Shannon, TESPAR And Approximation Strategies

R. A. King and T. C. Phipps

Domain Dynamics Limited
Cranfield University, RMCS
Swindon SN6 8LA
United Kingdom
email: ddl@rmcs.cranfield.ac.uk

Abstract

This paper outlines the development and application of an alternative embodiment of Claude Shannon's celebrated sampling theorem qualified by Shannon in 1949 and tested more recently by the authors, via the classification of a wide variety of real-world band-limited waveforms. The work of Voelcker, Requicha et. al. is called upon and developed, to indicate key features of the basic coding concept, designated "Time-Encoded Signal Processing And Recognition" (TESPAR).

TESPAR coding is based upon approximations to the locations of the 2TW Real and Complex Zeros, derived from an analysis of the band-limited waveforms under examination. Progressively informative numerical descriptors of the waveform may be obtained via a ranking of the 2TW samples ("Shannon Numbers"), derived from the analysis.

TESPAR concepts are illustrated by reference to recent case-studies involving signal classifications across a spectrum from very-high frequency nano-second waveforms to very-low frequency waveforms in the sub-Hz range. The key features of TESPAR in the speech processing arena are emphasised, illustrating:-

- A capability to separate and classify many signals of interest that are indistinguishable in the frequency domain.

- An ability to code time-varying speech waveforms into optimum configurations for processing by Artificial Neural Networks.

- An ability to deploy, economically, massively parallel architectures for productive data fusion.

Introduction

Probably no single work this century has more profoundly altered man's understanding of communication than Claude Shannon's *A Mathematical Theory Of Communication*, first published in 1948 [1]. The ideas in Shannon's papers have been addressed by communication engineers and mathematicians around the world and have been elaborated on, extended and complemented with many new related ideas.

In a follow-up paper in 1949 [2], Shannon formalised the ***sampling theorem*** which now forms the basis of the majority of today's digital speech and data transmission systems. In his paper he defines a **THEOREM 1** as follows:

If a function f(t) contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced 1/2W seconds apart.

We note, however, that Shannon carefully describes his THEOREM 1 as, ***one answer*** which satisfies the requirements previously addressed. Thus, in this model, ***sampling*** involves determining the ordinates of a waveform at a series of points equally spaced 1/2W seconds apart.

From this basic theorem, exemplified in Figure 1, a universe of practical manifestations have been developed. This set of sampled data has formed the basis for a variety of mainly linear transformations e.g. Fourier, Linear Prediction, Wavelet and Walsh Coders as informative transformations for describing and classifying key features of the sampled data set.

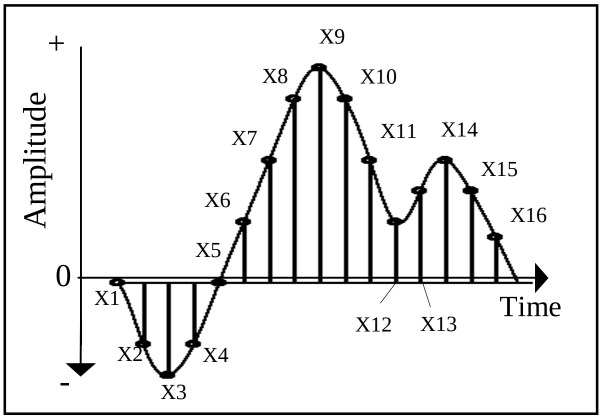

Figure 1 - Regular Sampling

Thus, current coding strategies involve an essential trio of requirements, viz.:

a) The use of amplitude descriptors.

b) The use of regular sampling.

c) An approximation domain which is magnitude or amplitude-based, i.e. dependent upon the number of bits per sample used to define the sampled ordinate values.

Infinite Clipping

At about the same time that Shannon's celebrated work was being published, two researchers: Licklidder and Pollack [3], were investigating the effects of ***amplitude clipping*** on the intelligibility of the speech waveform, a process then in wide use by the amateur radio community. Licklidder and Pollack extended this process to produce the so-called ***infinite clipping*** format whereby ***all amplitude information*** was removed from the waveform, resulting in a binary transformation that preserved only the zero-crossing points of the original signal - see Figure 2.

Surprisingly, when the speech waveform was differentiated, prior to infinite clipping, ***mean random-word intelligibility scores of 97.9% were achieved***.

From these observations it would appear that a substantial proportion of the information of interest in a speech waveform - i.e. its ***intelligibility*** - is contained solely in its zero-crossings.

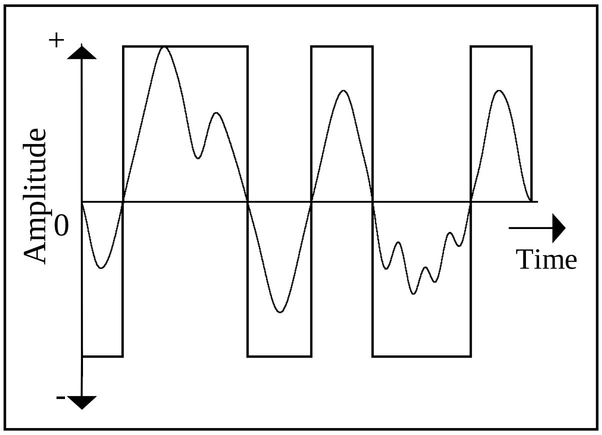

Figure 2 - Infinite Clipping Preserves Only Zero-Crossing Information

This evidence calls into question the special status of the three properties a), b) and c) described previously. It is clear that, having discarded all information relating to intermediate ordinate values, the infinitely-clipped samples represent the ***duration*** of the intervals between the zero-crossings of the waveform. These ***zero-based*** durations are thus ***signal-derived*** and not generated at regular 1/2W second time intervals. Clearly, for speech waveforms these samples will be ***irregularly spaced***.

These surprising observations provided the intellectual catalyst for the current work which has resulted in the creation and development of Time Encoded Signal Processing And Recognition (TESPAR).

A Qualification Of *Infinite Clipping* From Shannon's THEOREM 1

The authors are aware that any disagreement with Shannon is likely to be painful ! If, however, and mindful of the infinite-clipping experience, we revisit the last paragraph of Shannon's description of the sampling theorem, we note that he makes the following statements:-

d) “The 2TW numbers used to specify the function ***need not be the equally-spaced samples*** above. For example the ***samples can be unevenly spaced***.”

and:-

e) “Generally speaking, ***any set of 2TW independent numbers associated with the function can be used to describe it***.”

From this we may conclude that the lack of features a) and b) in the Infinite-Clipping phenomenon do not necessarily disqualify it, according to Shannon. Indeed, we may feel that unevenly-spaced zero-based durations fit quite comfortably alongside the requirements of his THEOREM 1.

Item c) however, raises important questions relating to the ***approximation domain*** associated with infinite clipping. For example:-

f) Are the unevenly-spaced zeros of the infinitely-clipped signal perhaps a necessary but not-sufficient data set for a fully effective approximation of speech waveforms ?

and more generally:-

g) Can zeros provide a respectable set of information-bearing attributes for the analysis and synthesis of waveforms in general ?

Approximation Issues

We observe that the infinite clipping data set represents some form of approximation to the original waveform which is surprisingly effective in preserving the intelligibility of band-limited speech. We now seek an insight into the features of this approximation with a view to extending it for further exploitation.

In a paper on *Approximation Spaces* [4], Preston C. Hammer provides some comfort for us by confirming that:

h) “All models and simulations of systems are approximations.”

i) “There is no *a priori* measure of goodness of fit which is satisfactory for all purposes.”

Hammer also identifies, as a key modelling objective, an answer to the following question:

j) “What properties of a function should an approximation to it preserve ?”

It would appear therefore from Shannon, Licklidder and Pollack, and Hammer, that zeros may be worthy of further investigation as a means of providing approximations of merit for some signals.

Zero-Based Analysis Of Signals

The beginnings of a formal zero-based theory of signals were indicated by Bond and Cahn [5] in 1958. By applying a set of algorithms developed by Titchmarsh [6] , they were able to show how the obvious inadequacies of the simple infinitely-clipped model could be overcome by the introduction of the concept of ***complex zeros***. They concluded that:

k) The aggregate of zeros (***real and complex***) occur in the limit at the Nyquist (i.e. Shannon 2TW) rate.

and:-

l) Continuous band-limited functions generated by natural information sources will include ***complex zeros that are not physically detectable.***

It is now apparent that the elements of a zero-based theory of signals have been available to mathematicians for many years. Their application to engineering problems has, however, not yet been fully realised. Voelcker [7][8] and Requicha [9] investigated the zero properties of signals. In two key papers, Voelcker brought together many of the relevant aspects of analytic functions into a framework which he then used to study the modulation of signals. Requicha was subsequently able to demonstrate the application of these techniques to many important areas of engineering.
The practical problem of zero extraction however, i.e. the construction of a real and complex zero sequence that corresponds to a given ***waveform*** , is generally non-trivial. The real zeros of a function, which are its conventional zero crossings, are easy to determine by, for example, ***clipping***, however the complex zeros of a function are not easy to determine. Visual inspection of the waveform may provide information on the location of ***some*** of its complex zeros, but the only known quantitative method for finding the location of ***all*** of the complex zeros involves the

numerical factorisation of a 2TWth-order trigonometric polynomial.

Consider a waveform of bandwidth W and time T which contains 2TW zeros, and where typically 2TW exceeds several thousand. Although a numerical factorisation of such a 2TWth-order polynomial is of academic interest, its calculation is computationally infeasible. This fact has proved a serious deterrent to the exploitation of the zero model.

The key to the TESPAR-based exploitation of the formal zero-based mathematical analysis is the realisation that a progressively informative ***approximation*** to the original signal need not involve the numeric factorisation of a 2TWth-order polynomial.

Instead, a highly informative zero-based model may be embodied via procedures that, segment the waveform between successive real zeros, and combine this duration information with simple approximations to the shape of the waveform in between these two locations. That is to say, recording and classifying those complex zeros that ***can*** be identified directly from the waveform.

In such a decomposition of signals into their real and complex zeros, it is noted that real zeros in the time domain are identical to the locations of the real zeros in the zero domain. Complex zeros appear in conjugate pairs and these are associated with perturbations - such as minima, maxima, points of inflexion etc. - in the wave ***shape*** that appear between the well-defined real zeros. Thus a sufficient number of important subsets of complex zeros may be identified by examining the features of the waveshape between its successive real zeros.

In the simplest implementation of a TESPAR coder, two descriptors associated with each such segment - or ***epoch*** - may be used to generate a basic TESPAR symbol alphabet, these are:

m) The ***duration*** between successive real zeros.

n) The ***shape*** between successive real zeros.

In the simple TESPAR model not all complex zeros can be identified from the shape, ***the approximation is therefore limited to those zeros that can be so identified.***

It will be apparent that the band-limited nature of a signal imposes significant restrictions upon the maximum and minimum ***duration*** of any epoch and also upon the maximum number of significant waveform ***features*** (extrema, points of inflexion etc.) that each epoch may contain. The longest epoch may have a duration approximately equal to half the period of the lowest frequency component allowed by band-limiting. Similarly, the shortest epoch may have a duration approximately equal to half the period of the highest frequency component allowed within the band of interest. Thus epoch length may be further approximated by quantisation within some predetermined range.

Clearly, short epochs cannot exhibit a multiplicity of minima as this would imply frequency components outside the bandwidth defined. Thus short epochs will in general be simple in nature, i.e. contain either no, or few, features, whilst long epochs may be either simple or contain many features.

The maximum number of ***extrema*** which any epoch can contain is given approximately by a function of the ratio of, the fundamental interval of the highest allowed frequency component, to the duration of the epoch. Again, the range of values over which the minima count for each epoch can vary is bounded. Thus, to a first-order approximation, each epoch may be classified in terms of its duration (D) and the number of minima, i.e. shape (S), that it contains.

TESPAR Alphabet

The TESPAR ***alphabet*** results from a vector quantisation process in which a generalised code book - or ***symbol table*** - is used to map the duration/shape (D/S) attributes of each epoch to a single descriptor - or ***symbol***.

	“S” = 0	**“S” = 1**	**“S” = 2**	**“S” = 3**	**“S” = 4**	**“S” = 5**
“D” = 1	1					
“D” = 2	2					
“D” = 3	3					
“D” = 4	4	4				
“D” = 5	5	5				
“D” = 6	6	6	6			
“D” = 7	6	6	6			
“D” = 8	7	8	8	8		
“D” = 9	7	8	8	8		
“D” = 10	7	8	8	8	8	
“D” = 11	9	10	10	10	10	
“D” = 12	9	10	10	10	10	10
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
“D” = 37	23	24	25	26	27	28

Figure 3 - Standard 29 Symbol TESPAR Alphabet

The symbol table is normally configured to cater for the most probable D/S pairings and can be derived statistically by processing typical exemplar signals *a priori*. To achieve further data compression, (D/S) pairings that are deemed to be ***similar*** in some way may be assigned the same symbol.

For most applications a standard TESPAR alphabet comprising 29 different symbols - see Figure 3 - has proven sufficient to represent the original waveform ***to a given approximation***. Holbeche [10] gives an account of one version of this coding procedure, for speech.

TESPAR Matrices

The output from a TESPAR coder is a stream of symbols, typically in the range 1 to 29, calculated from the D/S attributes of each epoch - see Figure 4.

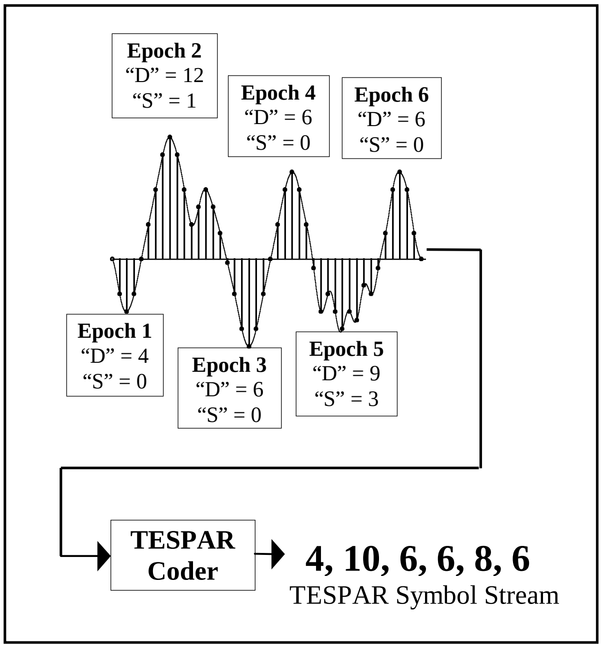

Figure 4 - Symbol Stream Produced By A TESPAR Coder based on Figure 3

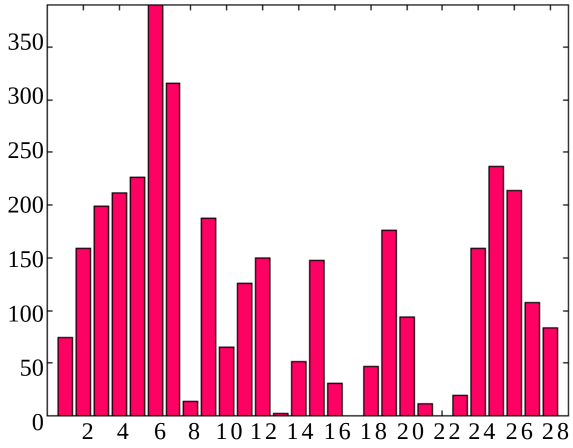

Figure 5 - TESPAR S-Matrix

This series of simple numerical descriptors may be readily converted into a variety of progressively informative fixed-dimension TESPAR matrices. For example, the ***S-Matrix*** is single-dimension 1x29 vector that records the frequency with which each TESPAR-coded symbol appears in the data stream - see Figure 5. A more discriminating data structure is the ***A-Matrix***. This is a two-dimensional 29x29 vector that records the number of times each pair of symbols in the alphabet appears *n* symbols apart. Employing the attribute *n* - known as the ***lag*** - allows temporal information to be represented in the matrix - see Figure 6. The use of small lags i.e. $n \leq 10$ tends to emphasise the short-term properties of the signal, whilst the use of larger lags tends to highlight those features which are more globally defined.

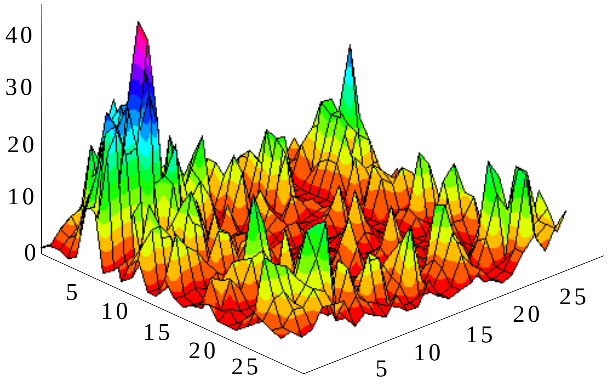

Figure 6 - TESPAR A-Matrix

Extending these ideas to three-dimensional vectors and beyond would obviously produce more informative TESPAR matrices. However, the considerable discriminating capabilities observed so far with comparatively simple S-matrices and A-matrices has meant that, to date, the use of such higher-order matrices has proven unnecessary.

TESPAR Classification Using Archetypes

A significant advantage of representing time-varying signals such as speech using TESPAR matrices - as opposed to using more conventional frequency domain descriptors - is that TESPAR matrices remain fixed in size regardless of the duration of the signal to be coded. In other words, even if the different conditions in a classification problem span variable time frames, all may be represented using TESPAR matrices of a common dimension. The use of fixed-size data structures greatly simplifies the process of generating exemplar templates at the training stage, and renders TESPAR matrices amenable to a wide range of comparison and classification procedures.

This fixed-sized nature of TESPAR matrices also opens up the prospect of using ***archetypes*** as a mechanism for increasing classification accuracy. An archetype is created by adding together, and then ***averaging***, matrices generated from several different versions of a particular condition. The averaging process tends to emphasise the consistent characteristics of the condition but reduces the significance of anomalies that may exist in the individual examples.

Archetypes for all conditions of interest may be generated and stored in a ***reference database***. At the classification stage, a new matrix is created ***live*** and compared, on an individual basis, with the archetypes in the reference database. Some form of statistical correlation can be used in the decision-making process to identify a ***winner***. Providing the winning score breaches a pre-defined ***acceptance threshold***, the identity of the live matrix is assumed to be the same as the identity of the highest-scoring archetype. Winning scores that fall below the acceptance threshold result in a ***not-recognised*** decision being given.

This process has been found to be surprisingly effective in separating and classifying signals, the spectrograms of which, may be identical in the frequency domain.

TESPAR Classification Using Artificial Neural Networks (ANN)

TESPAR matrices are ideally matched to the processing requirements of artificial neural networks (ANN) [11] for which the use of fixed-sized training and interrogation vectors is typically essential.

S-Matrices and A-Matrices have been successfully applied to supervised neural architectures - such as the ***Multi-Layer Perceptron*** (MLP) [12], and also to unsupervised neural architectures - such as the ***Kohonen Self-Organising Map*** [13].

The reliability of the final classification can usually be increased by combining several MLPs into a ***multiple network***. By changing the type of TESPAR matrix and/or varying the source of the signals used during training, each individual MLP in the multiple network will tend to base its classification decisions on a unique subset of properties of the training material and consequently exhibit a different classification behaviour. Adding MLPs that have been trained in this way to a multiple network will enhance the overall discriminating powers of the multiple network and give rise to an increase its classification accuracy.

At the classification stage, each individual MLP is interrogated separately to produce a set of ***individual decisions***. A final ***overall decision*** is obtained by polling the individual decisions and, applying to them, some sort of ***multiple decision logic***. The simplest multiple decision logic is a ***winner-takes-all*** strategy in which the overall decision is deemed to be the most common individual decision given. More sophisticated strategies include: weighting the individual decisions so that the most reliable/significant networks are given a greater status; ignoring individual network decisions that produce very low scores; and using *a priori* knowledge to form a higher-order decision logic that reflects any known relationships that may exist between subsets of the multiple network.

Massively Parallel Network Architecture (MPNA)

The size of the data needed to perform TESPAR-based signal classification has been found to compare most favourably with competitor systems, both in terms of cost, ease of implementation and overall system performance. However, a new architecture, called a Massively Parallel Network Architecture (MPNA) [14] has recently been developed primarily, but not exclusively, for use in the field of speaker verification. The principal strengths of the MPNA are:

- It embodies the immense power of massively parallel networks and data fusion to achieve the classification accuracy normally associated with a very large number *N* of networks.

- The reference template for each condition is very small, typically occupying circa. 40 bytes of data, irrespective of the size, dimensionality and complexity of the input data matrices.

- Once the MPNA has been created, no further network training is required. This helps keep the overall registration time to a minimum.

To construct an MPNA, an ordered set of N networks, (typically $100 \leq N \leq 1500$), all of which are different in some way, may be trained in non-real time using a database of exemplar signals collected *a priori*. Assuming that each network is trained using a subset of 8 different signals, this provides an MLP with 8 output nodes in the output layer. These N networks are then used as an interrogation set, against which all live signals are to be compared, both at registration and interrogation.

At registration, several examples of the first signal class to be learned, i.e. Condition 1, will be converted appropriately into TESPAR matrices, and compared with each of the N networks in turn. Each of the networks will give an individual decision indicating to which of its 8 output nodes the TESPAR matrix was closest. This decision is expressed as a winning node whose index, in this example, will be in the range 0 to 7 and whose identity can therefore be expressed in 3 bits. Assuming N = 100, then the profile of 100 individual network decisions can be expressed in 100 x 3 bits = 300 bits, i.e. circa. 40 bytes. Once all examples of Condition 1 have been processed, a modal averaging process is used to determine the numerical profile of winners that is likely to be generated by the Condition 1 during subsequent interrogations. This 40-byte profile of winners is then used as the reference template for Condition 1. Templates are subsequently produced for the remaining conditions of interest in a likewise manner and all templates are added to a reference database.

At the interrogation stage a profile for the live signal is produced and, in some way, compared with each of the templates in the reference database. As with the archetype classification, some form of statistical correlation can be used measure a degree of closeness and hence identify a ***winner***.

TESPAR At Work

TESPAR has been successfully applied to signals ranging from very-high frequency nanosecond waveforms associated with power transformers [15] to very-low frequency waveforms of a few fractions of a Hertz resulting from structural oscillations in bridges [16]. However, a significant work programme has also been directed towards deploying TESPAR in speech applications such as speech recognition and speaker verification.

A proprietary procedure, known as Fast Artificial Neural Networks (FANN), has also been developed for training ANNs quickly, i.e. in less than a few minutes, and storing them in relatively small memory sizes, i.e. in under 1 kbyte.

A TESPAR/FANN combination has been used to implement a very successful speaker verification capability [17]. In extensive trials, consuming some 16 man-months of work, a database comprising 150 male and 68 female speakers in which each speaker spoke the common phrase “Sir Winston Churchill” on 20 separate occasions, was used to test the system. Each utterance was recorded within a fixed 3-second window against a variety of ambient background noises. A multiple network, comprising 15 individual 6-output FANNs, trained using S-matrices, was constructed for each of the 218 speakers. Using unsupervised registration procedures 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%)

It should be noted though, that these results were obtained despite the fact that some 8% of the FANNs used did not fully converge. In addition, no FAR-reduction strategies were employed during registration.

A pilot study into the effects of commonly occurring ***benign traumas*** on system performance has also been conducted. The results indicate that potential pitfalls such as: effects of influenza, changes in speech that might occur over long periods of time, drunkenness (!) etc. need not adversely affect the accuracy of the TESPAR/FANN combination.

TESPAR/FANN was the chosen biometric technology in the European Union CASCADE Esprit Smart Card project. The outcome of this project was the successful development, implementation and subsequent demonstration of a 32-bit RISC processor, 20 square mm in area, with a voice biometric capability, for deployment in the next generation of smart cards and Pocket Intelligent Device applications [18].

Conclusions

The TESPAR/FANN combination is a powerful, robust, flexible and economic technology that is suitable for application to a wide range of speech processing tasks such as speaker verification and speech recognition. Significant trials show exceptionally low equal-error rates (EER) when compared with currently-reported conventional methodologies. The TESPAR/FANN procedures described permit system errors to be made vanishingly small by design over a wide range of speaker verification and speech recognition scenarios.

In many cases TESPAR architectures can normally be embodied so cheaply in DSP devices that only the spare capacity of the processor board is required for implementation. For example, a real-time TESPAR/FANN phoneme recognition system has already been embodied using the 4-kbyte fast external memory of a DSP32C. The possibility also exists for transferring this architecture to the processor of an inexpensive high-volume commercial A-D card such as a SoundBlaster-compatible PC audio card.

The MPNA is a new and exciting TESPAR-based methodology that has the potential to offer a very robust speaker verification performance using a speech biometric of only 40 bytes without requiring additional network training during registration.

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 TESPAR/FANN has been developed.
- The principal of Cranfield University (RMCS) for his permission to publish this paper.
- Ashley Elkins for his advice and assistance during the preparation of this paper.

References

[1] C. E. Shannon, “A Mathematical Theory Of Communication”, Bell Syst. Tech. J., vol. 27, pp. 379-423, July 1948.

[2] C. E. Shannon, “*Communication In The Presence Of Noise”*, Proc. IRE, vol. 37, pp. 10-21, Jan. 1949.

[3] J. C. R. Licklidder, I. Pollack, *“Effects of Differentiation, Integration, and Infinite Peak Clipping Upon The Intelligibility Of Speech”*, Journal Of The Acoustical Society Of America, vol. 20, no. 1, pp. 42-51, Jan. 1948.

[4] P. C. Hammer, *“Approximation Spaces”*, Advances In Mathematical Systems Theory, The Pennsylvania State University Press, p122, p132, 1969.

[5] F. E. Bond, C. R. Cahn, *“A Relationship Between Zero Crossings And Fourier Coefficients For Bandwidth-Limited Functions”*, IRE Trans. Information Theory, vol. IT-4, pp. 110-113, Sept. 1958.

[6] E. C. Titchmarsh, *“The Zeros Of Certain Integral Functions”*, Proc. progres. Math. Soc., vol. 25, pp. 283-302, May 1926.

[7] H. B. Voelcker, *“Toward A Unified Theory Of Modulation Part 1: Phase-Envelope Relationships”*, Proc. IEEE, vol. 54, no. 3, pp 340-353, March 1966.

[8] H. B. Voelcker, *“Toward A Unified Theory Of Modulation Part 2: Zero Manipulation”*, Proc. IEEE, vol. 54, no. 5, pp. 735-755, March 1966.

[9] A. A. G. Requicha, *“The zeros of entire functions, theory and engineering applications”*. Proceedings of the IEEE, vol. 68 no. 3, pp. 308-328, March 1980.

[10] J. Holbeche, R. D. Hughes and R. A. King, *“Time Encoded Speech (TES) descriptors as a symbol feature set for voice recognition systems”*, IEE Int. Conf. Speech Input/Output; Techniques and Applications, pp. 310-315, March 1986.

[11] S. R. Timms and R. A. King, *“Speaker Verification Routines For ISDN and UPT Access and Security Using Artificial Neural Networks And Time Encoded Speech (TES) Data.”*, IEE Int. Conf. Private Switching Systems And Networks, pp. 59-64, June 1992.

[12] M. H. George, and R. A. King, *“Time Encoded Signal Processing And Recognition For Reduced Data, High Performance Speaker Verification Architectures”*, 1st Int. Conf. Audio And Video-Based Biometric Person Authentication (AVBA 97), pp. 377-384, March 1997.

[13] T. Kohonen, *“Self-Organising Maps”*, Springer Series In Information Sciences, Second Edition, 1997.

[14] T. C. Phipps, and R. A. King, “A Speaker Verification Biometric In 40 Bytes”, CardTech/SecurTech 1997, vol. 1. pp-187-198, May 1997.

[15] 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, no. 6, December 1993.

[16] Domain Dynamics Limited, “*Application Of TESPAR In Identifying Structural Changes In Bridges*”, Internal Report For W. S. Atkins Science And Technology, 1996.

[17] R. A. King, *“TESPAR/FANN: an effective new capability for voice verification in the defence environment”*, Royal Aeronautical Soc. Conf. On The Role Of Intelligent Systems In Defence, pp. 5.1-5.8, 1995

[18] CASCADE Esprit Project EP8670 Data Sheet, 1995.