# Sistemas de Criptofonia sob Influência de Canais de Comunicações Móveis

José F. de Andrade Jr., Marcello L. R. de Campos e José A. Apolinário Jr.

***Resumo*— Técnicas de criptofonia são utilizadas para transformar sinais de voz em sinais ininteligíveis. Para sistemas de comunicações móveis que empregam Codificadores/Decodificadores (CODECs) comerciais, a encriptação digital não é uma opção adequada por não ser viável antes dos mesmos ou por requerer modificações de hardware e software. Se a encriptação for aplicada antes do CODEC, isto poderá resultar em uma baixa qualidade de voz, i.e., o CODEC tentaria codificar o sinal cifrado como se fosse sinal de voz. Alternativamente, alguns cifradores analógicos podem ser empregados sem causar grandes alterações no desempenho de codificação. Este trabalho investiga a degradação causada em sistemas de criptofonia analógicos no domínio da freqüência pelos efeitos de multipercursos de canais rádio presentes em telefonia celular. São apresentados resultados objetivos de qualidade pela metodologia PESQ (*Perceptual Evaluation of Speech Quality*) para sinais criptofonados e aplicados ao CODEC.**

***Palavras-Chave*— criptofonia, scramblers, comunicações móveis, GSM-FR, processamento de voz.**

***Abstract*— Speech privacy techniques are used to scramble clear speech into an unintelligible signal in order to avoid eavesdropping. Digital encryption may not be an option to provide speech privacy to mobile communication systems using commercial CODECs due to its inviable use before them or to the fact that it would require internal hardware and software modifications. If encryption is applied before the CODEC, poor voice quality may result, i.e, the CODEC would handle digitally encrypted signal as a noise. Alternatively, some scramblers may be placed before CODEC without causing much penalty to its performance. This work investigates the quality degradation of frequency domain scrambled signals caused by multipath effects present in cellular radios channels. Objective results in terms of PESQ versus channel SNR are presented to scrambled signals applied to the GSM-FR CODEC.**

***Keywords*— speech privacy, scramblers, mobile communications, GSM-FR, speech processing.**

## I. Introdução

Técnicas de criptofonia são utilizadas para transformar um sinal de voz em sinal ininteligível, cujo propósito é evitar escutas não autorizadas. Equipamentos denominados cifradores analógicos vêm sendo sistematicamente substituídos por equipamentos de criptofonia digital, que possuem maior grau de segurança, mas exigem técnicas de implementação mais complexas e maior largura de banda para transmissão. Quando se deseja implementar sigilo em sistemas móveis comerciais que empregam CODEC, tais como GSM-FR (*Global System for Mobile communications-Full Rate*) e AMR (*Adaptive Multirate*) [1], a encriptação digital pode não ser uma opção adequada devido à necessidade de alterações internas de hardware e software, bem como da expansão da largura da banda base do sinal. Se o sinal encriptado por técnicas digitais for aplicado diretamente ao CODEC, em decorrência do sinal não preservar as características de um sinal de voz, a codificação produzirá um sinal de baixa qualidade.

Por outro lado, cifradores analógicos no domínio freqüência podem ser empregados antes de codificadores de voz sem causar grandes alterações no desempenho do processo de codificação [2]; contudo, a qualidade do sinal recuperado (decifrado) é fortemente influenciada pelas condições do canal.

Este trabalho apresenta resultados de avaliação de qualidade pelo método PESQ para sinais decifrados em função da relação sinal ruído do canal. São apresentados resultados para os canais tipicamente urbano (TU6) e rural.

José F. de Andrade Jr., Marcello L. R. de Campos e José A. Apolinário Jr¸ Programa de Engenharia Elétrica, Coordenação de Pos-graduação em Engenharia, Universidade Federal do Rio de Janeiro e Instituto Militar de Engenharia, Rio de Janeiro, Brasil, E-mail: ps7jfa@urbi.com.br; campos@lps.ufrj.br; e apolin@ime.eb.br. Este trabalho foi parcialmente apoiado pelo CNPq (Processos 306445/2007-7 e 310258/2006-5) e Marinha do Brasil.

## II. Cifradores Analógicos no Domínio da Freqüência

Os primeiros cifradores analógicos no domínio da freqüência (CAF) empregavam a técnica de inversão de freqüência, que consiste na inversão do espectro do sinal ou de parte deste com o intuito de tornar o sinal ininteligível aos ouvintes que não possuam receptores capazes de desfazer a inversão espectral do sinal. Estes inversores, devido à simplicidade de se desfazer o processo de criptofonia, não são mais empregados.

Com o surgimento de novos processadores digitais de sinais (DSPs), capazes de realizar tarefas complexas com alto nível de miniaturização, foi possível projetar sistemas de CAF implementados com bancos de filtros e transformadas ortogonais [2], [4], [5].

Se o número de sub-bandas (ou subfaixas) for pequeno, o sinal apresentará uma inteligibilidade residual mais elevada. Para superar este problema, deve-se escolher um número mínimo de sub-bandas e uma chave (permutação) dentre aquelas que geram baixa inteligibilidade residual. Outra forma de melhorar o desempenho dos sistemas de CAF, é realizar alterações das chaves de maneira periódica e aleatória, de acordo com um polinômio gerador de seqüências pseudo-aleatórias.

Um sistema de CAF possui nível de segurança que varia de casual a tático e, para o caso em que se empregam seqüências pseudo-aleatória de chaves, consegue-se melhorar a segurança pouco acima do nível tático [6]. A mudança pseudo-aleatória de chaves produz uma distribuição aleatória da energia do

sinal pelas suas diversas subfaixas no período de tempo considerado, o que aumenta a resistência à criptoanálise.

O diagrama da Figura 2 representa um banco de filtros [7] com $M$ subfaixas capazes de cobrir todo o espectro de sinal de voz a ser cifrado. Após a filtragem pelo conjunto de filtros de análise $H_k(z)$ e decimação crítica por um fator $M$, as subfaixas são permutadas de acordo com a matriz de permutação $\boldsymbol{P}$.

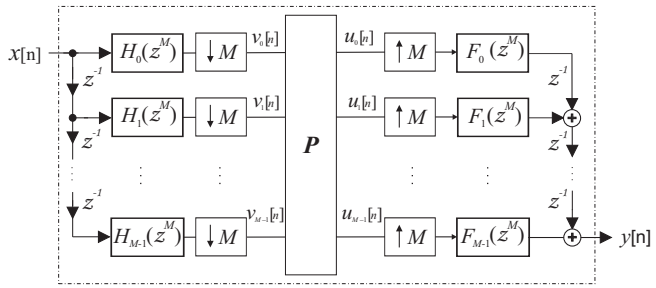


Fig. 1. CAF baseado em banco de filtros.

Para um sinal de voz dividido em blocos, onde o *i*-ésimo bloco pode ser representado pelo vetor $\boldsymbol{x}_i$, as amostras $V_k^i[z^M]$ pertencentes ao i-ésimo bloco representam a k-ésima subfaixa no domínio $z$, expressa como:

$$V_k^i[z] = H_k(z^M)X_k^i[z^M] \quad ,\forall\ k = 0, 1, \ldots, M-1. \tag{1}$$

Os vetores $V_k^i[z^M]$ são, então, rearranjados na forma de uma matriz com $M \times N$ elementos, onde $N-1$ é a ordem do filtro $H_k(z)$:

$$\boldsymbol{V}_i = [V_0^i\ V_2^i\ \cdots\ V_{M-1}^i]^T. \tag{2}$$

A multiplicação da matriz $\boldsymbol{V}_i$ pela matriz de permutação $\boldsymbol{P}$ resulta na matriz $\boldsymbol{U}_i$, cujas linhas estão dispostas segundo à permutação aplicada. Cada linha de $\boldsymbol{U}_i$ representa uma das sub-bandas do i-ésimo bloco cifrado.

$$\boldsymbol{U}_i = \boldsymbol{P}\boldsymbol{V_i} \tag{3}$$

$$\boldsymbol{U}_i = [U_0^i\ U_2^i\ \cdots\ U_{M-1}^i]^T \tag{4}$$

$$Y_k^i[z^M] = F_k(z^M)U_k^i[z^M] \quad ,\ k = 0, 1, \ldots, M-1 \tag{5}$$

$$\boldsymbol{Y}_i = [Y_0^i\ Y_2^i\ \cdots\ Y_{M-1}^i]^T \tag{6}$$

O sinal cifrado referente ao i-ésimo bloco é, portanto, obtido após a interpolação e efetuando-se o somatório elemento-a-elemento de cada linha de $\boldsymbol{Y}_i$. O resultado, no domínio da transformada $\mathscr{Z}$, é:

$$Y(z) = \sum_{k=0}^{M-1} z^{-k}Y_k^i[z^M]. \tag{7}$$

O sinal a ser aplicado ao CODEC deverá estar na forma analógica, portanto, antes de submeter o sinal cifrado $y[n]$ ao CODEC, o mesmo deverá ser reconvertido para a forma analógica.

Para decifrar o sinal criptofonado, pode-se utilizar o mesmo processo supra-mencionado, tomando cuidado de substituir a matriz de permutação $\boldsymbol{P}$ por sua inversa $\boldsymbol{P}^{-1}$.

A adoção de filtros com atenuação abrupta a partir da freqüência de corte confere à modalidade de CAF um importante diferencial, que é a imunidade à perda de sincronismo de quadro [8], tornando-a bastante atrativa para projetos de baixo custo aplicados a equipamentos de arquitetura fechada.

## III. Canal de Rádio em Comunicações Móveis

O canal de rádio, no contexto das comunicações móveis, é de grande importância no desempenho do sistema e, sempre que possível, deve ser caracterizado fisicamente. A maioria dos canais pode ser caracterizada pelos efeitos causados pela mobilidade da estação móvel (EM). As características de mobilidade da EM são responsáveis pela maioria dos desvanecimentos e por parte da dinâmica dos multipercursos. Para realizar uma caracterização mais realista, deve-se levar em conta, também, os efeitos provocados pelos fenômenos atmosféricos.

Como principal efeito da atmosfera sobre o canal de rádio, tem-se a variação do índice de refração. Esta variação influencia diretamente os fenômenos de reflexão, difração e dispersão. Devido à dinâmica do índice de refração, portanto, são criados os vários percursos de comunicação entre o transmissor e o receptor, os quais não igualmente percorridos pela mesma onda eletromagnética, resultando em uma composição de ondas no receptor.

Cada percurso possui um atraso e uma atenuação particular, que variam segundo o índice de refração e a mobilidade da EM. Esta variação temporal do sinal provoca flutuações de amplitude e fase do sinal no receptor. Este efeito é conhecido como *fading* ou, em português, desvanecimento. Existem diversos tipos de desvanecimentos, sendo de interesse para este trabalho os desvanecimentos multipercurso que não afetam a potência média do sinal nem provocam perdas de percursos. Estes desvanecimentos causam apenas "cintilações" na amplitude e fase do sinal.

O desvanecimento provocado pelos múltiplos percursos é fortemente influenciado por obstáculos, sejam estes naturais ou artificiais. Desta maneira, na caracterização de canais de comunicações móveis, deve-se levar em consideração a configuração do conjunto de obstáculos existentes no ambiente e o movimento relativo entre o transmisssor/receptor e estes obstáculos. Este fato resulta na caracterização particular de cada tipo de ambiente. A Tabela I apresenta o conjunto das atenuações ($A[dB]$) e atrasos ($\tau[\mu s]$) referentes aos ambientes tipicamente rural e urbano [9].

TABELA I
Atrasos e Atenuações Referentes a Canais Típicos para Áreas Rural e Urbana.

| Tipicamente Rural | | Tipicamente Urbano | |
|---|---|---|---|
| $\tau$ [$\mu s$] | $A$ [dB] | $\tau$ [$\mu s$] | $A$ [dB] |
| $0,0$ | $0$ | $0,0$ | $-3$ |
| $0,1$ | $-4$ | $0,2$ | $0$ |
| $0,2$ | $-8$ | $0,5$ | $-2$ |
| $0,3$ | $-12$ | $1,6$ | $-6$ |
| $0,4$ | $-16$ | $2,3$ | $-8$ |
| $0,5$ | $-20$ | $5,0$ | $-10$ |

Conforme esquematizado na Figura 1, a resposta do canal ao impulso é dada pelo somatório das respostas de todos os percursos, levando-se em consideração os atrasos e atenuações correspondentes à cada percurso. O coeficiente de atenuação $A_{i,j}$, no geral, é um número complexo, que representa a atenuação e a diferença de fase que o $j$-ésimo percurso provoca para o $i$-ésimo usuário.

Devido à imprevisibilidade do número de percursos, que surgem e desaparecem dinamicamente, pode-se modelar um canal rádio para comunicações móveis como um processo estocástico. A função densidade de probabilidade adotada neste trabalho para as simulações dos efeitos multipercursos foi a distribuição *Rayleigh* [10].

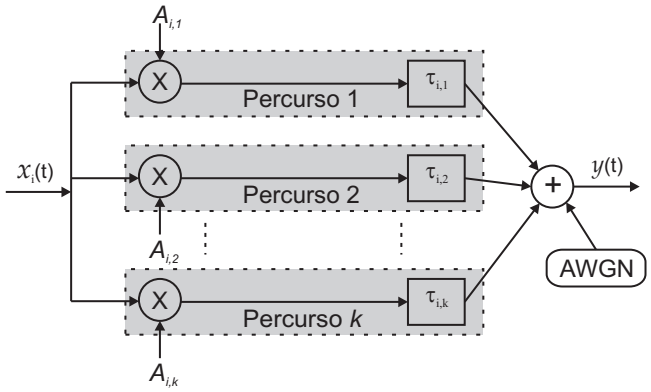


Fig. 2. Modelagem de canal multipercursos do $i$-ésimo usuário.

O sinal $y(t)$ ao chegar no receptor pode ser expresso matematicamente como:

$$y(t) = \sum_{i=1}^{n} \sum_{j=1}^{k} A_{i,j} x_i(t - \tau_{i,j}) \; + \; \eta(t), \qquad (8)$$

onde $n$ representa o número de usuários que compartilham o canal, $k$ é o número de percursos para um dado instante de tempo e $\eta(t)$ a tensão gerada pelo ruído branco aditivo gaussiano (AWGN).

De uma maneira geral, os efeitos provocados pelo canal podem ser classificados como espalhamento temporal e variações temporais. Em relação ao período de símbolo transmitido, os desvanecimentos podem ser classificados como rápidos ou lentos, de acordo como a sua duração. Em relação à magnitude do espalhamento temporal, pode-se classificar o canal como seletivo em freqüência, quando se tem grandes valores de espalhamentos, e não-seletivo, quando pequenos espalhamentos estão presentes.

## IV. Medida Objetiva de Qualidade

Este trabalho se limitará ao emprego do algoritmo PESQ [11] como método de avaliação perceptual de qualidade de sinal.

Medidas perceptuais são medidas obtidas por meio de algoritmos que fazem uso de modelos psico-acústicos com o propósito de reproduzir parcialmente características do ouvido humano.

O algoritmo PESQ é um algoritmo para medida de qualidade de voz em sistemas de telefonia, cuja descrição se encontra nas recomendações da ITU-T P.862 e P.862.3 [11], [12]. Este padrão incorpora recursos que permitem a avaliação de sistemas de telefonia como GSM, VoIP e ISDN. Um fator de relevância no algoritmo PESQ é que a medida de qualidade é apresentada diretamente na escala *Mean Opinion Score* (MOS) [13].

TABELA II
ESCALA MOS.

| MOS | Qualidade |
|---|---|
| 5 | Excelente |
| 4 | Bom |
| 3 | Razoável |
| 2 | Pobre |
| 1 | Ruim |

A Tabela II apresenta uma associação do índice MOS, resultado do algoritmo PESQ, com os resultados subjetivos traduzidos em termos de avaliação de qualidade.

Para uma conversação realizada em uma rede de telefonia celular do padrão GSM, o valor típico obtido pelo algoritmo PESQ é de $3,50$.

Devido ao bom desempenho em relação aos outros algoritmos existentes para avaliação da qualidade dos sinais de voz, o algoritmo PESQ tornou-se o padrão (IUT-T P862) para avaliação de qualidade de voz em redes de telefonia de banda estreita e CODEC de sinais de voz.

## V. Descrição do Simulador Utilizado

A simulação da influência do canal sobre sinais de voz criptofonados, foi realizada reproduzindo-se o canal de tráfego de voz do sistema para o codificador de voz GSM *Full Rate* (TCH-FS) [14]. Os efeitos do ruído branco aditivo gaussiano e os desvanecimentos por multipercurso também foram considerados.

O codificador de voz *Full Rate*, baseado no codificador de predição linear com excitação residual (RELP) e associado a um dispositivo de predição de longo prazo (LTP) fornece 260 bits, a uma taxa de $13,0$ $kbps$ ao TCH-FS. Estes 260 bits, que correspondem a um quadro de $20$ $ms$ do sinal de voz, são, então, reordenados segundo o grau de importância dos seus respectivos bits. Os bits são rearranjados de maneira que na disposição resultante a maior probabilidade de ocorrência de erros aconteça para os bits que afetam menos a qualidade da voz decodificada. A qualidade da voz é mais afetada pelos bits de coeficientes mais significativos do que pelos bits de coeficientes menos significativos.

É importante mencionar que após a permutação das subbandas do sinal original, a reordenação de bits realizada pelo codificador não resultará obrigatoriamente em uma melhoria da qualidade do sinal recuperado, pois a versão cifrada do sinal não possui puramente características de um sinal de voz.

Os bits de menor importância (78 bits), denominados bits classe 2 , não recebem mecanismos de correção nem detecção de erros. Os bits mais importantes (50 bits), denominados bits classe $1a$, recebem códigos de detecção de erros do tipo *Cyclic*

*Redundancy Check*-CRC. Para os bits da classe $1a$ e os bits de média importância (132 bits), classificados como classe $1b$, ocorre a inclusão de um código convolucional de meia taxa (378 bits) para correção de erros. Após a manipulação dos bits originais com a inclusão dos códigos corretores e detectores de erros, 456 bits são produzidos a partir de 20 $ms$ do sinal de voz. Este conjunto de 456 bits é submetido a uma matriz de *interleaving* que divide este conjunto em 8 grupos de 57 bits. Estes oito grupos são utilizados para composição dos *burst* ortogonais, formados por quatro grupos de 114 bits, que são, então, modulados e transmitidos pelo canal.

A demodulação e decodificação do sinal é realizada pelo processo inverso, tendo como destaque o emprego de um decodificador de *Viterbi* [15].

## VI. Resultados

Os resultados desta seção foram obtidos para um conjunto de 50 frases fonadas na língua portuguesa por 10 locutores distintos. A captação das frases foi realizada por meio de microfones de cápsula de eletreto em ambiente de baixo ruído.

As frases foram cifradas pela metodologia descrita seção II e qualidade avaliada na Referência [2]. O número de subfaixas empregado foi de 8 e a matriz de permutação (chave) fixa.

Para o cálculo do efeito Doppler, foi considerada a velocidade da EM de 30 $km/h$, que corresponde a um desvio máximo de $37,5$ $Hz$.

A Tabela III mostra os valores PESQ (MOS) versus SNR obtidos para o sinal decifrado (degradado), quando comparado com o sinal original. As simulações consideraram modelos de canais para os ambientes tipicamente urbano e rural.

Para se ter uma idéia somente influência do ruído, sem associação ao desvanecimento, foram realizadas simulações acrescentando-se apenas um canal do tipo AWGN.

TABELA III
VALORES PESQ (MOS) VERSUS SNR PARA OS CANAIS AWGN, TIPICAMENTE RURAL (TR6) E TIPICAMENTE URBANO (TU6) .

| SNR (dB) | TU6 | TR6 | AWGN |
|---|---|---|---|
| 5 | 1,44 | 1,38 | 2,96 |
| 10 | 1,64 | 1,51 | 2,98 |
| 15 | 1,97 | 1,95 | 2,99 |
| 20 | 2,23 | 2,29 | 3,00 |
| 25 | 2,35 | 2,57 | 3,01 |
| 30 | 2,48 | 2,77 | 3,02 |

## VII. Conclusões

A existência de um canal com efeitos multipercursos e AWGN (Aditive White Gaussian Noise) influencia na qualidade de sinais que tenham sido criptofonados por meio de sistemas de CAF e recuperados após a passagem por este tipo de canal. De acordo com os resultados do gráfico apresentado pela Figura 2, pode-se observar que a recuperação do sinal com qualidade aceitável ocorre a partir de uma relação sinal ruído SNR maior ou igual a 15 dB. Testes subjetivos de audição dos vários arquivos utilizados na simulação comprovaram este resultado objetivo. Para uma SNR=10 dB, é possível

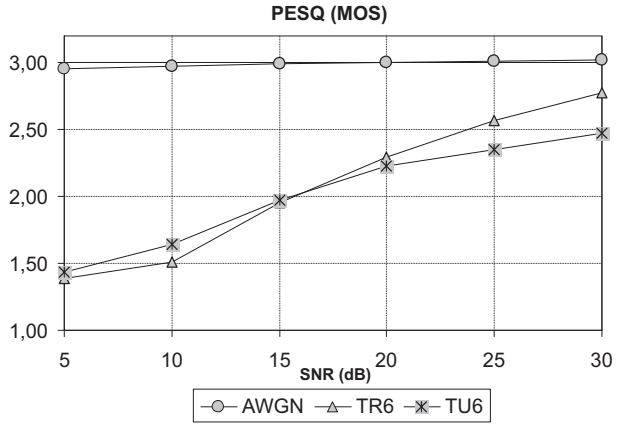


Fig. 3. Avaliação de qualidade para canais do tipo AWGN, TR6 e TU6.

compreender o conteúdo de voz com uma qualidade somente aceitável para situações momentâneas durante um processo de conversação real. Para SNR=5 dB não é possível compreender o conteúdo de voz.

Quando apenas o canal AWGN é considerado, quaisquer valores de SNR, maiores ou igual a 6 dB, permitem a recuperação do sinal com qualidade considerada razoável. Este fato é justificado pelo emprego de códigos detectores de erro do tipo *Cyclic Redundancy Check* para os bits da classe $1a$ e pela codificação convolucional com taxa $1/2$ dispensada aos bits da classe $1a$ e $1b$.

Com base nos resultados apresentados, pode-se concluir que, para condições normais de enlace, é possível empregar técnicas de CAF sobre um canal GSM-FR. Como sugestão para trabalhos futuros, pode-se implementar e simular a influência de canais típicos de telefonia móvel sobre sinais criptofonados com técnicas de CAF e codificados por CODEC do tipo AMR (taxa variável) bem como a avaliação subjetiva (MOS levantado a partir de diversos ouvintes).

## Agradecimentos

Ao Laboratório de Voz do Instituto Militar de Engenharia, pela cessão dos arquivos de voz utilizados nas simulações apresentadas neste trabalho.

## Referências

[1] 3GPP TS 26.071 V6.0.0 (2004-12) Technical Specification 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Mandatory speech CODEC speech processing functions; AMR speech CODEC.

[2] J. F. de Andrade Jr., M. L. R Campos, and J. A. Apolinário Jr.,"Speech privacy for modern mobile communication systems," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008)*, pp. 1777–1780, April 2008.

[3] GSM 06.60 version 8.0.1 (2000-11) Technical Specification 3rd Generation Partnership Project; TDigital cellular telecommunications system (Phase 2+); Enhanced Full Rate (EFR) speech transcoding.

[4] B. Goldburg AND S. Sridharan, "Design and cryptanalysis of transform-based scramblers,"*Jornal on Selected Areas on Communications IEEE*, vol. 11, no. 5, pp. 735–744, June 1993.

[5] M. S. Ehsani, S. E. Borujeni, "Fast Fourier transform speech scrambler," *2002 First International IEEE Symposium Intelligent Systems*, pp. 248–251, September 2002.

[6] J .A. Apolinário Jr., *Criptoanálise de Sinais de Voz Cifrados por Permutação de Segmentos Temporais*, Tese de Mestrado, Universidade de Brasília, Brasilia-DF, Brasil, Junho 1993.

[7] P .P. Vaidyanathan, *Mulitrate Systems and Filter Banks*. Englewood Cliffs NJ, Prentice-Hall, 1993.

[8] L. S. Lee, G .C. Chou, and C .S. Chang, "A new frequency domain speech scrambling system which does not require frame synchronization,"*IEEE Trans. Commun.*, vol. COM-32, no. 4, pp. 444–456, April 1984.

[9] ETSI/TC GSM. Recommendation GSM 05.05 - radio transmisssion and reception. 1.991. L .Levi, F. Muratori, V. Palestine, G. Romano, Performance of DS-CDMA System in a Multipath Fading Environment. Proc. IEEE ICUPC, pp. 28–32, 1.993.

[10] P. Z. Peebles, *Probability, Random Variables and Random Signals Principles*. McGraw-Hill Higher Education, New York , $4^{th}$ Edition, 2000.

[11] 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*, International Telecommunication Union, Geneva,February 2001.

[12] ITU-T Recommendation P.862.3, *Series P: Telephone transmission quality, telephone installations, local line networks. Methods for objective and subjective assessment of quality.* International Telecommunication Union, Geneva, November 2005.

[13] ITU-T Recommendation P.800, *Methods for subjective determination of transmission quality*, International Telecommunication Union, Geneva, August 1996.

[14] ETSI TS 100.909 V8.8.0 (2004-11) Technical Specification 3rd Generation Partnership Project; Digital cellular telecommunications system (Phase 2+); Channel coding.

[15] J. G. Proakis, *Digital Communications*. McGraw-Hill, New York , $4^{th}$ Edition, 2001.