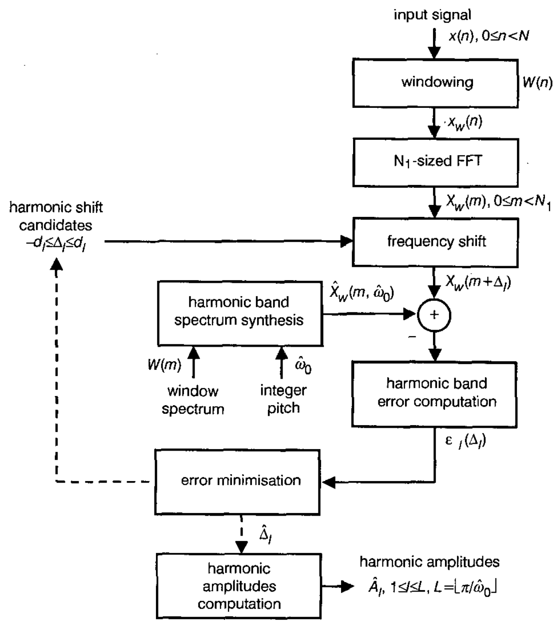

Fig. 2 *Block diagram of DA method*

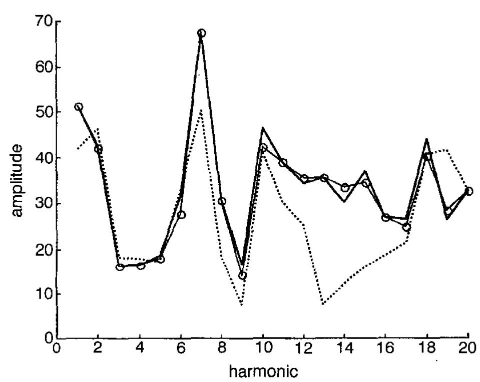

Fig. 3 *Example of harmonic estimations by three methods*

——— fractional pitch method
—○— delta adjustment method
...... integer pitch without adjustment

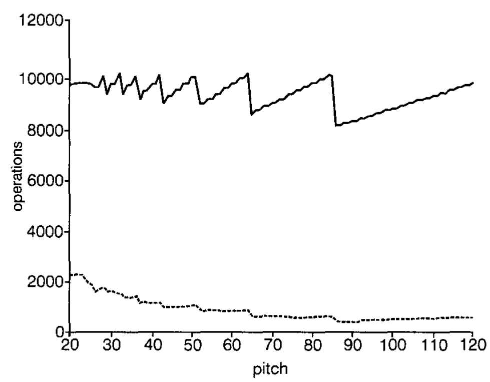

Fig. 4 *Complexities of FP and DA methods against pitch*

——— Y: fractional pitch method
...... X: delta adjustment method

Table 1: Approximated operations of FP and DA

Method	Approximated operations
FP	$Y = M\left(5\left\lfloor\frac{N_1}{P_0}\right\rfloor L + 3\frac{N_1}{2}\right)$
DA	$X = \sum_{l=1}^{L} 5\left\lfloor\frac{N_1}{P_0}\right\rfloor(2\Delta_l + 1)$

To demonstrate performances of the proposed DA method, the spectral distance (*SD*) between the harmonic amplitude estimates by the FP and DA methods and segmental signal-to-noise ratio (*segSNR*) were measured between the synthesised speech signals by the two methods. In calculating *SD* and *segSNR*, only voiced speech frames are taken into consideration. Results of the measures are summarised in Table 2 and show that the DA method is more efficient for female than male speech. Additionally, subjective preference tests were performed informally and any degradation was not perceivable.

Table 2: Results of objective *SD* and *segSNR* tests

	Female	Male	Average
SD, dB	0.39	0.75	0.57
segSNR, dB	35.55	29.53	32.54

Conclusions: We have proposed a fast harmonic estimation for harmonic speech coders. The method matches a frequency-warped version of the original spectrum that conforms to a fixed pitch at all harmonic bands. From the experimental results, the proposed method showed good performance while requiring very low complexity. The method can be easily applied to most harmonic coders that use the conventional FP method.

17 October 2001
Electronics Letters Online No: 20020245
DOI: 10.1049/el:20020245

Yong-Soo Choi (*Network Research and Development Laboratory, LG Electronics Inc., 60-39 Kasan-dong, Kumchon-gu, Seoul 153-023, Korea*)

Dae-Hee Youn (*Department of Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-dong, Sudaemun-ku, Seoul 120-749, Korea*)

References

1 MCAULAY, J., and QUATIERI, T.F.: 'Sinusoidal coding' in KLEIJN, W.B, and PALIWAL, K.K. (Eds.): 'Speech coding and synthesis' (Elsevier, 1995), Chap. 4, pp. 121–173
2 GRIFFIN, F., and LIM, J.S.: 'Multiband excitation vocoder', *IEEE Trans. Acoust. Speech Signal Process.*, 1988, **36**, (8), pp. 1223–1235
3 ISO/IEC FCD 0.1 Subpart 2: 'Information technology—very low bit rate audio-visual coding' 1998
4 YOO, C.D., and LIM, J.S.: 'Speech enhancement based on the generalized dual excitation model with adaptive analysis window'. IEEE Proc. Int. Conf. Acoust. Speech and Signal Process., Detroit, MI, USA, 1995, pp. 832–835
5 KLEIJN, W.B., KROON, P., and NAHUMI, D.: 'The RCELP speech-coding algorithm', *Eur. Trans. Telecommun.*, 1994, **5**, (5), pp. 573–582
6 LEE, K.S., and COX, R.V.: 'TTS based very low bit rate speech coder'. IEEE Proc. Int. Conf. Acoust. Speech and Signal Process., Phoenix, AZ, USA, 1999, pp. 181–184

Unequal error protection and source-channel decoding of CELP speech

A. Nazer and F. Alajaji

The reliable communication of FS CELP 1016 encoded speech over very noisy channels is investigated. Using second-order Markov chains it is shown that over one-quarter of the CELP bits in every frame of speech are redundant. An unequal error protection coding scheme, which exploits this residual redundancy, is proposed for sending the CELP parameters over Gaussian and Rayleigh fading channels. Simulations indicate substantial coding gains over conventional systems.

Introduction: During the last decade, several pieces of work have shown the practical advantages of joint source-channel coding (JSSC) systems, where the source and channel codes are jointly designed, over tandem systems where the two codes are treated separately. We consider a JSSC method for the robust communication of FS CELP 1016 encoded speech [1]. We propose an unequal error protection (UEP) and source-optimised channel coding scheme for the transmission of all the CELP parameters over additive white Gaussian noise (AWGN) and Rayleigh fading channels. This extends the work in [2], where equal error protection (EEP) coding focusing only on the CELP line spectral pair (LSP) coefficients was presented. Our method exploits the residual redundancy within all the CELP parameters (including the LSPs) and employs rate-compatible punctured convolutional (RCPC) codes under maximum *a posteriori* (MAP) decoding. This study can also be applied to other recently developed low rate vocoders such as MELP and others.

CELP 1016 redundancy: CELP 1016 is a frame oriented vocoder that samples the input at 8 kHz with each frame (of 30 ms duration) producing 144 bits at a rate of 4.8 kbit/s [1]. We examine the redundancy in the three most significant bits (MSBs) of each set of CELP parameters: LSPs, pitch gains, pitch delays, codebook gains and indices. Each parameter set is modelled by a second-order block stationary Markov process to capture both inter- and intra-frame redundancies. A large training sequence (83,826 frames) from the TIMIT speech database [3] is applied to the vocoder to estimate the Markov probabilities of each parameter set and compute its entropy rate, its redundancy ρ_D due to its non-uniformity, its redundancy ρ_M due to its memory [2] and its total redundancy ρ_T: $\rho_T = \rho_D + \rho_M$. The results are shown in Table 1 where ρ_D, ρ_M and ρ_T are provided for each parameter and for the entire frame. We see that of the considered 78 high-order CELP bits, more than 21 bits (or $\simeq 27\%$) are redundant.

Table 1: CELP 1016 redundancy (bits/frame)

CELP parameter	Redundancy		
	ρ_D	ρ_M	ρ_T
LSP	5.2747	7.2105	12.4852
Codebook gain	4.0478	1.2544	5.3022
Pitch gain	0.1832	1.4910	1.6742
Pitch delay	0.7064	0.8266	1.5330
Codebook index	0.0323	0.0321	0.0644
Total frame	10.2444	10.8146	21.0590

System structure: In addition to being redundant, the CELP parameters contribute differently to the speech reconstruction. We hence apply UEP to allow various levels of protection for different parameters. Our system consists of a family of RCPC codes [4, 5]. This is achieved by ordering the information by importance, and applying lower rate codes to more important bits and higher rate codes to less important ones. At the receiver, MAP soft-decision decoding that exploits their residual redundancy is employed. This decoder is based on the Viterbi algorithm and chooses the code sequence $\hat{x}^K = (\hat{x}_1, \ldots, \hat{x}_K)$ that minimises $\Pr(y^K | \hat{x}^K) \Pr(\hat{x}^K)$, where $y^K = (y_1, \ldots, y_K)$ is the received sequence of length K, which is the number of sent CELP parameters. We consider binary phase-shift keying modulated AWGN and fully interleaved Rayleigh fading channels with noise variance $N_0/2$. Thus, the above minimisation metric reduces to

$$\sum_{k=1}^{K} \| \boldsymbol{y}_k - \boldsymbol{a}_k \hat{\boldsymbol{x}}_k \|^2 - N_0 \ln \Pr(\hat{x}^K)$$

where $\boldsymbol{a}_K$ denotes the Rayleigh coefficients (for AWGN, $\boldsymbol{a}_k$ is the all-one vector) which we assume to be available at the decode. The prior distribution $\Pr(\hat{x}^K)$ is estimated using the Markov model discussed earlier on the TIMIT training sequence.

We implement three coding systems: uncoded, EEP using a 32-state rate 3/4 convolutional code as in [2] and UEP using a 32-state base rate 1/3 RCPC code with period $p = 8$ [5, 6]. In the EEP and UEP schemes, only 78 bits per frame are convolutionally encoded: the three MSBs of all the 10 LSP parameters, the six MSBs of all the four pitch delay parameters, the three MSBs of all the four codebook gain parameters, and the three MSBs of all the four pitch gain parameters. The remaining 60 bits are sent uncoded and are hard decision decoded for all transmission schemes. The RCPC coding rates, chosen based on the sensitivity of the CELP parameters, are shown in Table 2.

Table 2: UEP scheme using mother rate – 1/3 family of RCPC codes

CELP parameter	Code rate
LSP 1–10	8/24
Pitch delay 1 and 3	8/22
Pitch delay 2 and 4	8/20
Codebook gain 1–4	8/18
Pitch gain 1–4	Uncoded

A TIMIT testing sequence consisting of 4753 frames was used in our simulations which were performed using a practical decoding delay of one frame. The performance criterion used is the average speech distortion measure, which is an average of seven different speech distortion measures of two different types – cepstral and cosh measures [7]. Note that the minimum possible speech distortion (for a noiseless channel) is 4.79 dB.

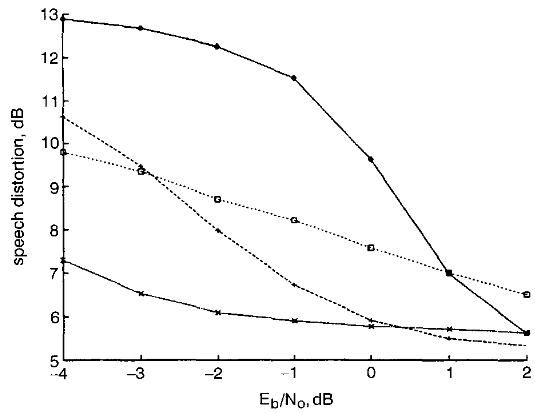

Fig. 1 *Average speech distortion for AWGN channel*

—◇— EEP-ML
- -+- - EEP-MAP
- -□- - uncoded
—×— UEP-MAP

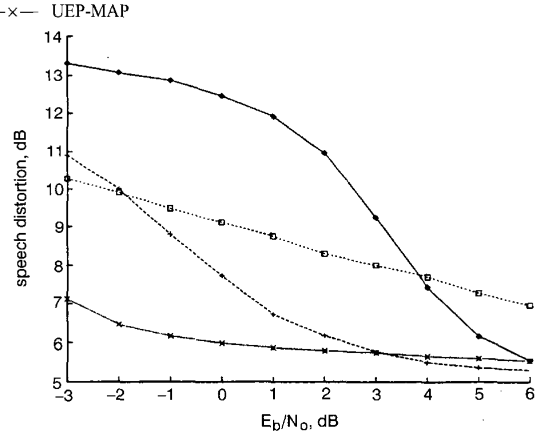

Fig. 2 *Average speech distortion for Rayleigh channel*

—◇— EEP-ML
- -+- - EEP-MAP
- -□- - uncoded
—×— UEP-MAP

The rate of the UEP scheme is 138/252, while that of the EEP scheme of [2] is 138/162. The performance of the two schemes with maximum likelihood (ML) and MAP decoding, as well as that of the uncoded scheme are shown in Figs. 1 and 2 for various E_b/N_0, where E_b is the energy per information bit. We clearly observe that the UEP-

MAP scheme provides the best performance, particularly during severe and medium channel conditions. At a speech distortion of 6 dB, the gains for UEP-MAP against EEP-MAP are 1.42 dB in Fig. 1 and 2.55 dB in Fig. 2; also, the gains for exploiting the CELP redundancy of EEP-MAP over EEP-ML are 1.82 and 2.47 dB in Figs. 1 and 2, respectively. Additional simulations using other performance criteria (such as the symbol error rate) and subjective listening tests are available in [6]. A listening demonstration can be accessed at the Internet site http://markov.mast.queensu.ca/ ~ nazera/.

Acknowledgment: This work was supported in part by NSERC of Canada.

28 August 2001
Electronics Letters Online No: 20020242
DOI: 10.1049/el:20020242

A. Nazer and F. Alajaji (*Mathematics and Engineering, Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, K7L 3N6*)

E-mail: fady@mast queensu.ca

References

1 National Communications System (NCS): 'Details to assist in implementation of Federal Standard 1016 CELP' (Office of the Manager, NCS, Arlington, VA, USA, Tech. Info. Bull. 92-1), 1992
2 ALAJAJI, F., PHAMDO, N., and FUJA, T.: 'Channel codes that exploit the residual redundancy in CELP-encoded speech', *IEEE Trans. Speech Audio Process.*, 1996, **4**, (5), pp. 325–336
3 National Institute of Standards and Technology (NIST): 'The DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM' (NIST, 1990)
4 HAGENAUER, J.: 'Rate-compatible punctured convolutional codes (RCPC codes) and their applications', *IEEE Trans. Commun.*, 1988, **36**, (4), pp. 389–400
5 LEE, L.H.C.: 'New rate-compatible punctured convolutional codes for Viterbi decoding', *IEEE Trans. Commun.*, 1994, **42**, (12), pp. 3073–3079
6 NAZER, A: 'Joint source channel coding and unequal error protection of CELP 1016 encoded speech over very noisy channels'. M.Sc. Thesis, Queen's University, Kingston, ON, Canada, 1997
7 GRAY, A.H., and MARKEL, J.D.: 'Distance measures for speech processing', *IEEE Trans. Acoust., Speech Signal Process.*, 1976, **24**, (8), pp. 380–391

Error-centre-based algorithm for support vector machine training

L. Meng and Q.H. Wu

A new error-centre-based algorithm for support vector machine training is proposed. While the standard techniques require a memory space growing quadratically with the problem size, the new algorithm requires only a linearly growing memory space and may thus be applied to much larger data sets compared with the standard techniques.

Introduction: Training a support vector machine (SVM) amounts to solving a quadratic programming (QP) problem [1]. Given a training sample $\{(\vec{x}_i, y_i)\}_{i=1}^{l}$, where $y_i = \pm 1$ is the target response indicating which pattern the input example $\vec{x}_i$ belongs to, the QP problem for SVM training has the form:

$$\min_{\vec{\alpha}} \quad -\vec{\alpha}^T \vec{1} + \tfrac{1}{2}\vec{\alpha}^T Q \vec{\alpha} \tag{1}$$

$$\text{subject to} \quad \vec{\alpha}^T \vec{y} = 0 \tag{2}$$

$$0 \leq \vec{\alpha} \leq C\vec{1}$$

where Lagrange multipliers $\vec{\alpha}$ are the variables to optimise, and C is a parameter defined to penalise a soft margin. The quadratic term Q in SVM training is defined as $Q_{ij} = y_i y_j K(\vec{x}_i, \vec{x}_j)$, where $K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j)$, $i, j = 1, \ldots, l$, is the kernel function defining the inner-products of the nonlinearly mapped examples $\Phi(\vec{x})$ in the feature space. With the optimal values of $\vec{\alpha}$, α_i^* $(i = 1, \ldots, l)$, the separating hyperplane is given by

$$f(\vec{x}) = \sum_{i=1}^{l} \alpha_i^* y_i K(\vec{x}_i, \vec{x}) + b^* = 0 \tag{3}$$

where b^* is found by enforcing the empirical risk to be zero. The margin is bounded by the set of examples $\{\vec{x} \mid f(\vec{x}) = \pm 1\}$.

The size of the quadratic term Q is l^2, which grows quadratically with the data size of the problem. Although standard QP techniques may be applied directly to SVM training, the fact that they require an explicit storage of Q prohibits their application to problems with large training sets. To tackle this, a new algorithm is proposed for SVM optimisation to reduce the memory space requirement.

Proposed error-centre-based optimisation (ECO) algorithm: Karush-Kuhn-Tucker (KKT) conditions are the necessary and sufficient conditions for the optimal solution of a positive definite QP problems [2]. Applying KKT conditions to SVM optimisation, we have

$$\alpha_i = 0 \Leftrightarrow y_i f(\vec{x}_i) \geq 1$$
$$0 < \alpha_i \leq C \Leftrightarrow y_i f(\vec{x}_i) \leq 1 \tag{4}$$

i.e. only for the examples that lie inside or on the margin are the corresponding α_i^* nonzero. These examples are called 'support vectors' (SVs). In many practical applications, only a small percentage of training examples are SVs. This feature is referred to as the 'sparseness' of the solution to SVM optimisation.

Note that the value of the objective function (1) remains unchanged if the rows and columns of matrix Q corresponding to zero Lagrange multipliers are removed. This means that the set of SVs is all we need to fully define the QP problem and ensure the correct solution. Moreover, the sparseness of the solution implies that both the memory space requirement and the computation time may be largely reduced if we train the machine on the set of SVs. However, it is impossible to achieve this straightforwardly since we do not know beforehand which training examples would turn out to be SVs. The proposed algorithm takes a heuristic approach to find the set of SVs. The implementation steps are listed below:

Given training set **S**, treat each pattern of **S** as a cluster and add the centres of these clusters to the working set $\hat{\mathbf{S}}$.
Repeat
    Train the SVM on $\hat{\mathbf{S}}$.
    Set $\hat{\mathbf{S}}$ to the support vectors.
    For each cluster of **S**
        Split the current cluster into two sub-clusters by identifying the margin errors.
        Add the cluster centre of the margin errors to $\hat{\mathbf{S}}$.
Until no new error cluster is formed.

KKT conditions (4) imply that no training example for which $\alpha_i = 0$ may lie inside the margin. Inspired by this, the splitting of a cluster is achieved by the steps shown below:

For each example $\vec{x}_i$ in the current cluster $\mathbf{C}_p$
    set $\alpha_i = 0$.
    If $y_i f(\vec{x}_i) \leq 1$, assign $\vec{x}_i$ to the new cluster $\mathbf{C}_q$.
Cluster centre $\vec{c}_s = (\sum_{\vec{x}_j \in \mathbf{C}_s} \vec{x}_j) / (\sum_{\vec{x}_j \in \mathbf{C}_s} 1)$, $\forall s,\ s \in \{p, q\}$

The examples for which $y_i f(\vec{x}_i) \leq 1$ violate KKT conditions, and are thus referred to as margin errors in ECO. KKT conditions are the necessary and sufficient conditions for the optimal solution and the ECO procedure iterates until no margin error is found. Hence, the optimality of the solution, with respect to the original training set **S**, is guaranteed by the new algorithm. To further reduce the problem size at each iteration, only the centres of the margin error clusters are added to the working set of training. The rest of the clusters are represented by the SVs of the previous iteration. Since only are the centres of margin errors involved in the training (SVs of previous iterations must have been error centres), this new algorithm is called error-centre-based optimisation (ECO).

Experiments and results: The ECO algorithm is implemented in Matlab. The quadratic programming subroutine provided in the Matlab optimisation tool-box has been used as the standard technique for comparison purposes. The QP problem at each iteration of ECO is