in the first part of the preceding Section. Again we neglect the temperature dependence of the parasitic elements of the device.

$$\frac{G_C(T_1)}{G_C(T_2)} = \left(\frac{f_T(T_1)}{f_T(T_2)}\right)^2 \qquad (6)$$

Concerning the measurement of the mixer, using a measurement of test devices with eqn. 6, a drop in conversion gain of 1.7dB was predicted due to the change in transconductance. Unfortunately, this is not the only significant effect contributing to mixer performance degradation. Another important factor that has to be considered is the drift of pinch-off voltage with temperature.

Since the mixer relies on being biased very close to the pinch-off point, conversion gain is very sensitive to pinch-off voltage variations. The drift in pinch-off voltage was measured to be –0.71mV/°C for the 6 × 15μm device used for the mixer MMIC.

By making a measurement of conversion gain against gate bias point, we estimated the drop in conversion gain resulting for the gate being biased 99mV below optimum bias point at 140°C. The resulting increase in conversion loss above that predicted by eqn. 6 was found to be 0.7dB.

At room temperature the mixer exhibits ~1dB conversion loss with an LO power level of 7dBm and 3dB conversion gain with 13dBm LO power (*RF* = 65.5GHz, *LO* = 64GHz). This compares well with other published room temperature results, e.g. [6]. The behaviour of the mixer over temperature can be seen in Fig. 1 for the operating point with 7dBm LO power. A second measurement with 13dBm LO power showed a similar drop. Over the temperature range from 0°C to 140°C, the conversion gain drops by 2.9dB, while the predicted value was 2.4dB.

Conclusion: We have developed a simple way for modelling the temperature dependence of millimetre-wave front end MMICs and presented a method by which to predict the drop of gain using only the change in cut-off frequency thereby obviating the need for full temperature dependent device models. This showed that the temperature dependence of amplifiers becomes more acute as the number of stages increases. Furthermore, we found that mixer MMICs will be much more severely affected by temperature change than amplifier circuits owing to conversion gain sensitivity to drift in pinch-off voltage with temperature.

Acknowledgments: The authors wish to thank IRTU for funding this project under grant ST173, and OMMIC (PML) for fabricating the GaAs MMIC with their D01PH process.

© IEE 2000 *1 August 2000*
Electronics Letters Online No: 20001353
DOI: 10.1049/el:20001353

T. Brabetz, N.B. Buchanan and V.F. Fusco (*The Queen's University of Belfast, Electrical and Electronic Engineering, Ashby Building, Stranmillis Road, Belfast BT9 5AH, United Kingdom*)

E-mail: V.Fusco@ee.qub.ac.uk

References

1 ALI, F., and GUPTA, A.: 'HEMTs & HBTs: devices, fabrication, and circuits' (Artech House, Dedham, MA, 1991), pp. 292–294

2 ROBERTS, G.W., and SEDRA, A.S.: 'SPICE' (Oxford University Press, 1997), 2nd edn., p. 126

3 ANHOLT, R.: 'Electrical and thermal characterization of MESFETs, HEMTs, and HBTs' (Artech House, Norwood, MA, 1995)

4 SOARES, R., GRAFFEUIL, J., and OBREGON, J. (Eds.): 'Applications of GaAs MESFETs' (Artech House, Dedham, MA, 1983), pp. 95–100

5 MAAS, S.A.: 'Microwave mixers' (Artech House, Norwood, MA, 1993), 2nd edn.

6 MADIHIAN, M., DESCLOS, L., MARUHASHI, K., ONDA, K., and KUZUHARA, M.: '60-GHz monolithic down- and up-converters utilizing a source-injection concept', *IEEE Trans. Microw. Theory Tech.*, 1998, **46**, (7), pp. 1003–1006

Independent component analysis applied to feature extraction for robust automatic speech recognition

L. Potamitis, N. Fakotakis and G. Kokkinakis

The authors explore independent component analysis (ICA) as a statistical technique for deriving suitable data-driven representational bases for the projection of spectra and cepstra in the context of automatic speech recognition (ASR). Based on the close link between the independent mechanisms of speech variability and the concept of statistical independence they derive a new feature transformation that effects consistent improvement in recognition performance.

Introduction: The feature extraction stage of current ASR systems converts the input speech waveform in a series of low-dimensional vectors, each summarising a short segment of the acoustical speech input to minimise the computational demands of the hidden Markov model (HMM) classifier. The resulting feature vector produced by subsequent transformations is driven in a final decorrelation stage that permits the use of diagonal covariance mixture density HMMs.

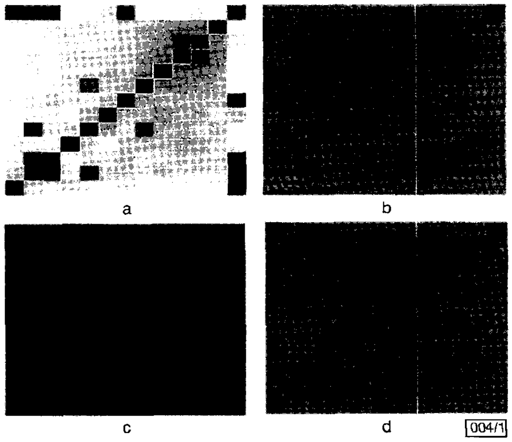

Fig. 1 *Covariance of DCT, PCA, LDA and ICA projected spectra*
a DCT
b PCA
c LDA
d ICA

Several successful decorrelation strategies have been applied to a log-spectrum and cepstrum [1]. However, there are certain shortfalls inherent in these strategies. The discrete cosine transform (DCT) is a nonadaptive procedure that projects a log-spectrum in the direction of global variance and achieves only partial decorrelation of features (Fig. 1*a*). Linear discrimination analysis (LDA) is reported to provide insufficient degrees of freedom for discrimination between classes [2] and is very sensitive to SNR mismatches in training and test data [3] (Fig. 1*c*). Principal component analysis (PCA), based on the principle of minimum reconstruction error, projects a log-spectrum in the directions of maximum variability (Fig. 1*b*). Unfortunately, there is no guarantee that the sources of variability explained by PCA are all useful in speech recognition. Last, but not least, the minimum reconstruction error does not imply minimum classification error.

Our attempt, based on ICA, is a departure from (rather than an extension of) the second-order statistics of PCA, DCT and LDA, which cannot handle the high-order, nonlinear dependencies between the variables of the feature vectors. The training observational data **X** (log-spectrum/cepstral features) can be seen as a multivariate time series resulting from an unknown hidden linear mixing process **A** of independent functions **S**; that is, **X** = **A*S**. ICA finds a demixing matrix **W** such that **U** = **W*X** = **W*A*S** = **P*D*S**, where **P** is a permutation matrix and **D** is a diagonal scaling matrix. The matrix **W** is formed using an unsupervised iterative algorithm based on the principle of optimal information transfer of observational data through a nonlinear squashing function ϕ [4]. The algorithm maximises the mutual information between input (**X**) and output (**U**). Minimum mutual information as a cost function is a measure that is described by all the higher order cross-statistics. Optimising it results in a complete cross-statistics optimisation. In [4] it is shown that an updating algorithm to obtain **W** is

$$\Delta W \propto \frac{\partial I(u,x)}{\partial W} \propto \frac{\partial I(u,x)}{\partial W} W^T W = \left[I - \phi(u)u^T\right] W \tag{1}$$

where

$$\phi(u) = -\frac{\frac{\partial p(u)}{\partial u}}{p(u)} = \left[-\frac{\frac{\partial p(u_1)}{\partial u_1}}{p(u_1)}, \dots, \frac{\frac{\partial p(u_N)}{\partial u_N}}{p(u_N)}\right]^T \tag{2}$$

p(*u*) stands for the nonlinear density function. We tried several nonlinearities with comparable performance. The logistic function was used so that $\phi(u) = 2\tanh(u)$.

The assumptions of ICA conform to the framework of homomorphic analysis [5]. Viewing the speech spectrum $|S(w)|$ as consisting of a quickly varying part $|E(\omega)|$ and a slowly varying part $|\Theta(\omega)|$ we can form the following equation:

$$\log|S(\omega_i)| = \log|E(\omega_i)| + \log|\theta(\omega_i)| \qquad i = 1, \dots, \text{number of filterbanks} \tag{3}$$

We assume that the slowly and fast-varying parts pertain to filter functions of the glottal pulse, vocal cord transfer function, mouth radiation and transmission line distortion, all of which can be added in the log-spectrum domain. Since these filters correspond to different mechanisms, the log-spectrum can be considered the result of an unknown linear mixing process of independent sources of variation, which is exactly the assumption that the ICA framework postulates. Our aim is to identify and exclude from the projection matrix **W** useless variability sources, thus achieving better feature discrimination.

Table 1: Word recognition accuracy (%)

MFCC	56.57
MFCC-CMN*	70.68
ICA from MFCC-CMN	74.21
ICA from log FBANK	**76.22**

* CMN = cepstral mean normalisation,
FBANKS = mean normalised mel-filterbank features

Simulation and results: To assess the effectiveness of the ICA derived projection functions, we trained Entropics' HMM Toolkit using part of the identity card corpus of the SpeechDat database. One thousand speech files formed the training set and 200 the testing set. We designed two sets of experiments: one with 20 mel scale log filterbank coefficients and one with 20 mel frequency cepstral coefficients (MFCCs). For both experiments the ICA procedure resulted in the corresponding 20 × 20 weight matrix **W** derived solely from the training set. We retained the 13 most dominant projection vectors according to their absolute norm, which reduced **W** to 13 × 20. We projected each training and testing feature file by applying the appropriate **W** matrix. We did not employ first and second order derivatives in the comparison of the new feature set as we were not interested in absolute performance.

Table 1 depicts the recognition results achieved with the different feature sets. The fourth result demonstrates a clear and considerable gain by decorrelating all moments beyond the second and brings out the insufficiency of the DCT (third result). Since HMMs are based on Gaussian models which have full statistics of the second order, we can attribute the fourth result to the more effective decorrelation properties of ICA (see the figures of the projected covariance matrices) and to the more meaningful selection of features.

Conclusions: The concept of statistical independence gives a better insight into the feature selection process due to the homomorphic property of a log-spectrum. We wish to emphasise the practical advantage of our method, which furthers considerably the discriminative ability of the MFCC and log-spectrum front-end by applying an additional transformation matrix without inflicting significant computational overhead, since **W** is derived off-line.

© IEE 2000 29 August 2000
Electronics Letters Online No: 20001365
DOI: 10.1049/el:20001365

I. Potamitis, N. Fakotakis and G. Kokkinakis (*Wire Communications Laboratory, Electrical and Computer Engineering Department, University of Patras, 261 10 Rion, Patras, Greece*)

E-mail: potamitis@wcl.ee.upatras.gr

References

1 EISELE, T., HAEB-UMBACH, , and LANGMANN, D.: 'A comparative study of linear feature transformation techniques for automatic speech recognition'. ICSLP, 1996, pp. 252–255

2 HERMASNKY, H., and NABYATH, N.: 'Spectral basis functions from discrimination analysis'. ICSLP, 1988, pp. 616–620

3 SIOHAN, O.: 'On the robustness of linear discrimination analysis as a preprocessing step for noisy speech recognition'. ICASSP, 1998, pp. 125–128

4 LEE, T.W., GIROLAMI, M., and SEJNOWSKI, T.J.: 'Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources', *Neural Comput.*, 1999, **11**, (2), pp. 409–433

5 JANG, G.-J., YUN, S.-J., and OH, Y.-H.: 'Feature vector transformation using ICA and its application to speaker verification', *Eurospeech*, 1999, pp. 767–770

Adaptive multi-rate speech coder for VoIP transmission

V. Abreu-Sernández and C. García-Mateo

A speech coder with three different bit rates adapted for the transmission of voice-over IP (VoIP) networks is presented. Continuously, a rate control device analyses the traffic congestion of the network and orders the speech coder to switch among five operation modes if necessary. These modes include mitigation techniques of packet losses.

Introduction: VoIP is an emergent area for real-time speech transmission, not only for phone calls but also for documents teleconferencing, for videoconferencing or for calling and helping Internet centres using a click-to-talk button on WWW browser screens. In this way, audio and video signals, data, images and text are running on the same network, allowing the use of multimedia services, but IP networks were designed to convey information without real-time needs. Therefore, when the traffic congestion of the network causes high values of latency (defined as the total end-to-end delay in a packet transmission) or causes the loss of a packet with valid information, the communication is not turned off. In such cases, when the acknowledgment message of a packet has not arrived to the transmitter, the packet is retransmitted by means of an automatic repeat request (ARQ) mechanism. However, speech conversations are very sensitive to delays: the RFC2354 [1] document defines an interactive speech session when the latency is lower than 250ms. For this reason, in a VoIP communication when a packet is thrown away in any router, there is no possibility of a retransmission and the speech information carried by this packet is lost. Until recently there have been no quality of service (QoS) functions running over IP networks. In 1997, the Internet Engineering Task Force (IETF) standarised the resource reservation protocol (RSVP) [2] for providing the QoS requested for every user. However for the right behaviour of the RSVP, most of the everywhere routers must be replaced by a new generation of devices, and this expensive change will come gradually.

There are more simple solutions to mitigate the damaging effects of the packet losses from the speech coder/decoder point of view. First of all, the speech coder must include a voice activity detector (VAD) which reduces the transmitted information by ~50–60%. Secondly, the speech coder must switch in a faster way between different bit rates depending on the statistics of the present transmission. Thirdly, the coder/decoder must use concealment and correction strategies when the packet loss rates go beyond a certain value. In [3] we presented a multipulse speech coder which perfectly complies with the first and second condi-