A Robust Speaker Verification Biometric

M.H. George* and R.A. King†

*Domain Dynamics Limited. †Cranfield University (RMCS).
Shrivenham, Swindon, England SN6 8LA.

Abstract — The need for simple, ubiquitous security procedures to verify the identity of authorised system users is assuming increasing importance in the expansion and exploitation of new technology. High performance *Biometric Verification* methods offer important tools for this purpose, enhancing the security, reliability and integrity of transactions conducted electronically. This paper describes TESPAR/ FANN, a new *digital data / artificial neural network* combination which is proving highly effective in the Speaker Verification arena, and in other key non-speech applications.

INTRODUCTION

Secure systems, data and locations are currently protected from unauthorised access by a variety of devices. These may include PIN pads, keys both conventional and electronic, identity cards, cryptographic and dual control procedures. Whilst such "terminal to terminal" procedures are well established and largely automatic, the "human to terminal" link remains vulnerable to attack from unauthorised individuals often by simple theft of access devices and codes, or by other well-known criminal impersonation strategies.

Biometric verification

As electronic and computer systems increasingly dominate the gatekeeping function to transactions of high value and strategic importance, so the need for system and network security assumes high priority. Even in the relatively benign commercial arena, projected ISDN services and UPT systems demand high integrity access control and security. The need for intelligence to verify the identity of users of such systems has been emphasised as "*a most urgent need*." See Pandya [1]. In a more hostile security environment, this capability may constitute a vital operational requirement.

Given the limitations of conventional security procedures, a range of *biometric* verification options are currently under consideration. The idea is to enable automatic verification of identity by computer assessment of one or more behavioural and/or physiological characteristics of an individual.

Conventional methods

In the case of human speech recognition, conventional *frequency domain analysis* of speech has to date been almost exclusively applied in attempts to extract a set of statistical biometric features associated with the voice of a particular speaker. These are usually coded into a data set or *template*

The idea of automatic verification of a person's identity by the acoustic analysis of their speech is of course not new.

Voice biometrics

Speaker Verification is a biometric which offers an ability to provide positive verification of identity from an individual's voice characteristics. For example, system access may be authorised simply by means of an *enrolled* user speaking over the telephone or into a microphone attached to the system. The system analyses the characteristics of that voice sample to determine if there is a sufficient match to a set of characteristics analysed and extracted at the time of enrolment of that user.

For commercial applications this idea is compelling, given the increasing use of Smart Cards, telephone banking, share dealing, home shopping, and the potential for new communication networks such as GSM and UPT to conduct high value business transactions - all targets for the resourceful criminal.

In the Public Sector a large number of applications are amenable to speaker biometric verification, including:

- Access control to secure systems, data records and physical locations, where there may often be a requirement for "hands free" operations
- Radio, Mobile Telephone and other communications systems
- Prison Payphones
- Passport control
- Benefits payments
- Eligibility and enrolment for Health and Social Services
- Enforcement of Bail, Parole and non-custodial activities such as Community Service

While biometric verification techniques may appear highly attractive technically, social issues must not be ignored. Market research studies have shown that voice methods rate high in public acceptability, while some other options suffer a lower acceptability rating [3].

DEVELOPING A VOICE BIOMETRIC CAPABILITY

requiring time normalisation. Such systems suffer from a number of serious limitations, associated with, for example, tedious enrolment procedures, high computational complexity, and a vulnerability to artefacts such as background noise or so called benign traumas common in the real world operating environment.

Having extracted these templates, the classification task is by no means easy. To overcome the time variability of the template, complex techniques such as dynamic time warping must be applied in an effort to normalise the data set from one speech sample to another. Mathematical correlation distance scoring techniques are routinely used to differentiate among templates, but suffer from a number of limitations. More recently both Hidden Markov Models (HMMs) and Artificial Neural Networks (ANNs) have been investigated for the classification task. Unfortunately the time variability inherent in templates generated by current conventional frequency domain data analysers creates formidable difficulties for the convenient and effective application of Neural Network architectures to the verification task. See for example Morgan and Scofield [2].

TESPAR/FANN TECHNOLOGY

The work we describe has capitalised upon TESPAR/FANN methods, a new data / neural network combination which is proving highly effective in the speaker verification arena. TESPAR/FANN involves the integration of novel *Time Encoded Signal Processing And Recognition* (TESPAR) waveform coding procedures with orthogonal *Fast Artificial Neural Networks* (FANNs) structured for this purpose, in a decision making / data fusion hierarchy which enables verification of the identity of individuals, by means of them speaking a simple common phrase.

TESPAR coding

TESPAR is a new simplified digital language, first proposed by King and Gosling [4] for coding speech. The process however may be extended to any information bearing entity that can be represented in terms of a band-limited signal. The range so far investigated encompasses seismic signals with frequencies and bandwidths of fractions of a hertz, to radio frequency signals in the gigaHertz region, and beyond.

TESPAR is based upon a precise mathematical description of waveforms, involving polynomial theory, which shows how a signal of finite bandwidth ("band-limited") can be completely described in terms of the locations of its real and complex zeros. This contrasts with the more conventional approach of linear transformations based on "amplitude" sampling at regular intervals, as has been described by Fourier, Nyquist, Shannon and others. The real and complex zero descriptors of TESPAR and the time-bandwidth data produced by a Fourier transform are mathematically equivalent, and both result in 2TW (the vital Shannon Number) of digital sample data points describing the waveform. The mathematical underpinnings of this zero-based approach are outlined in Voelcker [5] and Requicha [6].

Given the real and complex zero locations of the signal, a vector quantisation procedure has been deployed to code these data into a small series of discrete numerical descriptors, typically around 30 (the TESPAR *symbol alphabet*). Holbeche [7] gives an account of one version of this coding.

Matrix formation

The output from a TESPAR coder is a simple numerical symbol stream which may be converted into a variety of progressively informative matrix data structures. For example, the single-dimension vector (or *S-matrix*) is a histogram recording the frequency with which each TESPAR coded symbol occurs in the data stream. A more discriminating data set is the two-dimensional histogram or *A-matrix* which is formed from the frequency of symbol pairs, which need not necessarily be adjacent. Extending this to 3 dimensions would improve the discrimination power still further. Typical A and S matrices are shown in figures 1 and 2. See also King [8].

Classification

TESPAR data structures are of fixed size, dependent upon the alphabet used. This makes for regimes of processing that are both stable and simple to implement. In the speaker verification task, TESPAR matrices for several utterances of an individual speaker may be collected during the enrolment process, and used to produce a reference matrix or *archetype* which embodies the unique characteristics of that speaker.

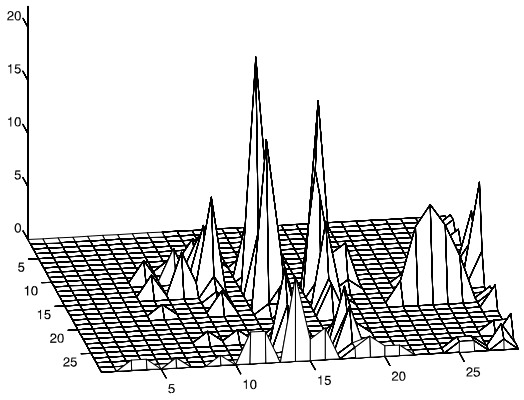

Figure 1
Un-normalised A-matrix, male speaker

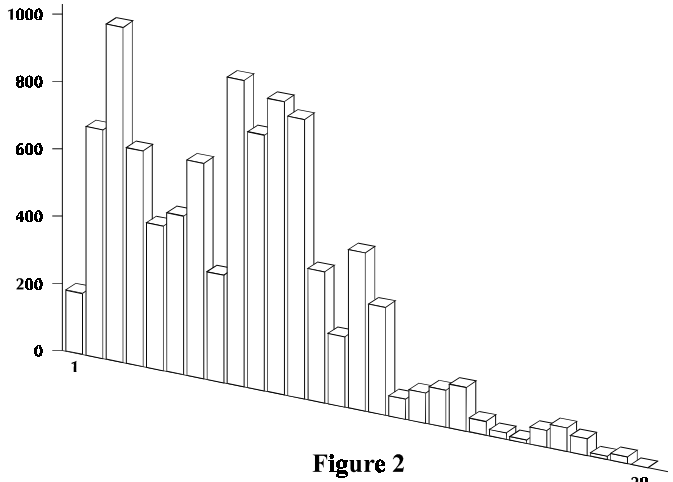

Figure 2
Un-normalised S-Matrix, male speaker

Subsequently at the time verification is required, a new matrix is created live and compared against the reference for a decision to be made. Standard correlation statistical methods can be applied in the decision making process, and yield useful results.

The Artificial Neural Network dimension: Potentially far more powerful is the possibility of applying Artificial Neural Network methods of pattern classification to the TESPAR matrices. Because TESPAR matrices are of fixed size and dimension, they are ideally matched to the input requirements of Neural Networks. Recent practical experience confirms that the TESPAR/FANN combination enables the introduction of very powerful classification procedures, producing system performances previously considered infeasible.

Enrolment strategies: There are many possible enrolment options: a universal phrase, a set of random words or digits, unique passwords, and various other combinations are all available. We have based our work on the use of *a simple common universal phrase*. By this means TESPAR/FANN classification methods focus on the differences and similarities in the characteristic data structure of each speaker, rather than on attempting to recognise both the words and the speakers. Techniques of prompted random words and digits can be very attractive for some applications, and these are equally amenable to TESPAR/FANN methods.

Performance advantages

TESPAR-based verification techniques are presenting significant performance advantages over conventional Fourier based methods. For example:

- they have typically 2 orders of magnitude lower computer processing power requirement, with consequent lower power consumption.
- they use and form simple data structures which are both compact and of known size so that limited memory resources in embodiments such as Smart Cards can be employed efficiently. This has important benefits for data storage and transfer operations.
- the data structures are optimally matched for classification methods that use FANN architectures.
- samples can be obtained direct from low cost analogue sensors such as telephone handset microphones.
- they offer extremely high degrees of discrimination.
- classification procedures and architectures can, by routine design, enable system errors to be made vanishingly small over a wide range of real world applications and environments.
- verification speed is minimal, e.g. less than 1 second using current popular microprocessor technology for a single pass interrogation.

For these and other productive reasons TESPAR/FANN is implementing the biometric functions required in the European Union CASCADE Esprit Smart Card project which is

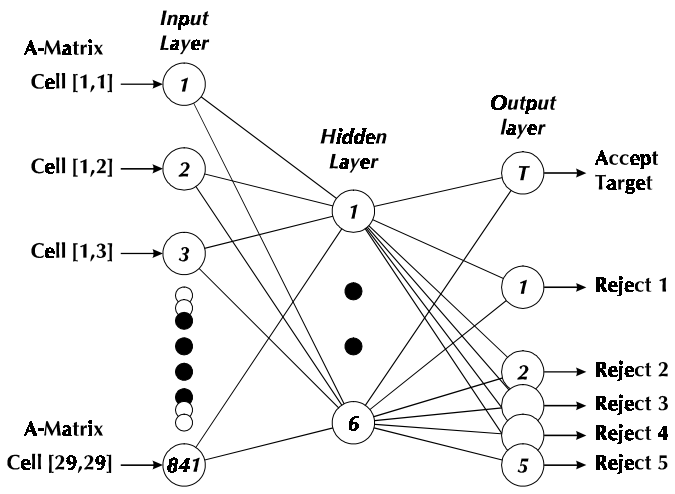

Figure 3
Typical Neural Network Configuration (A-Matrix)

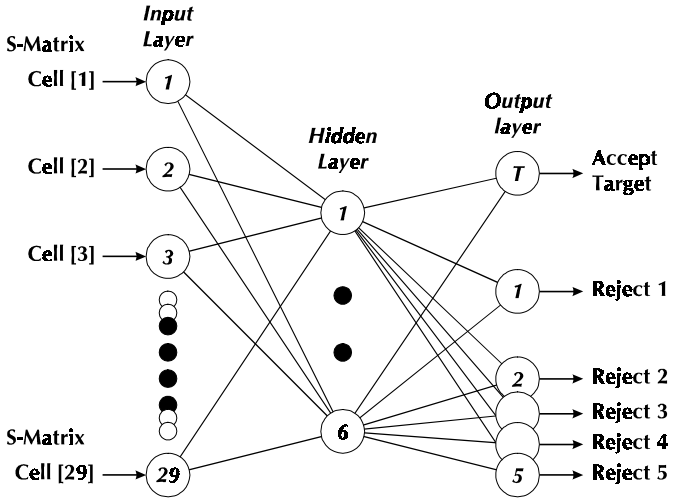

Figure 4
Typical Neural Network Configuration (S-Matrix)

developing a 32-bit RISC processor 20 square mm in area for a new generation of Smart Card and secure Pocket Intelligent Device applications [9].

Silicon issues

The TESPAR coding and vector quantisation process is already available both as a software algorithm, and in a low power ASIC silicon design. Beyond this, negotiations are currently well advanced with TriTech Microelectronics to produce a range of very low cost, low power TESPAR embodiments in silicon which offer a high degree of flexibility for integration into a wide range of potential high volume TESPAR applications.

In parallel and in association with this activity a collaboration with Kings College and University College London is under way to adapt their pRAM Neural Network architecture [10, 11] to the task of classifying TESPAR data structures. pRAM technology provides Neural Networks that *can be trained on the silicon itself.* Thus the realisation of complete TESPAR/FANN single chip solutions is in sight, capable of training in situ and adaptable to widely differing low cost, high volume applications.

Comment: These results were obtained despite the fact that some 8% of the FANNs created did not converge fully.

EVALUATIONS OF THE TESPAR/FANN VOICE BIOMETRIC

To assess the TESPAR/FANN voice biometric capability, we have conducted extensive trials, including testing on a database of 150 male and 68 female speakers in an evaluation consuming 16 man months of effort [8].

The database provided 20 versions of a single common phrase, “Sir Winston Churchill”, recorded for each speaker in a 3 second time interval under a variety of ambient acoustic noise conditions - a total of 4360 samples. Each utterance was successively converted into PCM files and TESPAR S-matrices, from which 15 different 3-layer FANNs (figure 4) were constructed and trained for each target speaker. On interrogation, the 15 individual FANN output classifications were combined using a simple vote-taking “winner takes all” method (figure 5).

Using *supervised registration* procedures [8] the following results were obtained:

0 x False Reject errors out of 4360 interrogations (FRR < 0.023%)

4 x False Accept errors out of 2616 interrogations (FAR = 0.153%).

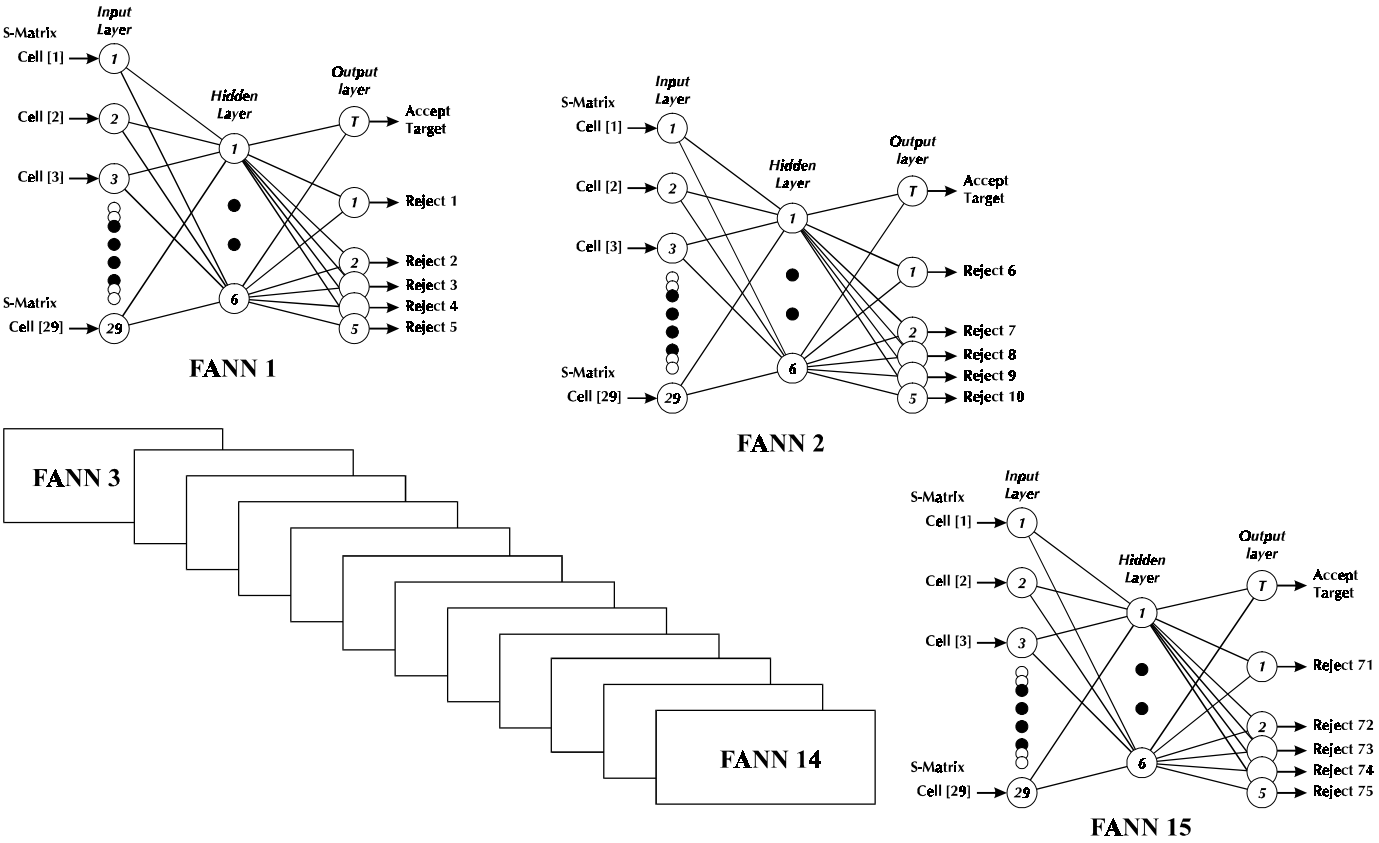

Figure 5
Typical TESPAR/FANN Speaker Verification Architecture

The phrase “Sir Winston Churchill” is not especially suitable (phonetically rich) for a typical real world application.

No FAR reduction strategies were deployed during registration.

One only of the numerous bespoke options [8] available was invoked to illustrate supervised registration.

The results appear significant and compare favourably with those currently reported in the literature for equivalent state of the art competitor systems.

The effects of benign traumas

As a preliminary to the extensive investigation described above, a pilot study was conducted into the effects of commonly occuring benign traumas on a typical TESPAR/FANN based speaker verification system [12]. Each speaker provided samples of the phrase “My name is Charles Westlake”. Testing was conducted under normal conditions, and when the speakers were affected by alcohol, by dental anaesthesia, by a cold, with an oral obstruction, and over an 11 month period of time.

The results of this trial indicated that the TESPAR/FANN combination appears substantially invulnerable to the effects of such conditions.

Design strategies

Bespoke strategies, such as supervised registration, admit a system which tailors the TESPAR/FANN data for classification to the individual characteristics of the target speaker. This gives the system great flexibility in dealing with idiosyncratic members of a target population (the “goats”).

Adoption of a multiple network architecture with the verification decision based on a data fusion / vote taking decision logic across the network set offers the possibility of making both FAR and FRR system errors vanishingly small by design. In practice, the verification architecture may consist of 15 or more networks, with a predicted error performance likely to meet or better the 1 in 100,000 target FRR performance figure set as a “requirement” by the UK banking community for biometric methods.

Development tools

All the work described has been conducted using a Domain Dynamics’ proprietary PC-based development system, the TADS-XS 50. The system includes an extensive library of both conventional and TESPAR signal processing and data analysis software, operating under the popular MATLAB™ graphical user interface. FANN classification architectures are created, trained, tested and interrogated within the system using the proprietary FasTEST software suite. This development facility is proving extremely valuable in enabling third parties to evaluate TESPAR/FANN architectures in a wide range of real world classification tasks.

Extension Of TESPAR/FANN To Other Security Applications

TESPAR/FANN techniques are applicable to the classification of any entity whose underlying information can be represented as a band-limited signal. Domain Dynamics has already applied this technology in over 50 case studies ranging from the condition monitoring of critical valves in a reciprocating compressor to the classification of nanosecond waveforms resulting from high voltage partial discharge defects in power transformers [13].

In the Security arena, TESPAR has been applied to the design of Perimeter Intrusion Detection Systems (PIDS) which discriminate amongst a variety of realistic hostile and benign conditions.

Signals from military vehicles and from sonar systems have also been classified, thus demonstrating the capability of identifying different types of target and their range.

Video

On the BBC Television Tomorrow’s World programme in April 1994 a demonstration of a TESPAR/FANN speaker verification system was shown in which the system correctly verified the presenter’s identity in the presence of highly intrusive background noise (church bells).

It also correctly rejected a high quality tape recorded sample of the presenter’s voice, demonstrating that the TESPAR/FANN technique is mathematically capable of differentiating between speech presented live and a “replay attack” from a high quality recording of the authorised person’s voice.

Such demonstrations illustrate the discrimination power of the TESPAR/FANN process, which is exemplified in its ability to classify many signals that remain indistinguishable in the frequency domain.

CONCLUSIONS

Experience to date with TESPAR/FANN hardware and software indicates:

1. The TESPAR/FANN combination is a powerful, robust, flexible and economic technology for a wide range of speaker verification and security applications.

2. Significant trials have confirmed exceptionally low error rates when compared with the currently reported conventional methodologies.

3. The TESPAR/FANN procedures described permit system errors to be made vanishingly small over a wide range of operational speaker verification applications.

4. TESPAR/FANN technology is available now for developing a wide range of powerful, cost-effective operational speaker verification embodiments.

5. TESPAR/FANN may be applied to other security objectives where the classification of hostile and benign events is of critical operational importance.

6. Hardware and software development tools are readily available for solving speaker verification and other real world signal classification problems.

ACKNOWLEDGEMENTS

Thanks are due to:

- Domain Dynamics Limited for their permission to publish this paper and for their support and funding of the research work under which the TESPAR/FANN technology has been developed.
- The Principal of Cranfield University (RMCS) for his permission to publish this paper.

REFERENCES

[1] R. Pandya, “No escape from the global telephone”. New Scientist, 19 October 1991, p. 24

[2] D.P. Morgan and C.L. Scofield, Neural Networks and Speech Processing. Mass., USA: Kluwer Academic Publishers, 1991

[3] Editor E. Newham, “A Basic Comparison of Biometric Methods”, Biometric Technology Today, vol. 1 (1), p. 7, April 1993

[4] R.A. King and W. Gosling. Electronics Letters, vol. 14 (15), pp. 456-457, 1978

[5] H.B. Voelcker, “Toward a unified theory of modulation”. Proceedings of the IEEE, vol. 54 (3), pp. 340-353; and vol. 54 (5), pp. 735-755, 1966

[6] A.A.G. Requicha, “The zeros of entire functions. theory and engineering applications”. Proceedings of the IEEE, vol. 68 (3), pp. 308-328, March 1980.

[7] J. Holbeche, R.D. Hughes, and R.A. King, Proceedings of the IEE International Conference on Speech Input/Output: Techniques and Applications, pp. 310-315, 1986

[8] R.A. King, “TESPAR/FANN: an effective new capability for voice verification in the defence environment”, presented at the Royal Aeronautical Society Conference on The Role of Intelligent Systems in Defence, London, March 1995, p. 5.1-5.8

[9] CASCADE Esprit Project EP8670 Data Sheet, 1995

[10] D. Gorse and J.G. Taylor, “A review of the theory of pRAMs”, in the Proceedings of the Weightless Neural Network Workshop ‘93, University of York, April 1993.

[11] T.G. Clarkson, C.K. Ng and J. Bean, “A Review of Hardware pRAMs”, in the Proceedings of the Weightless Neural Network Workshop ‘93, University of York, April 1993.

[12] R.A. King et al, “The Effects of Commonly Occuring Benign Traumas on TESPAR/FANN based Speaker Verification Systems”. Internal Report for The Woolwich Building Society, 1992.

[13] J. Fuhr, M. Haessig, P. Boss, D. Tschudi and R.A. King, “Detection and location of internal defects in the insulation of Power Transformers”, IEEE Transactions on Electrical Insulation, vol. 28 (6), pp. 1057-1067, December 1993.