The scatter of the data points in Fig. 3 indicates that there might be room for further lowering the Rayleigh scattering of PCVD fibres by optimising the preparation processes.

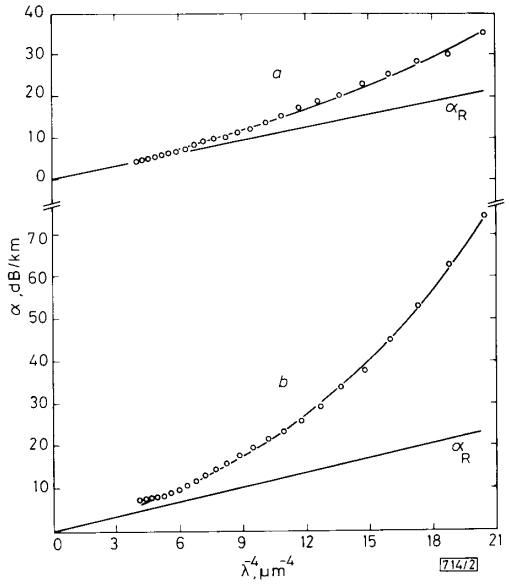

Fig. 2 *Short wavelength attenuation of Ge + F-doped SM fibres*
a $x_{Ge}^{eff} = 4\%$ *b* $x_{Ge}^{eff} = 10\%$
——— 3-parameter fits on A, K_{Ge}, λ_{Ge} (eqns. 1 and 4)

Summary: The attenuation spectra of Ge, F and Ge + F-doped PCVD fibres have been analysed with respect to scattering, absorption and profile design. The UV absorption is dominated by the Ge dopant and increases with its concentration. The increase of the Rayleigh scattering is proportional to the sum of the F and Ge-dopant concentrations. The results obtained approach the intrinsic absorption and scattering properties as estimated from theory and bulk silica scattering data. No significant influence of the index profiling itself has been found. Thus minimum attenuation of the fibres can be achieved only by minimising the total dopant concentration of the fibre core.

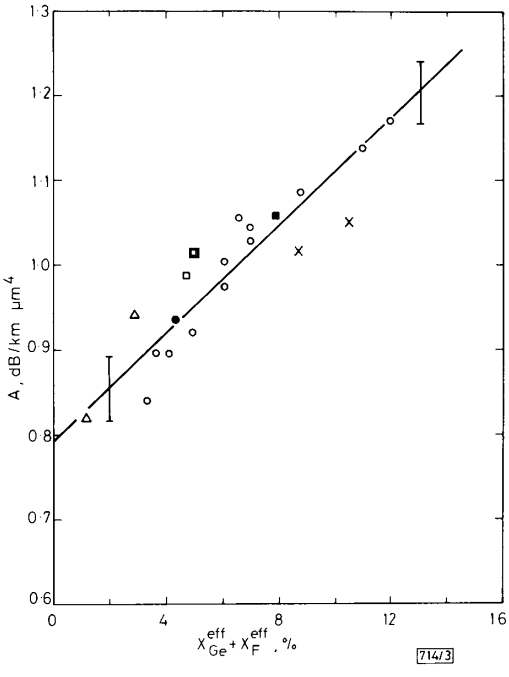

Fig. 3 *Rayleigh scattering coefficients of DC, MC, DFSM and GIMM fibres against total, effective, dopant concentration*
○ Ge △ F □ Ge + F-doped × multimode
● and ■ averages from many fibres

P. GEITTNER
H.-J. HAGEMANN
D. LEERS
Philips Research Laboratories Aachen
PO Box 1980, D-5100 Aachen, Federal Republic of Germany

31st October 1988

References

1 OLSHANSKY, R.: 'Propagation in glass optical wave guides', *Rev. Modern Phys.*, 1979, **51**, pp. 341–367
2 MUKHERJEE, S. P., and MOHR, R. K.: 'Light scattering in gel-derived glasses in the SiO_2–B_2O_3–Na_2O system', *J. Non-Cryst. Solids*, 1984, **66**, pp. 523–527
3 SCHULTZ, P. C.: 'Ultraviolet absorption of titanium and germanium in fused silica'. Proc. of 11th Int. Congress on Glass, Prague, 1977, **3**, pp. 155–163

RADIAL BASIS FUNCTION NETWORK FOR SPEECH PATTERN CLASSIFICATION

Indexing terms: Neural networks, Speech processing, Speech recognition

A neural network model incorporating radial basis functions is used in a speech-pattern classification problem. The method is compared with a back-propagation neural network model and with a vector-quantised hidden Markov model of the same problem. Training times are over an order of magnitude faster, with similar classification results.

Introduction: Recently, several researchers have applied neural network models to speech recognition problems.[1] The most frequently used algorithm has been the back-propagation of error algorithm,[2] which performs pattern classification by partitioning the input space using hyperplanes defined on the hidden nodes of the network. This algorithm has several drawbacks, a major one being the extremely long training times required for most problems. For the present work, a neural network model utilising the radial basis function[3] is developed and tested on a phonetic labelling problem. Comparisons with a back-propagation neural network model and with a vector-quantised hidden Markov model are made with reference to this problem.

Radial basis functions: A network using hidden nodes defining hyperellipsoids (rather than hyperplanes) may offer superior pattern classification. However, if the network is trained using the back-propagation method (using the chain rule to back-propagate errors) then training times are likely to be little better than for conventional back-propagation networks. An alternative philosophy would be to use many more hidden nodes, but with fewer degrees of freedom per node. This would give a potentially faster training algorithm.

Consider a network with hidden nodes computing their output as follows:

$$y_{pi}^H = \Phi(\|\boldsymbol{x}_p - \boldsymbol{c}_i\|) \tag{1}$$

where y_{pi}^H is the output of hidden node i in response to the pth input vector, $\boldsymbol{x}_p$, and $\boldsymbol{c}_i$ is a vector of similar dimension, representing the centre of a radially symmetric function Φ. Typically Φ is chosen as a Gaussian, so define

$$\Phi(\|\boldsymbol{x}_p - \boldsymbol{c}_i\|) = \exp\left\{-\sum_j \left[\frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right]\right\} \tag{2}$$

where σ_{ij} are the elements of a covariance matrix, which here is taken to be diagonal.

The set of hidden units consists of a set of functions which are presumed to constitute an arbitrary 'basis' for the patterns to be classified when expanded into hidden unit space; these

are referred to as radial basis functions. It is hypothesised that expansion of input vectors into a hidden unit space of relatively high dimension (i.e. by defining many radial basis functions) will result in a greater likelihood of the classification problem becoming linearly separable. (The properties of high-dimensional binary spaces are discussed by Kanerva[4] and a neural network model involving expansion into high dimensions inspired by Kanerva's work has been developed by Prager *et al.*[5]). In this case the output of the network may be defined as

$$y_{pi}^T = \omega_{i0} + \sum_j \omega_{ij} \Phi(\|\boldsymbol{x}_p - \boldsymbol{c}_j\|) \quad (3)$$

where y_{pi}^T is the output of the ith target node in response to the pth pattern, ω_{ij} is the weight from the jth radial basis function to the ith target node and ω_{i0} is the bias or threshold of the ith target node.

Such a network is trained by minimising the least squares error, E:

$$E = 0{\cdot}5 \sum_p \sum_i (Y_{pi}^T - y_{pi}^T)^2 \quad (4)$$

where Y_{pi}^T is the desired output of target node i in response to the pth input vector.

If the centres and widths of the radial basis functions are fixed then E may be minimised by an adaptive algorithm such as the LMS method.[5] However, it should be noted that eqn. 3 specifies that the outputs computed by the target units are linear in the weights. This is in contrast to most perceptron-like networks, in which the units execute some kind of nonlinear transfer function. Such a linear network is exactly soluble using a noniterative method that is similar to Wiener filtering. This method requires an error function that is purely quadratic in weight space (such as eqn. 4) and uses the inverse of the correlation matrix of radial basis function outputs, summed over patterns, to compute the optimal weight values.

Speech pattern classification: Phonetic labelling experiments were carried out on hand-segmented vowel tokens taken from a phonemically dense speech database of 98 sentences uttered twice by a male speaker. The input speech signals were sampled at 16 kHz prior to a 20th-order linear predictive analysis (with a pre-emphasis of 6 dB/octave), from which 20 cepstral coefficients were derived. The hand-segmented data consisted of approximately 750 vowel tokens (of 20 classes) extracted from the 98 sentences. These vowel tokens were represented by a feature vector consisting of the median cepstral coefficients for each third of the token (60 coefficients in total) plus 12 coefficients representing the duration of the token. The duration value in milliseconds was represented in a distributed fashion over the 12 coefficients using a Gaussian coarse-coding technique, enabling an arbitrary real number to be distributed over several units (coefficients) taking values in the range 0–1·0. Hence, each vowel token was represented as a static pattern of 72 coefficients.

The neural network model consisted of three layers: 72 inputs, 20 targets (representing the 20 vowel classes), and a variable number of radial basis functions. A standard 'one out of n' output coding was used for the targets, that is the desired target values for a given pattern were taken as $+\varepsilon$ for the target node corresponding to the desired class and $-\varepsilon$ for the other target nodes. In this work, the value of ε was taken to be 2·0. The centres of the radial basis functions were chosen by randomly choosing input vectors, in accordance with the sample distribution. Their widths were chosen using a nearest-neighbour rule, being equal to the Euclidean distance to the nearest function centre.

This neural network model was trained on all the vowel tokens extracted from one set of utterances and tested on the vowel from the second (unseen) set of utterances. Experiments were carried out on networks with varying numbers of radial basis functions: each network configuration was trained and tested 12 times (using a different random selection of function centres on each training) and the mean recognition score (%) and standard deviations of training and testing classification scores were computed over the 12 experiments. These results are given in Table 1, together with results obtained on the same problem using a back-propagation neural network model and a vector-quantised hidden Markov model. The back-propagation network had 36 hidden units and was trained using the standard gradient descent algorithm (with momentum).[2] The discrete hidden Markov model[7] was trained using the forward–backward algorithm and a size 256 codebook; additionally a codebook update technique was used, which approximates a discrete hidden Markov model to a continuous model. All classifiers were trained on input data based on 20th-order LPC cepstral coefficients.

Table 1 CLASSIFICATION SCORES FOR THREE CLASSIFIERS TRAINED ON 758 VOWEL TOKENS AND TESTED ON 759 (UNSEEN) VOWEL TOKENS

		Training set		Test set	
Classifier	*N(hid)*	Mean	SD	Mean	SD
		%	%	%	%
RBF	64	74·0	0·94	65·3	0·86
RBF	100	78·5	1·12	68·4	1·31
RBF	144	84·0	0·84	70·4	1·44
RBF	170	85·8	0·48	71·2	0·60
RBF	196	87·1	1·83	71·5	2·13
RBF	256	90·6	1·39	73·3	1·53
BP	36	98·0	0·30	71·6	0·44
HMM	—	92·1	—	69·4	—

RBF indicates radial basis functions neural network model, BP a back-propagation neural network model and HMM a vector-quantised hidden Markov model. Statistics from RBF classifier were computed from 12 sets of randomly chosen centres, those from BP classifier from 5 random initial states. *N(hid)* refers to number of hidden units or radial basis functions in neural network model

The results indicate that the recognition accuracy of the radial basis functions network is fully comparable with a back-propagation neural network model and a hidden Markov model. Performance of the radial basis functions network increases with the number of radial basis functions in the network, although this effect becomes negligible when the number of functions reaches a certain level (144–170 in this work). However, this algorithm does fail on some pathological choices of function centres, causing the optimal solution to involve extremely large weight values—in the experiments performed during this work, this situation was not observed in networks with less than 170 radial basis functions. This effect may be ameliorated by choosing an error function which incorporated a factor that is inversely proportional to the number of training examples in a class. That is, replace eqn. 4 by

$$E = 0{\cdot}5 \sum_p \sum_i (Y_{pi}^T - y_{pi}^T)^2 m_p \quad (5)$$

$$m_p = \frac{1}{P_c}$$

where there are P_c examples from target class c in the training set. This data-balancing term does solve the problem of 'weight explosion', but can cause other problems, in which the output of the network for some patterns has an output with all targets set to their minimum values.

It is also illuminating to compare training times. On this vowel classification problem, a network with 144 radial basis functions could be trained in less than 4 minutes. In comparison the building of the vector quantisation codebook for the hidden Markov model took approximately 3 hours and training the back-propagation neural network model took approximately 3 hours (on a pipelined array processor, which gives a 5–10 times speed increase compared with the serial hardware used for the other methods). On similar hardware, training a 36-node back-propagation network is over two orders of magnitude slower than training a radial basis functions network with 144 radial basis functions.

Acknowledgment: I wish to thank Dr. R. Rohwer for valuable insights and discussion during the course of this work, and for offering the method used to solve the linear network.

S. RENALS *12th December 1988*

Department of Physics
and Centre for Speech Technology Research
University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN, United Kingdom

References

1 WAIBEL, A., HANAZAWA, T., HINTON, G., SHIKANO, K., and LANG, K.: 'Phoneme recognition: Neural networks vs. hidden Markov model'. Proc. IEEE ICASSP-88, New York, 1988, pp. 107–110
2 RUMELHART, D. E., HINTON, G. E., and WILLIAMS, R. J.: 'Learning Internal Representations by Error Propagation', *in* RUMELHART, D. E., and MCCLELLAND, J. L. (Eds.): 'Parallel distributed processing: Volume 1' (MIT Press, 1987), pp. 318–362
3 BROOMHEAD, D. S., and LOWE, D.: 'Multi-variable functional interpolation and adaptive networks', *Complex Syst.*, 1988, **2**, pp. 321–355
4 KANERVA, P.: 'Self-propagating search: a unified theory of memory'. PhD thesis, CSLI, Stanford University, 1984
5 PRAGER, R. W., HARRISON, T. D., and FALLSIDE, F.: 'Boltzmann machines for speech recognition', *Comput. Speech Lang.*, 1986, **1**, pp. 3–27
6 WIDROW, B., and STEARNS, S. D.: 'Adaptive signal processing' (Prentice-Hall, Englewood Cliffs, 1985)
7 HUANG, X. D., JACK, M. A., and ARIKI, Y.: 'Parameter re-estimation in semi-continuous hidden Markov modelling of speech with feedback to the vector quantisation codebook', *Electron. Lett.*, 1988, **24**, pp. 1375–1376

BIAS DEPENDENCE OF LOW-FREQUENCY GATE CURRENT NOISE IN GaAs MESFETs

Indexing terms: Semiconductor devices and materials, FETs, Noise

The spectra of gate current noise are investigated in GaAs MESFETs between 10^2 and 10^4 Hz. A change in the white-noise behaviour is observed with the increase of the gate current. It is shown that the contribution of an ideal Schottky shot noise is associated with two thermal noise components. The thermal noise sources originate in different leakage conductances.

Introduction: The low-frequency noise levels of GaAs MESFETs have important implications even in their microwave properties. The noise behaviour of transistors is generally represented by two input noise generators in a conventional $e_n - i_n$ representation. In the common-source configuration these equivalent noise generators are connected to the gate. Usually the magnitude of i_n is estimated by measuring the output noise voltage as a function of R_g, the resistance of the gate biasing circuit.[1,2] Apart from this method, the direct measurement of the input equivalent noise current by means of a current preamplifier gives an effective alternative tool to obtain input characteristics whatever the conditions placed on the AC input and output load impedances.

The results indicate that various noise sources are successively dominant, depending on the gate voltage. Thermal noise due to a gate parallel conductance is evident at low voltage. At higher voltages a source of shot noise type dominates, marked by an equivalent current well beneath the full gate current.

Experimental data: Noise investigations were made on dual-gate MESFETs (NEC 41137). The driving gate and the control gate were connected together forming a single gate. For more convenient measurements, the drain and the source have been short-circuited so that the device can be seen as a diode.

The DC gate current varies by an order of 10^3 for voltages from 2 mV to 2 V following a behaviour significantly different from those of an ideal metal/SC interface. A linear relationship, associated with a conductance G_0, is first observed at low currents, followed by a variation corresponding to a weak decreasing conduction. Such dependence is generally caused by the barrier height reduction induced by a disturbed layer, due to surface states. According to the interfacial layer theory, the reverse current for GaAs diodes is of the form $J_R \propto \exp(-0{\cdot}26X^{1/2}\delta)$ in which the barrier height field dependence X and the layer thickness δ are used as adjusting parameters.[2,3]

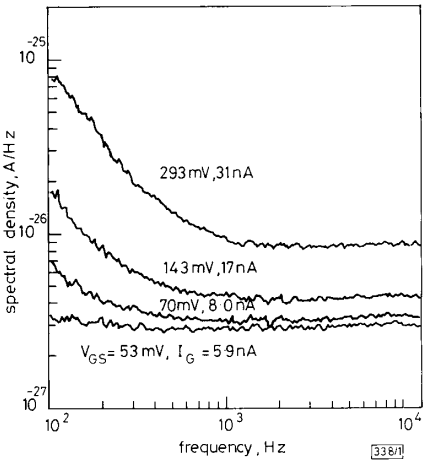

Fig. 1 *Spectra of gate current noise density for various bias voltages*

Fig. 1 shows typical direct spectra of the input noise current obtained at several gate currents. Only typical $1/f$ and white noise levels are present, without any significant contribution by G-R or by trapping effects. At lower currents the noise is plainly white over the entire frequency range and independent of the I_G value. At higher currents, the white level increases as I_G, combined with an uncorrelated $1/f$ spectrum in the low-frequency region. The flicker noise, developed in previous work,[5] is subjected to the input conductance given by the weak currents.

The problem is to identify the locations of various white noise sources, which can be either thermal or Schottky noises and their combinations. The spectral density S_i, formed only by the white noise component, is plotted in Fig. 2 in the form $I_{eq} = (S_i/2q)$ against I_G. The gate current allows comparison with the shot noise mechanism, which is the exclusive effect of an ideal Schottky diode.[6,7] This reference is then expressed in the practical form $I_{eq} = I_G$. The result clearly indicates two well separated zones when the gate current increases. From the equivalent noise level $4kTG_o$ we found $G_o = 1{\cdot}90 \times 10^{-7}$ S in full agreement with the value $1{\cdot}85 \times 10^{-7}$ S given by the $I_G - V_{GS}$ characteristic. A similar agreement has been

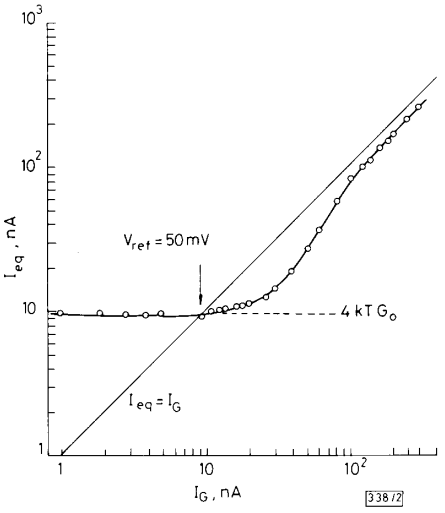

Fig. 2 *Equivalent current noise against reverse gate current*