications Dictionary*, disponível em: <http://mcms009a.erlm.siemens.de/data-lexicon/lexicon/index/indxg.htm>. Acesso: março 2009.
- [44] *Mean Opinion Score*, disponível em: http://en.wikipedia.org/wiki/Mean_Opinion_Score. Acesso março 2009.
- [45] *Codecs*, disponível em: <http://www.en.voipforo.com/codec/codecs.php>. Acesso: março 2009.
- [46] *Speech Codecs and Associated PSQM Values*, disponível em: http://www.vocal.com/psqm_data.html. Acesso março 2009.
- [47] *Methods for subjective determination of transmission quality*, Recommendation ITU-T P.800, 1996.
- [48] *Assessing VoIP Call Quality Using the E-model*, disponível em: http://www.ixiacom.com/library/white_papers/display?skey=voip_quality. Acesso: março 2009.
- [49] TELES , V. C., ROSINHA, A. C. U., *Análise Acústica dos Formantes e das Medidas de Perturbação do Sinal Sonoro em Mulheres Sem Queixas Vocais, Não Fumantes e*

Não Etilista, Revista @rquivos Internacionais de Otorrinolaringologia, vol. 12, n.4, p. 523-530, 2008.

- [50] *Praat: doing Phonetics by computer*, disponível em: <http://www.fon.hum.uva.nl/praat/>. Acesso: julho 2008.
- [51] SoX – Sound Exchange, disponível em: <http://sox.sourceforge.net/>. Acesso: dezembro 2008.
- [52] Imtelephone, disponível em: <http://www.imtelephone.com/>. Acesso: janeiro 2009.
- [53] VoiceAge, disponível em: <http://www.voiceage.com/>. Acesso: janeiro 2009.
- [54] iLBCfreeware, disponível em: <http://www.ilbcfreeware.org/>. Acesso: janeiro 2009.
- [55] Speex, disponível em: <http://speex.org/>. Acesso: janeiro 2009.
- [56] *Dicionário Michaelis*, disponível em: <http://michaelis.uol.com.br/>.
- [57] Cambridge Dictionaries Online, disponível em: <http://dictionary.cambridge.org/>.
- [58] OPPENHEIM, A. V., WILLSKY, A.S, NAWAB, S.H, *Signals and Systems*, Ed. Prentice Hall, Second Edition, 1997.
- [59] *International Phonetic Alphabet*, disponível em: <http://en.wikipedia.org/wiki/IPA>. Acesso: junho 2009.
- [60] OPPENHEIN, A. V., *Discrete-Time Signal Processing*, Ed. Prentice Hall, Second Edition, 1998.