tion in the channel, which in turn deteriorates the current driving capability of device.

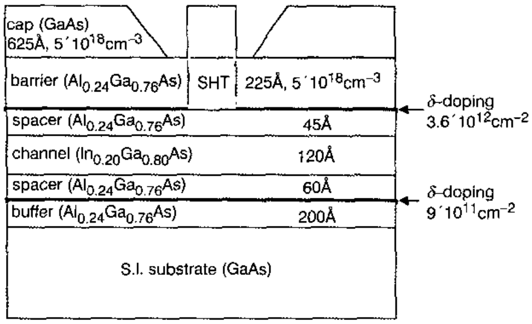

Fig. 1 *Cross-section of $Al_{0.24}Ga_{0.76}As/In_{0.2}Ga_{0.8}As$ double heterostructure enhancement/depletion-mode p-HEMT*

Cross-hatched region under gate indicates deactivated donors after SHT process to fabricate E-HEMT

Table 1: Effect of RF power of hydrogen plasma on threshold voltage V_{TH}, gate-grounded drain current I_{DSS}, cutoff frequency f_T, maximum oscillation frequency f_{max} of p-HEMT

RF [W]	V_{TH} [V]	I_{DSS} [mA/mm]	f_T [GHz]	f_{max} [GHz]
0	−0.77	123	7.5	23.36
30	−0.05	29	9	23.36
100	0.26	0.9	8.28	22.8

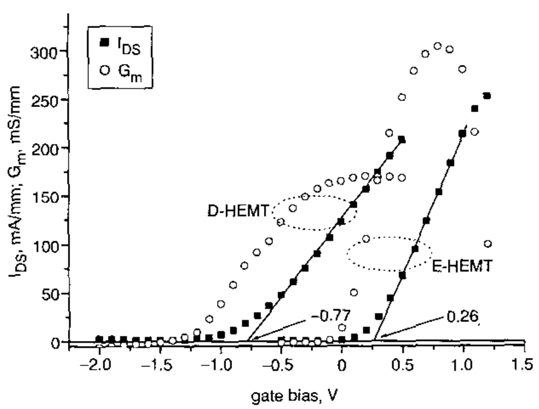

Fig. 2 *Drain current I_{DS} and transconductance G_m of E/D-HEMTs measured at $V_{DS}=2.5$ V*

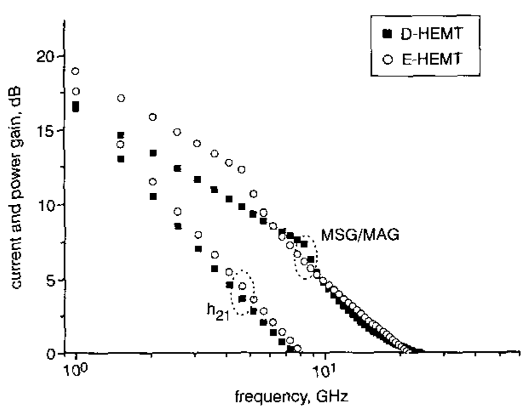

Fig. 3 *RF performances of E/D-HEMTs measured at gate bias displaying maximum transconductance and $V_{DS}=2.5$ V*

Conclusions: We have demonstrated E/D-mode HEMTs using selective hydrogen treatment composed of hydrogen exposure and RTA, prior to the gate metallisation. The selective hydrogen treatment method produced a large threshold shift without serious degradation in DC and RF performance of the p-HEMT. These results are attractive in various applications including fibre optic communication systems and power amplifiers requiring E/D-mode HEMTs.

Acknowledgments: This work was supported in part by MOE through the BK21 program and KOSEF through the UFON-ERC at the Kwangju Institute of Science and Technology.

3 December 2002
Electronics Letters Online No: 20030263
DOI: 10.1049/el:20030263

I.-H. Kang and J.-I. Song (*Department of Information and Communications, Kwang-Ju Institute of Science and Technology, 1 Oryong-dong, Buk-gu, Kwangju 500-712, Korea*)

E-mail: jisong@kjist.ac.kr

References

1 TKACHENKO, Y., KLIMASHOV, A., WEI, C., ZHAO, Y., and BARTLE, D.: 'E-PHEMT for single supply, no drain switch, and high efficiency cellular telephone power amplifiers'. GaAs IC Symp. 1999, Monterey, CA, USA, 1999, pp. 127–130

2 CHEUNG, R., PATRICK, W., and BAECHTOLD, W.: 'Depletion and enhancement-mode InP high electron mobility transistors fabricated by a dry gate recess process'. Solid-State and Integrated Circuit Technology Conf. 1998, Beijing, China, pp. 598–601

3 MAHAJAN, A., CUEVA, G., ARAFA, M., FAY, P., and ADESIDA, I.: 'Fabrication and characterization of an InAlAs/InGaAs/InP ring oscillator using integrated enhancement- and depletion-mode high electron mobility transistors', *IEEE Electron Device Lett.*, 1997, **18**, (8), pp. 391–393

4 KANG, I.H., KIM, J.H., SONG, H.J., and SONG, J.I.: 'Effect of selective hydrogen pretreatment on characteristics of AlGaAs/InGaAs p-HEMTs'. 9th Korean Conf. on Semiconductors, Chunan, Korea, 2002pp. 265–266

5 OKUMURA, T.: 'H-related issues in GaAs Schottky contacts'. GaAs Mantech, Vancouver, Canada, 1999

6 PEREIRA, R.G., VAN HOVE, M., DE POTTER, M., and VAN ROSSUM, M.: 'Influence of CH_4/H_2 reactive ion etching on the deep levels of Si-doped $Al_xGa_{1-x}As$ ($x=0.25$)', *J. Vac. Sci. Technol. B*, 1996, **14**, (3), pp. 1773–1779

Compensating additive noise and CS-CELP distortion in speech recognition using stochastic weighted Viterbi algorithm

N.B. Yoma, J. Silva, C. Busso and I. Brito

A solution to the problem of speech recognition with signals corrupted by additive noise and CS-CELP coders is presented. The additive noise and the coding distortion are cancelled according to the following scheme: first, the pdf of the clean coded-decoded speech is estimated with an additive noise model; secondly, the pdf of the clean uncoded signal is also estimated with a coding distortion model; finally, the hidden Markov model is compensated using the expected value of observation pdf in the context of the stochastic weighted Viterbi algorithm.

Introduction: Improving robustness to noise is one of the most important problems that need to be solved to make speech recognition successful in real applications. Moreover, the evolution and popularity of mobile and TCP/IP networks have created the problem of improving the recognition accuracy for speech distorted by low-bit rate coders. The distortion of coding schemes in speech recognisers cannot be solved by applying conventional noise cancelling techniques [1]. Some of the techniques that have been proposed to cancel or compensate additive or/and convolutional noise are: spectral subtraction (SS) [2]; RASTA [3]; parallel model combination (PMC) [4]; and, cepstral mean normalisation. A stochastic version of the weighted Viterbi [5] (stochastic weighted Viterbi (SWV)) algorithm was proposed and successfully applied to a text-dependent speaker verification task with signals corrupted by additive noise [6]. In that paper the observation parameters were considered as being random variables with normal distributions, and the hidden Markov model (HMM) observation pdf was replaced with its expected value. In this Letter the coding–decoding distortion is modelled with a Gaussian distribution. An HMM compensation method is then proposed by considering the original and unseen uncoded cepstral parameter also as a random variable and by estimating the expected value of the output pdf as in [6]. Finally, the additive noise and the coding distortion are cancelled according to the following scheme: the pdf of the clean coded–decoded speech is estimated; the pdf of the clean uncoded signal is then also estimated; and, the HMM is compensated using the expected value of observation pdf. The approach leads to reductions as high as 40 or 50% in word error rate (WER).

Stochastic weighted Viterbi (SWV) algorithm: In the ordinary HMM topology the output pdf of observing the frame O_t at state s, $b_s(O_t)$, is computed considering O_t as being a vector of constants. In this Letter the observation vector is composed of static, delta and delta-delta cepstral coefficients, and according to [6] these parameters should be considered as being random variables with normal distributions. Therefore, to counteract this incompatibility, [6] proposes to replace $b_s(O_t)$ with $E[b_s(O_t)]$ in the Viterbi algorithm, where $E[b_s(O_t)]$ denotes the expected value of the output pdf. The expected value and the uncertainty in noise cancelling variance of the static and delta cepstral coefficients are estimated according to [6]. The same approach employed with the delta cepstral parameters is applied to the delta-delta cepstral coefficients in frame t defined as the difference between the delta parameters in $t+1$ and $t-1$. If the HMM output pdf, $b_s(O_t)$, is modelled with a mixture of Gaussians with diagonal covariance matrices, $E[b_s(O_t)]$ is [6]:

$$E[b_s(O_t)] = \sum_{g=1}^{G} p_g \cdot \prod_{n=1}^{N} \frac{1}{\sqrt{2 \cdot \pi \cdot Vtot_{s,g,n,t}}} e^{-1/2 \cdot (E[O_{t,n}]-E_{s,g,n})^2 / Vtot_{s,g,n,t}} \quad (1)$$

where s, g, n are the indices for the states, the Gaussian components and the coefficients, respectively; p_g is a weighting parameter; $O_t = [O_{t,1}, O_{t,2}, \ldots, O_{t,N}]$ is the observation vector; $E_{s,g,n}$ and $Var_{s,g,n}$ are the HMM mean and variance, respectively; the mean, $E(O_{t,n})$, and the uncertainty in noise cancelling variance, $Var(O_{t,n})$, are estimated as described above; and

$$Vtot_{s,g,n,t} = Var_{s,g,n} + K \cdot Var(O_{t,n}) \quad (2)$$

K is a constant that adapts the decreasing rate of the output pdf discrimination ability to the language model.

Compensation of CS-CELP coding distortion: The coding–decoding distortion, D_n, is modelled in the cepstral domain as an additive random variable with Gaussian distribution $f_{D_n}(D_n) = N(m_n^d, v_n^d)$ where n, m_n^d and v_n^d denote the cepstral coefficient, the mean and the variance, respectively. Therefore, the cepstral coefficient n in frame t of the original signal, $O^o_{t,n}$, is given by:

$$O^o_{t,n} = O^d_{t,n} + D_n \quad (3)$$

where $O^d_{t,n}$ is the cepstral coefficient corresponding to the decoded–coded speech signal. In a real application $O^d_{t,n}$ is the observed cepstral parameter. From (3), the expected value of $O^o_{t,n}$, is given by:

$$E[O^o_{t,n}] = O^d_{t,n} + m_n^d \quad (4)$$

Consequently, the coding–decoding distortion scheme is represented by the mean vector $M^d = [m_1^d, m_2^d, \ldots, m_n^d \ldots, m_N^d]$ and the variance vector $V^d = [v_1^d, v_2^d, \ldots, v_n^d \ldots, v_N^d]$. This distortion does not depend on the phonetic class, and is consistent with the analysis presented in [1]. The original and unseen uncoded cepstral parameter $O^o_{t,n}$ is a random variable, so the output pdf $b_s(O_t)$ should also be replaced with its expected value $E[b_s(O_t)]$ as in [6]:

$$E[b_s(O_t)] = \sum_{g=1}^{G} p_g \cdot \prod_{n=1}^{N} \frac{1}{\sqrt{2 \cdot \pi \cdot Vtot^d_{s,g,n}}} \cdot e^{-1/2 \cdot (E[O^o_{t,n}]-E_{s,g,n})^2 / Vtot^d_{s,g,n}} \quad (5)$$

where $E[O^o_{t,n}]$ is given by (4) and

$$Vtot^d_{s,g,n} = Var_{s,g,n} + v_n^d \quad (6)$$

Joint compensation of additive noise and coding distortion: As can be seen in Fig. 1, the problem presented here corresponds to a clean signal $s(t)$ firstly corrupted by additive noise in the temporal domain, $x(t)$, and then coded and decoded, $x^D(t)$. The observation parameter vectors of the signals $s(t)$, $x(t)$ and $x^D(t)$ are $O_t^{S,U}$, $O_t^{X,U}$ and $O_t^{X,D}$, respectively. S and X denote the clean and noisy signal, respectively; U and D correspond to the signals before (uncoded) and after (distorted) the coding–decoding process. As shown in Fig. 2, the proposed method first compensates the additive noise by applying SS and estimating the variance in noise cancelling with $x^D(t)$. As a result, the pdf of the distorted clean speech, $f_{O^{S,D}}(O^{S,D})$, is generated. The compensation in (5)(6), CDC (coding-distortion compensation), is then applied generating the pdf of the clean uncoded signal, $f_{O^{S,U}}(O^{S,U})$. Finally, by taking the expected value of the output pdf, the compensation of the additive noise and the coding distortion will be incorporated in the Viterbi decoding.

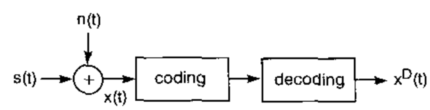

Fig. 1 *Additive noise and coding distortion*

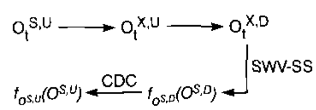

Fig. 2 *Joint compensation of additive noise and coding distortion*
$f_{O^{S,D}}(O^{S,D})$ denotes pdf of clean distorted signal; $f_{O^{S,U}}(O^{S,U})$ corresponds to pdf of clean uncoded signal

Table 1: WER (%) with signal corrupted with additive noise (car noise) and coded by 8 kbit/s CS-CELP

SNR	18 dB	12 dB
Baseline	107.0	147.4
SWV-SS	61.4	93.2
SWV-SS-CDC	53.0	89.0

Table 2: WER (%) with signal corrupted with additive noise (speech noise) and coded by 8 kbit/s CS-CELP

SNR	18 dB	12 dB
Baseline	106.4	147.0
SWV-SS	57.1	90.4
SWV-SS-CDC	52.0	87.0

Experiments: The compensation method was tested with speaker-independent continuous speech recognition experiments using LATINO-40 database [7]. This database is composed of 40 Latin American native speakers, each reading 125 sentences from newspapers in Spanish. The training utterances were 4500 uncoded clean sentences provided by 36 speakers and context-dependent phoneme HMMs were employed. The vocabulary is composed of almost 6000 words. The testing database was composed of 500 utterances provided by four testing speakers (two females and two males). Each context-dependent phoneme was modelled with a three-state left-to-right topology without skip transition, with eight multivariate Gaussian densities per state and diagonal covariance matrices. Bigram language modelling was employed, and K in (2) was made equal to 10. The band from 300 to 3400 Hz was covered with 14 Mel DFT filters, at the output of each channel the energy was computed, SS was applied and the log of the energy was estimated. The frame energy plus ten static cepstral coefficients, and their first and second time derivatives were computed. The noise was estimated using only 10 non-speech frames before the beginning of the utterance. The 500 testing clean signals were used to create the noisy utterances by adding car and speech noise from the Noisex database. The noisy signals were then coded and decoded using the 8 kbit/s CS-CELP (ITU-T G.729). The coding–decoding distortion parameters M^d and V^d were estimated by directly aligning uncoded and coded–decoded training utterances. The techniques are indicated as follows: *Baseline* without any HMM compensation; *SWV-SS*, the *SWV* algorithm with *SS*; and, *SWV-SS-CDC*, the SWV algorithm with both SS and CDC. The results are shown in Tables 1 and 2. The baseline system with clean signal gave a WER equal to 38.9%.

Discussion and conclusion: As can be seen in Tables 1 and 2, the additive noise and the coder dramatically degrade the word accuracy of the system at SNR = 18 and 12 dB. SWV and SS substantially reduced the WER, but the highest improvement was achieved when CDC was also applied. Reductions as high as 50 and 40% in WER are observed at 18 and 12 dB. Nevertheless, the degradation of the system at 12 dB is still too severe. The additive noise has probably a more significant effect on the increase of the WER. As a result, improving the accuracy of the additive noise model [5] at low SNR should certainly increase the effectiveness of the approach proposed here. Finally, the coding distortion compensation should be applicable to other speech compression schemes.

Acknowledgments: The authors wish to thank F. McInnes, University of Edinburgh, UK, for proofreading our manuscript. The research described here was supported by Conicyt/Fondecyt-Chile.

25 November 2002
Electronics Letters Online No: 20030252
DOI: 10.1049/el:20030252

N.B. Yoma, J. Silva, C. Busso and I. Brito (*Electrical Engineering Department, University of Chile, Av. Tupper 2007, P.O. Box 412-3, Santiago, Chile*)

E-mail: nbecerra@cec.uchile.cl

References

1 HUERTA, J.M.: 'Speech recognition in mobile environments'. PhD Thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, April 2000
2 VASEGHI, S.V., and MILNER, B.P.: 'Noise compensation methods for hidden Markov model speech recognition in adverse environments', *IEEE Trans. Speech Audio Process.*, 1997, **5**, (1), pp. 11–21
3 HERMANSKY, H., *et al.*: 'Compensation for the effect of the communication channel in auditory-like analysis of speech (RASTA-PLP)'. Proc. Eurospeech91, Geneva, Italy, 1991, pp. 1367–1370
4 GALES, M.J.F., and YOUNG, S.J.: 'HMM recognition in noise using parallel model combination'. Proc. Eurospeech93, Berlin, Germany, 1993, pp. 837–840
5 YOMA, N.B., *et al.*: 'Improving performance of spectral subtraction in speech recognition using a model for additive noise', *IEEE Trans. Speech Audio Process.*, 1998, **6**, (6), pp. 579–582
6 YOMA, N.B., and VILLAR, M.: 'Speaker verification in noise using a stochastic version of the weighted Viterbi algorithm', *IEEE Trans. Speech Audio Process.*, 2002, **10**, (3), pp. 158–166
7 LDC. Latino-40 database provided by Linguistic Data Consortium (LDC), University of Pennsylvania

Explicit multicast over mobile IP with routing optimisation

Sung-Hee Kim and Ki-Jun Han

A new scheme is proposed, called explicit multicast over mobile IP with routing optimisation (XMIPRO), to solve the problem due to triangle routing and serious packet losses during the registration delay in explicit multicast over mobile IP (XMIP). Analytical and simulation results show that the scheme offers more efficient routing and a greater robustness than the existing schemes in mobile IPv4 networks.

Introduction: Recently, explicit multicast over mobile IP (XMIP) [1] was proposed to support the mobile multicast using explicit multicast (Xcast) [2] in mobile IP networks. When the mobile node is away from its home, however, Xcast packets are first routed to the home agent. XMIP thus suffers from inefficient delivery due to triangle routing and serious packet losses during the registration delay. To solve this problem, we propose a new mobile multicasting scheme called explicit multicast over mobile IP with routing optimisation (XMIPRO).

XMIPRO: The XMIPRO is based on two-level, hierarchical mobility management architecture as shown in Fig. 1. Xcast router (XR) located at a higher level in the network hierarchy is the Xcast-capable router and temporarily stores the binding information for the mobile node. When the mobile node roams to other networks, it performs a general registration procedure, as in mobile IP. On receipt of a *registration reply* message from home agent (HA), foreign agent (FA) sends a *redirection* message to its XR as indicated by (1) (see below). Upon receiving the *redirection* message, the XR creates a binding cache, temporarily storing the mapping information that indicates the mobile node a new point of attachment and then forwards the *redirection request* message to the correspondent node's mobility agent (CMA) as indicated by (2). If the XR has a binding cache entry for the mobile node's new point of attachment when receiving an Xcast packet during the *redirection* delay, it can send the Xcast packet directly to mobile node's new care-of address (COA). mobility agent (MA) is the Xcast-capable mobility agent which manages an individual subnet and temporarily stores the Xcast packet received from correspondent node (CN). The routing table maintained by MA is extended to directly send the Xcast packet, as shown in Fig. 1. The Xcast binding cache temporarily stores the binding information to indicate the mobile node's new point of attachment. Upon receiving the *redirection request* message from mobile node's new FA, correspondent node's mobility agent (CMA) creates an Xcast binding cache entry for the mobile node. Since all packets transmitted by mobile nodes are routed towards the destination via MA, if CMA has an Xcast binding cache entry for the destination node when sending an Xcast packet to the mobile node, it sends the Xcast packet directly to the COA as indicated in the cached mobility binding. XMIPRO is also free from packet loss due to roaming. Upon receiving the Xcast packet, mobility agent creates a cache (buffer) and temporarily stores the Xcast packet received from correspondent node (CN). Therefore, upon receiving the new Xcast packet in a visited subnet, the mobile checks whether there is any offset in the datagram sequence number of packets. If the Xcast packets received in a newly visited subnet are ahead (in terms of sequence number) of those in the old subnet due to network dynamic, the mobile sends the *clear* messages with an offset block, say $[a, b]$, to the old FA, where the offset block means the difference in data sequence to be recovered between two adjacent subnets. Upon receiving the *clear* message, the old FA first looks up the MN binding cache bounded with mobile node's home address. If the old FA has a cache entry, it tunnels the Xcast packet in the cache to the mobile. In this way, XMIPRO can solve the triangle routing problem and guarantees no packet loss due to roaming.

Cost analysis: We built a simplified analytical model for evaluation of the average costs using the method presented in [3]. We assumed that a CN generates Xcast packets at a mean rate of λ and mobile nodes move from one subnet to another at a mean rate of μ. The packet to mobility ratio (PMR) is defined as the mean number of Xcast packets received by a mobile node from a CN per movement. The PMR is denoted by $p = \lambda/\mu$. We introduced two cost components. One was the transmission cost, measured by the number of hops, the other was the processing cost, measured in processing time of the control packets. The distances between the entities involved in our proposed scheme are denoted by a, b, c, d, e and f as shown in Fig. 1. We defined some parameters to derive the network cost as shown below:

r: average cost for processing a control packet at any node;
L: ratio of the length of a data packet to the length of an ICMP control packet;
C_{reg}: cost of mobile IP registration at the new FA;