

# Codificador G729a orientado à avaliação da qualidade perceptual do sinal de voz

Flávio Luis Wisnevski e Rubem Dutra Ribeiro Fagundes

**Resumo** — Esta pesquisa tem como objetivo principal desenvolver um modelo de codificador de voz que avalia a qualidade perceptual. Estamos utilizando um codificador paramétrico definido pela recomendação ITU-T G729a, tecnicamente conhecido como CS-ACELP (Conjugate Structure Algebraic Code Excited Linear Prediction). Para este cenário foi realizado o controle de avaliação de qualidade do sinal de fala, com o algoritmo PESQ (Perceptual Evaluation of Speech Quality).

**Keywords**— Voice Codec, G729, PESQ

**Abstract** — This research has as main objective to develop the voice encoder model that evaluates the perceptual quality. We are using a parametric encoder defined by Recommendation ITU-T, G729a, technically known as Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP). For this scenario was performed to control quality assessment of the speech signal, the algorithm PESQ (Perceptual Evaluation of Speech Quality).

**Key words** — Voice Encoder, G729, PESQ

## I. OBJETIVO

O procedimento experimental deste trabalho consiste em avaliar as características do sinal de voz amostrado no bloco do codificador, através da implementação do algoritmo do PESQ (Perceptual Evaluation of Speech Quality) no código fonte do codificador G729a [1]. Isso possibilita a análise perceptual do sinal independente das variações do meio de transmissão entre os blocos *encoder* e *decoder*.

Implantar um modelo de treinamento para os padrões do algoritmo PESQ sobre o bloco do codificador, fazendo com que exista coerência no valor dos resultados do MOS (Mean Opinion Score)[3]. Usaremos como base as especificações da recomendação P.862 do ITU-T [4], a qual descreve o tamanho da amostra em relação ao tempo de envio dos pacotes.

Propor alterações na estrutura do bloco *encoder* do codificador CS-ACELP para obter uma melhoria do sinal codificado analisado pelo algoritmo PESQ.

Avaliar a codificação do sinal de voz e a sua qualidade perceptual baseado nos parâmetros restritos ao codificador, melhorando a qualidade perceptual nos frames codificados.

## II. FUNDAMENTAÇÃO TEÓRICA

### A. Codec G729a (CS-ACELP)

O modelo de codificação é baseado no algoritmo de codificação CS-ACELP, *Conjugate Structure Algebraic Code-Excited Linear Prediction*. Efetua a predição linear por excitação com código algébrico. Nesse intervalo de tempo, o sinal de voz (quase estacionário) é analisado e parametrizado, comparado através de uma análise por síntese em busca da melhor excitação do sinal que corresponda ao sinal alvo. Cada

frame de 10ms que é analisado, gera 80 bits por segundo, em que são extraídos os parâmetros por frame: coeficientes de predição linear do filtro, índices do *codebook* fixo, índices do *codebook* adaptativo e ganho. Esses parâmetros codificados e transmitidos são distribuídos de acordo com a tabela I, a qual explica a alocação de bit dos 8kb/s do algoritmo CS-ACELP em um frame em 10ms:

Parameter	Codeword	Sub frame 1	Sub frame 2	Total per frame
Line spectrum pairs	L0, L1, L2, L3			18
Adaptive-codebook delay	P1, P2	8	5	13
Pitch-delay parity	P0	1		1
Fixed-codebook index	C1, C2	13	13	23
Fixed-codebook sign	S1, S2	4	4	8
Codebook gains (stage 1)	GA1, GA2	3	3	6
Codebook gains (stage 2)	GB1, GB2	4	4	8
<b>Total</b>				<b>80</b>

TABELA I. PADRÃO DE FRAMES ENVIADOS PELO CODEC G729A

O Codec CS-ACELP, segundo sua descrição na recomendação ITU G729a [1], consiste em um codificador e um decodificador. Para uma amostra constituída de 80 bits (frame de 10ms), as principais etapas são executadas conforme os estágios de codificação do CS-ACELP:

1) O sinal de entrada passa por um filtro passa-alta para eliminação de ruído, é pré-processado e esse sinal é utilizado para as análises subseqüentes da codificação.

2) A análise de Predição Linear [5] é realizada uma vez a cada frame de 10ms calculando-se os coeficientes de filtro de Predição Linear. Esses coeficientes são convertidos para *Line Spectrum Pairs* (LSP) [6] e sofrem quantização vetorial.

3) O sinal de excitação é escolhido por meio de uma análise por síntese, procedimento de busca em que o erro entre o sinal de voz original e o reconstruído é minimizado de acordo com uma medida de distorção perceptualmente ponderada. Isso é feito através da filtragem do sinal de erro com um filtro de ponderação perceptual, cujos coeficientes são derivados da Predição Linear não quantizada do filtro.

4) Os parâmetros de excitação (parâmetros dos dicionários fixos e adaptativo) são determinados por um subframe de 5ms (40 bits) cada um. Os coeficientes quantizados e não quantizados do filtro de Predição Linear são usados no segundo subframe.

5) Um valor de *pitch* é estimado com base no sinal resultante da avaliação do filtro de ponderação, em malha aberta.

6) Os estados iniciais do filtro são atualizados pela filtragem do erro residual. Calcula-se o circuito em malha fechada para estimar o ganho e o valor do dicionário adaptativo.

7) Com os valores do dicionário adaptativo e com as duas estimativas de *pitch*, busca-se no dicionário fixo um valor para encontrar a solução ótima.

8) Os ganhos dos dicionários (fixo e adaptativo) são quantizados, o filtro é atualizado e o sinal de excitação é determinado.

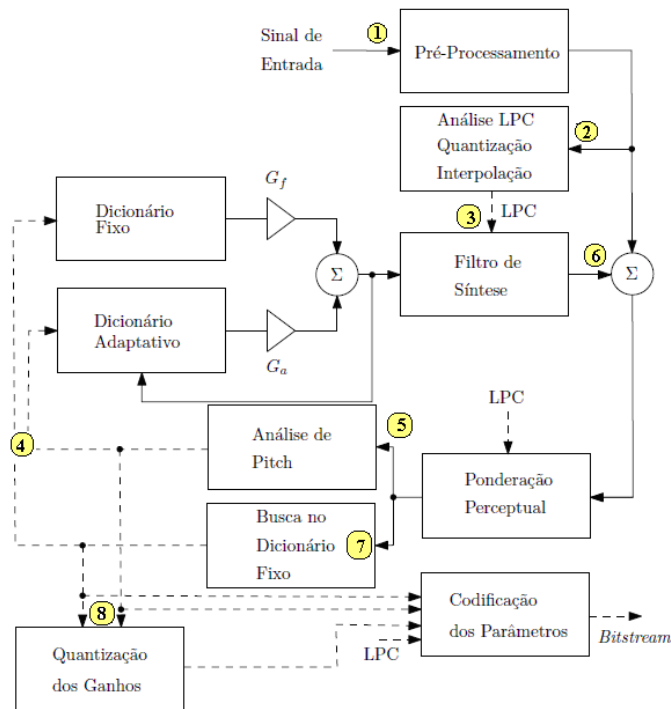


Figura 1. Diagrama de blocos do encoder CS-ACELP (G729)

Na decodificação do sinal os parâmetros do fluxo de bits recebidos são extraídos conforme descrição dos estágios de decodificação do CS-ACELP:

- 1) Os índices são decodificados para obter os parâmetros do codificador correspondente a um quadro de fala de 10 ms.
- 2) Esses parâmetros são os coeficientes LSP e duas frações de *delay* do *pitch*, dois vetores do *codebook* fixo e os dois conjuntos de ganhos referente ao *codebook* adaptativo e o *codebook* fixo.
- 3) Efetua-se a síntese de curto prazo: Os coeficientes LSP são interpolados e convertidos em coeficientes do filtro de predição linear para cada sub frame.
- 4) Efetua-se a síntese de longo prazo: É avaliado um subframe de 5ms, em que são executadas as seguintes etapas:
  - a) A excitação é calculada adicionando os vetores do codebook fixo e do codebook adaptativo, dimensionado pelos respectivos ganhos.
  - b) O sinal de fala é reconstruído através da filtragem da excitação, realizada pelo filtro de síntese de predição linear.
  - 5) A reconstrução do sinal de fala ocorre após o estágio de pós filtragem, que inclui um filtro adaptativo baseado nos filtros de síntese a curto e longo prazo, seguido por um filtro passa-alta que completa a operação de filtragem e dimensionamento da operação.

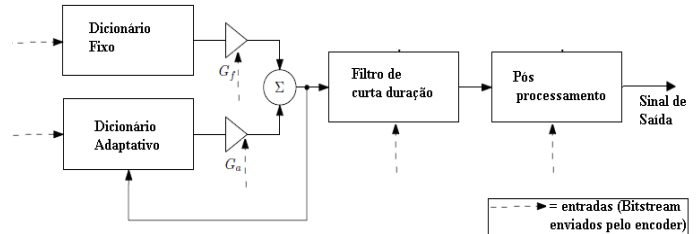


Figura 2. Diagrama de blocos do decoder CS-ACELP (G729)

### B. MOS (Mean Opinion Score)

É um método de avaliação perceptual do sinal, fornece uma medida numérica da qualidade da voz humana do circuito. Este processo faz uso de testes subjetivos através de opiniões que são matematicamente computados e que pela sua média são obtidos indicadores de qualidade do desempenho do sistema. Para determinar o MOS[3], um número de ouvintes percebe a qualidade da voz através da leitura em voz alta de sentenças de oradores masculinos e femininos ao longo da avaliação. O ouvinte fornece as notas de cada sentença com a seguinte definição: (1) ruim; (2) pobre; (3) razoável; (4) bom; (5) excelente. O MOS é a média aritmética de todas as notas individuais e pode variar de 1 (pior) a 5 (melhor).

### C. PESQ (Perceptual Evaluation of Speech Quality)

Esse algoritmo, PESQ, é um padrão do ITU-T para medida de qualidade descrito na recomendação P.862 [4], foi desenvolvido especificamente para ser utilizado em redes TCP/IP utilizando VoIP e em telefonia convencional, já que o algoritmo da recomendação P.861 o PSQM(Perceptual Speech Quality Measure) não era eficiente para tratar dos problemas específicos destas redes. O PESQ apresenta a medida da qualidade com resposta relacionada a medida do MOS.

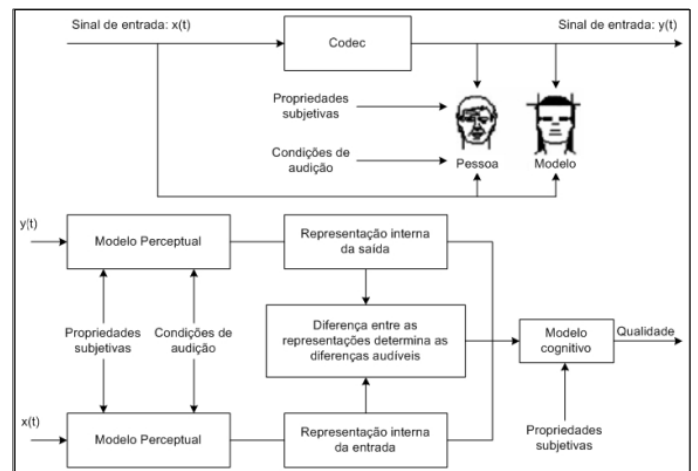


Figura 3. Modelo perceptual do PESQ

O algoritmo PESQ é uma ferramenta de avaliação que faz uma compensação do ganho nos sinais original e degradado, apresentando o mesmo nível de potência. Essa calibração é executada tanto no domínio do tempo quanto na frequência. Realiza-se após uma filtragem nos sinais, em que o sinal a ser analisado (enviado sobre a rede) permanece com mesma característica do sinal de origem (fonte de transmissão). Depois disso, os sinais são alinhados no tempo em que é definido o intervalo de análise. A conversão para o domínio da frequência é realizada utilizando uma janela *Hanning* sobre quadros de 32ms e 50% de sobreposição. A análise é feita sobre as

diferenças no domínio da frequência na escala Bark [7], após serem realizadas as compensações de percepção psicológica do som correlacionado com a sua intensidade em termos de amplitude e de variações de ganho (*loudness*).

### III. PROPOSTA DE TRABALHO

Para verificação do estudo, foi implementado a recomendação P.862 (PESQ) do ITU-T no *encoder* do *codec* CS-ACELP (G729a), com a finalidade de mensurar uma ponderação do MOS. Essa métrica é a atual referência para avaliação de qualidade do sinal de áudio percebido nos sistemas de telecomunicações.

A figura abaixo representa o bloco de origem do codificador que fará a análise do processo de avaliação perceptual durante a transmissão de dados:

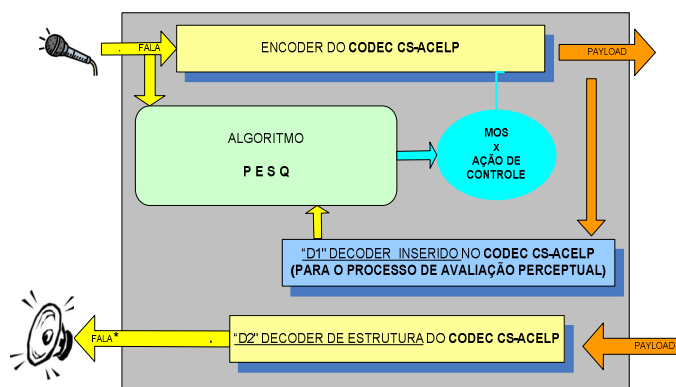


Figura. 4. Diagrama de blocos simplificado de análise do codec

- o bloco do *encoder* realiza a codificação do sinal, em que foi realizada a segmentação para leitura de um número específico de frames do sinal de entrada;

- o bloco do *decoder* foi duplicado na estrutura. O bloco “D2” é estrutura normal do codificador (*encoder/ decoder*), serve para decodificar o sinal recebido. O bloco “D1” inserido na estrutura do codificador é responsável pela decodificação dos frames codificados para transmissão. Esses frames necessitam ser decodificados para avaliação perceptual do sinal;

- o bloco do PESQ é responsável por receber o sinal de referência da entrada do codificador e o sinal que passou pelo *encoder* e foi codificado para envio e transmissão. Esse mesmo sinal passa por um *decoder* “D1” e é decodificado para avaliação perceptual;

- a indicação de MOS, ou seja, a pontuação dos sinais avaliados incide sobre uma ação de controle no bloco do *encoder*, alterando os frames seguintes à avaliação presente.

O principal objetivo desta análise é avaliar o sinal do ponto de vista perceptual e esta verificação ocorre em tempo real com ações de controle no bloco de origem de codificação, o *encoder*, para garantir e melhorar perceptualmente os frames codificados.

Durante essa verificação de parâmetros, o codificador deverá realizar o processo acima, ininterruptamente, executando uma heurística.

### IV. METODOLOGIA

Para a elaboração e o desenvolvimento de um padrão de avaliação perceptual aceitável que validasse a proposta, a

pesquisa foi realizada com base nos parâmetros avaliados no algoritmo do PESQ e nos parâmetros do codificador G729a. Para isso, a seguinte heurística foi utilizada:

1- A codificação inicializa com os ganhos originais nos três primeiros frames, para que o sinal de entrada seja estável, evitando a leitura de erro devido a uma característica interna do codificador.

2- Analisa-se a tendência da pontuação do MOS, como medida numérica de qualidade da voz humana, utilizando-se a leitura e escrita dos ganhos avaliados anteriormente. Esses ganhos são aplicados para cada conjunto de orador, conforme definido no experimento.

3- Como inicialização do modelo de comparação, é realizada a leitura do valor da análise do PESQ das amostras sequenciais dos arquivos comparados, verificando o valor de MOS da amostra atual em relação ao valor de MOS da amostra anterior.

4 - Essas amostras são avaliadas em fragmentos sequenciais de 28 frames do PESQ, menor quantidade de frames avaliados.

5- Não será realizada nenhuma ação de controle quando houver intervalos de frames avaliados pelo PESQ sem um valor de MOS.

6- O controle foi realizado em três estágios:

i) Se a tendência dessas amostras (intervalo de frames analisados sequencialmente) apresentarem um valor de MOS do intervalo de frames atual maior do que o valor de frames anterior, um ganho de *pitch* é adicionado ao processo.

ii) Se houver um valor de MOS regressivo durante três frames consecutivos, o codificador deverá alterar seus parâmetros internos para que o ganho retorne ao original, mantendo ao valores de codificação. A escolha por três intervalos de frames avaliados consecutivamente pelo PESQ é para manter a ação de controle dentro do tempo de avaliação padrão do algoritmo (100 frames).

iii) Quando o ganho estiver recebendo o sinal original, ele retorna para o passo “i” desta heurística e, assim, sucessivamente, até encerrar a avaliação dos frames recebidos.

Para solução da metodologia aplicada nesta pesquisa realizou-se testes e validação em software, efetuando as modificações propostas com base nas pesquisas dos artigos do IEEE (*Institute of Electrical and Electronics Engineers*), bibliografia relacionada e análise da recomendação ITU-T [1],[2],[3],e [4].

Para a elaboração deste projeto, não foram considerados os parâmetros ofensores da qualidade de voz inerente ao meio de transmissão (perda de pacotes, o *delay* e o *jitter*) [8]nem o tempo de processamento para a solução proposta. Os testes foram realizados sobre o bloco de origem do sinal de voz no *encoder*, não foram alterados os coeficientes LPC [9], pois são extremamente sensíveis ao erro de quantização.

### V. TESTES REALIZADOS

Os testes realizados foram baseados na utilização de arquivos de voz em formato PCM de 16 bits, amostrados a 8KHz, que serviram de entrada para o codificador original CS-ACELP e sua versão modificada.

Para que o experimento tivesse uma análise criteriosa, foram estudadas e inseridas as premissas avaliadas

anteriormente na metodologia alterando parâmetros de acordo com a heurística proposta.

Os gráficos abaixo apresentam os valores analisados para cada um dos arquivos de teste extraídos do banco de dado Timit[10], formados por seis oradores. As frases foram agrupadas com o *software audacity*, o que possibilitou um período de análise suficiente para avaliação dos resultados.

Orador Masculino:

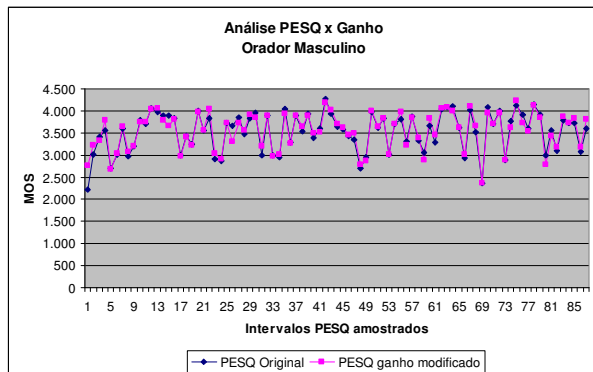


Figura 5. Gráfico de aplicação de ganho para oradores masculinos

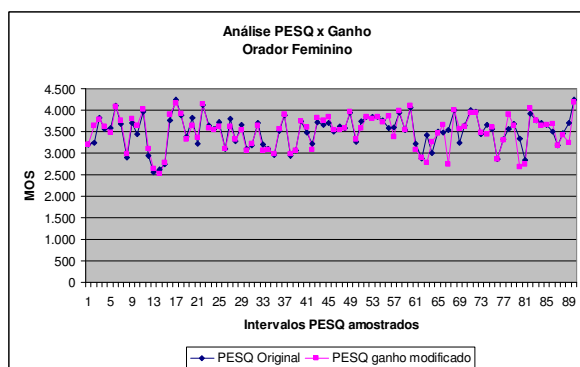


Figura 6. Gráfico de aplicação de ganho para oradores femininos

No primeiro teste incluiu-se o PESQ e aplicou-se um percentual de ganho de *pitch*, realizado de forma serial em todos os frames avaliados. Obtiveram-se os valores de ganho de 10% para o arquivo1 de oradores femininos e 1,54% para o arquivo1 de oradores masculinos. Aplicando esses valores obteve-se 61,54% de melhora de frames avaliados para o orador masculino e 58,62% de melhora de frames avaliados para o orador feminino. Isso ocorreu sem a heurística definida para o teste, pois esta foi a primeira avaliação.

## VI. RESULTADOS OBTIDOS

O estudo avaliou as respostas do codificador sob ponto de vista de qualidade perceptual, obtendo a melhora perceptual no sinal de entrada nas amostras do orador masculino e nas amostras do orador feminino. Os ganhos foram avaliados serialmente e observou-se que um acréscimo no ganho de *pitch* do sinal amostrado foi a forma mais significativa de avaliar o sinal.

Quando avaliado perceptualmente o sinal de voz, existem frames garantidos (frames que mantiveram sua pontuação MOS) e frames melhorados (frames com pontuação MOS superior ao arquivo original). Abaixo, o sinal avaliado a cada intervalo de 28 frames.

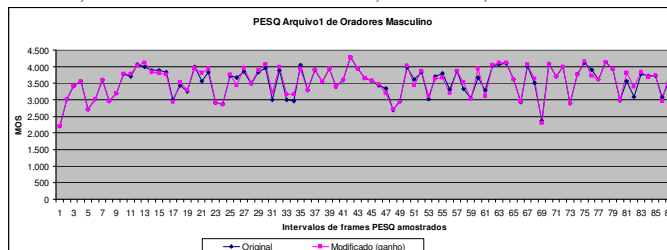


Figura 7. Aplicação da heurística no arquivo de oradores masculinos

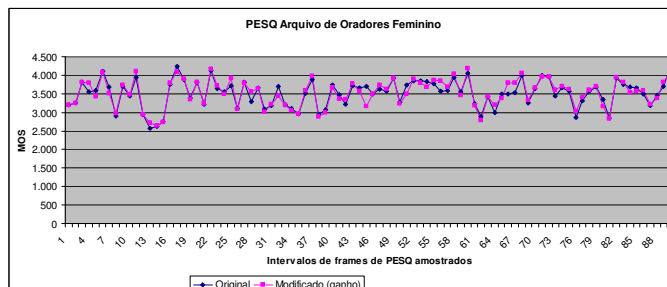


Figura 8. Gráfico de aplicação da heurística no arquivo de oradores femininos

A heurística realizada para obter resultados de melhoria perceptual do sinal de voz foi efetuada com sucesso. Abaixo, um resumo dos parâmetros avaliados para cada conjunto de oradores.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	<b>307,366</b>
somatório acumulado de frames do PESQ com ganho pitch adicional	<b>308,983</b>
somatório acumulado de frames do PESQ com escolha do melhor valor	<b>311,001</b>
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	<b>72,41%</b>
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	<b>44,83%</b>

TABELA II. RESULTADOS OBTIDOS PARA ORADORES MASCULINOS.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	<b>316,328</b>
somatório acumulado de frames do PESQ com ganho pitch adicional	<b>317,365</b>
somatório acumulado de frames do PESQ com escolha do melhor valor	<b>329,842</b>
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	<b>55,56%</b>
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	<b>50,0%</b>

TABELA III. RESULTADOS OBTIDOS PARA ORADORES FEMININOS

## VII. CONCLUSÕES

A avaliação foi realizada sobre um modelo de codificação da voz que analisa a qualidade perceptual. Foi inserido dentro do código fonte do codificador G729a, o algoritmo do PESQ. Este critério de decisão para as intervenções no sistema teve importantes considerações a ressaltar após estudos sobre o Processamento digital de Sinais.

1-O codificador, ao iniciar o processamento do sinal de voz para codificação, utiliza como referência o valor mais expressivo da escala de quantização do ganho de *pitch*. Esse processo quando constatado foi excluído do critério de avaliação de decisão, retirando a ação de controle dos dois primeiros frames, visto que a influência desses frames distorcia a análise do sinal, quando executando a heurística proposta.

2- A ausência de sinal também foi um item importante da análise. Segundo (ITU P862, 2001), a avaliação perceptual é obtida pelo algoritmo do PESQ somente quando existe um período de silêncio inferior a 25% do frame de voz analisado. Sendo este o entendimento para os frames não avaliados quando executado o algoritmo.

3- O mapeamento dos ganhos de *pitch* e de codificação não quantizados no codificador G729a é importante para avaliar a pontuação mais significativa. Em todas as análises realizadas, utilizou-se somente o ganho de *pitch*. Foi obtida uma melhor avaliação perceptual quando efetuada adição de ganho. Deste modo, desconsiderou-se a utilização do ganho de codificação como parâmetro para melhoria de qualidade do sinal.

4- Foram realizados testes com os dois ganhos simultâneos, *pitch* e codificação. Entretanto, com o padrão de avaliação adotado pela heurística, não foi obtido sucesso na melhora do processo, pois os valores de ganhos foram verificados separadamente e não estão automatizados.

5- Foram verificados parâmetros específicos do codificador como os valores do “GAMMA”(variável no código fonte do G729a). Esse parâmetro altera as características do filtro de ponderação do codificador. Entretanto, este é um valor fixo no *encoder* e no *decoder*. Assim, tendo como proposta desta pesquisa a avaliação do sinal na origem, não foram realizadas alterações em parâmetros tabelados pelo codificador para não alterar o processo de decodificação do sinal.

6- A proposta de modificar frames a frames o sinal de voz dinamicamente, ou seja, avaliados a cada intervalo de 28 frames pelo PESQ foi realizada com sucesso. Apesar de um valor baixo para o somatório do PESQ, a melhoria perceptual ocorreu. Constatou-se que o PESQ efetua a média das avaliações do sinal de voz amostrado. Na análise de um arquivo inteiro o valor do PESQ não altera muito a pontuação referida na escala MOS, com variação de 0 à 5 pontos. Neste sentido, cabe ressaltar que o algoritmo do PESQ aplica a pontuação máxima de 4,5 pontos para sinais idênticos.

7- A heurística realizada para obter resultados de melhoria perceptual do sinal de voz foi efetuada com sucesso. Abaixo, um resumo dos parâmetros avaliados para cada conjunto de oradores. O somatório acumulado é a soma do valor da pontuação MOS que cada intervalo de 28 frames do PESQ.

descrição do parâmetros avaliados	masc 01	masc 02	fem 01	fem 02
1) somatório acumulado de frames do PESQ com ganho do pitch do sistema	307,366	426,227	316,328	428,255
2) somatório acumulado de frames do PESQ com ganho pitch adicional	308,983	426,799	317,365	429,932
3) somatório acumulado de frames do PESQ com escolha do melhor valor	311,001	433,666	329,842	435,92
4) % de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	72,41%	52,94%	55,56%	61,65%
5) % de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	44,83%	46,32%	50,00%	50,38%
6) % de ganho de pitch aplicado, valor escolhido pela análise de ganho para os oradores	1,54%	2,86%	10,00%	1,82%

TABELA IV. ANÁLISE DO GANHO PITCH SOBRE A PROPOSTA DE AVALIAÇÃO

Na “linha 1” da tabela acima tem o resultado da avaliação do sinal de voz original. Na “linha 2”, o resultado do sinal de voz avaliado com a heurística proposta sempre obteve melhora no sinal analisado, conforme valores dos parâmetros da “linha 4” e da linha “5”.

Não foi implementada a heurística de controle do melhor sinal, devido a necessidade de armazenamento temporário(*buffer*) para escolha do sinal. De acordo com a tabela IV, “linha 3”, a melhoria de desempenho seria muito pequena para justificar a carga computacional necessária para aplicação da heurística de controle do melhor sinal.

8- O ganho aplicado ao sinal de voz para avaliação perceptual é verificado para cada orador ou conjunto de oradores, devido à variação do sinal, sendo essa análise, um item particular e importante no processo de melhoria. Conforme descrito na “linha 6”, os ganhos não possuem uma relação direta com o orador e sim com a frequência de *pitch*.

9- A avaliação do sinal de voz, tendo como origem o bloco de codificação do sinal, é uma análise a ser explorada, uma vez que o orador tenha o seu ganho adicional de *pitch* mapeado na fonte de codificação. O sistema tem condições de realizar interações no processo para melhoria do sinal de voz codificado.

Neste sentido, essa avaliação tem uma contribuição para a melhora dos sinais em equipamentos de uso pessoal como *smartphones* e *tablets*. O orador principal (usualmente o proprietário do equipamento codificador) pode ter suas características aprendidas pelo codificador, durante o uso prolongado pela contínua interação com o equipamento.

## VIII. REFERENCES

- [1] (ITU G729.A, 1996) ITU-T. *Reduce complexity 8 kbit/s CS-ACELP speech codec*. International Telecommunications Union, 11/1996. (ITU-T Recommendation G.729.0 Annex A). <http://www.itu.int/rec/T-REC-G.729-199611-1/AnnA/en>
- [2] (ITU G711, 1988) ITU-T. *Pulse code modulation (PCM) of voice frequencies*. International Telecommunications Union, 11/1988. (ITU-T Recommendation G.711.0). <http://www.itu.int/rec/T-REC-G.711-198811-1/en>
- [3] ITU-T Recommendation P.800.1, *Methods for subjective determination of transmission quality*. Genève, Agosto/1996.
- [4] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Genève, Novembro/2005.
- [5] (MAKHOUL, 1975) John Makhoul. *Publicação: Linear Prediction: A Tutorial Review*. Proceedings of the IEEE, Volume 63, nro 4, Abril/1975. 16/06/2010 IEEE Xplore.
- [6] (BULTHEEL e BAREL, ) Adhemar Bultheel and Marc Van Barel, *Paper: Linear prediction: mathematics and engineering*. Department of Computing Science, K.U.Leuven, Belgium. Acessado em junho/2010. <http://www.emis.de/journals/BBMS/Bulletin/bu1941/BULTHEEL.PDF>
- [7] (YNOGUTI) Carlos Alberto Ynoguti. *Notas de aula: Psicoacústica e Codificação Perceptual*, Inatel. Disponível em <http://cict.inatel.br/nova2/docentes/ynoguti/e724/Audio.pdf>
- [8] Demóstenes Zegarra Rodriguez, Miguel Arjona Ramirez “Determinação do comportamento de codificadores em diferentes condições de rede em um cenário experimental”, Disponível em [http://www.lps.usp.br/~miguel/publica/sbrt2008\\_42576.pdf](http://www.lps.usp.br/~miguel/publica/sbrt2008_42576.pdf) Acessado em Fevereiro/2010
- [9] (O'SHAUGHNESSY, 1999) Douglas O'Shaughnessy, *Speech communications, human and machine*, Wiley-IEEE Press, Montreal, 1999 (2 edition).
- [10] TIMIT Acoustic-Phonetic Continuous Speech Corpus, reference: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1 in University of Pennsylvania>.
- [11] A. M. KONDOZ, *Digital Speech: Coding for low rate communication systems*. New York: John Wiley & Sons, 2004.
- [12] (HAYKIN e VEEN, 1999) Simon Haykin e Barry Van Veen, *Signal and Systems*, John Willey & Sons Inc, 1999 (Re-impressão 2007).
- [13] (RABINER e SCHAFER, 1978) Laurence Rabiner e Ronald W. Schaffer, *Digital Processing of Speech Signals*, Prentice Hall, US, 1978.
- [14] (PROAKIS e MANOLAKIS, Jhon G. Proakis e Dimitris G. Manolakis, *Digital Signal Processing, Principles, Algorithms. And Applications*, Prentice Hall, 2007 (4° edition).
- [15] Flávio Luis Wisnevski, Rubem Dutra Ribeiro Fagundes, “Codificador G729a orientado à avaliação da qualidade perceptual do sinal de voz”, Disponível em [http://tede.pucrs.br/de\\_arquivos/11/TDE-2011-11-07T141109Z-3511/Publico/434559.pdf](http://tede.pucrs.br/de_arquivos/11/TDE-2011-11-07T141109Z-3511/Publico/434559.pdf)