EURASIP NEWSLETTER

ISSN 1687-1421, Volume 14, Number 2, June 2003

European Association for Signal, Speech, and Image Processing

Newsletter, Volume 14, Number 2, June 2003

Contents

President’s Message

Dear EURASIP members,

As you may have noticed, the Newsletter issue you are reading at this moment has changed with respect to the latest ones. This change is not only a formal one but it corresponds to the intention of giving a better service to our members through the Newsletter as well. The EURASIP AdCom estimated that this task would require an extra effort and therefore there was a call for applications to the position of EURASIP Newsletter coeditor. This call was publicized through the previous Newsletter issue, the EURASIP website, and direct mailing.

As a result of this call, we received sixty-two applications. First of all, I would like to thank all the people that expressed their interest in collaborating with EURASIP. The AdCom will take into account your interest and we hope to be able to cooperate with as many as possible of you in the near future in other EURASIP projects. Second, it is my pleasure to announce that, after analyzing all the applications, the AdCom contacted Luca Simone Ronga who accepted the position of EURASIP Newsletter coeditor. I would like to publicly welcome Luca to the AdCom and thank him for his commitment with EURASIP.

Finally, as I already commented in my previous message, the structure and contents of the Newsletter is currently being redesigned. If you have any suggestions to improve its usefulness, please send them to our Newsletter coeditors: Piet Sommen (Piet.Sommen@eurasip.org) and/or Luca S. Ronga (Luca.Ronga@eurasip.org).

Best regards,
Ferran Marqués

Call for Tutorial Papers

In an effort to upgrade the scientific content of the Newsletter, the EURASIP AdCom decided to include a tutorial paper in each Newsletter issue.

The Newsletter is being distributed to approximately 1500 EURASIP members, so it is envisaged that a tutorial paper will spread the author's views to a rather wide audience.

Prospective authors are invited to submit manuscripts (2–6 pages) on topics that are of a general interest to the Signal Processing community. The papers should be written in a style, also, readable by a nonspecialist reader.

The submitted papers will go through a reviewing process in order to guarantee a high scientific quality.

For additional information, please contact our Newsletter editor.

EURASIP Secretary-Treasurer's Report

Opening balance of **172,587.76€** was held as follows on 1st April 2002:

CHF cheque account (CHF 113,228.39) . . . . 76,734.87€
Euro cheque account . . . . 41,463.14€
Euro money market account . . . . 51,051.34€
Postal account on 31/12/2001 (CHF 4,926.09) . . . . 3,338.41€

The EURASIP main account movements during the financial period from 1st April 2002 until 31st December 2002 are documented in the table below in Euros (€).

Incoming balance of 1st April 2002			**172,587.76€**
Income:	Membership fees includes Journal subscriptions	23,732.71€	
	Nokia Awards	1,000.00€	
	Interests	1,078.64€	
Total:		25,811.35€	
Expenses:	Elsevier (various concepts)	33,328.67€	
	Hindawi (various concepts)	972.86€	
	Newsletter	2,694.20€	
	Awards	1,500.00€	
	Administrative expenses	9,163.4€	
	Taxes	364.72€	
Total:		48,023.85€	
Outgoing balance on 31st Dec. 2002 = **Incoming balance** on 1st Jan. 2003			**150,375.26€**

Total balances at 1st January 2003 in EURO:

CHF cheque account (CHF 114,615.80) . . . . 78,990.92€
Euro cheque account . . . . 13,593.81€
Euro money market account . . . . 51,837.36€
Postal account (CHF 8,638.05) . . . . 5,953.17€

Total in EURO . . . . **150,375.26€**

EURASIP's finances are looking favorable, as can be verified from the previous balances and accounts as given in the Newsletters.

Calendar of Events

Year	Date	Event	Location	EURASIP Involvement	Chairperson/Information
	June 8–11	Nonlinear Signal and Image Processing (NSIP-03)	Grado, Italy	Cooperation	G. Sicuranza and G. Vernazza http://ipl.univ.trieste.it/nsip03
	July 1–4	7th Int. Symposium on Signal Processing and its applications (ISSPA)	Paris, France	Cooperation	B. Boashash and A. Beghdadi http://isspa2003.univ-paris13.fr/
	July 2–4	4th EURASIP Conf. on Video/Image Processing and Multimedia Communications (EC-VIP-MC-2003)	Zagreb, Croatia	Sponsor	Branka Zovko-Cilhar http://www.vcl.fer.hr/ec2003/
	July 8–11	Visual Communications and Image Processing 2003	Lugano, Switzerland	Cooperation	T. Ebrahimi and T. Sikora www.vcip2003.ch
	September 2–5	Advanced Concepts for Intelligent Systems (ACIVS-2003)	Ghent, Belgium	Cooperation	J. Blanc-Talon http://eltodo.rug.ac.be/acivs2003/
	September 8–11	6th Int. Conference on Digital Audio Effects	London, UK	Cooperation	Mark Sandler http://www.elec.qmul.ac.uk/dafx03
	September 18–20	3rd Int. Symposium on Image and signal Processing and Analysis	Rome, Italy	Cooperation	A. Neri and H. Babic http://www.isispa.org
	September 18–19	International Workshop VLBV03	Madrid, Spain	Cooperation	Narciso Garcia http://vlbv03.upm.es
	September 22–23	IEE Colloquium on DSP Enabled Radio	Livingstone, Scotland, UK	Cooperation	Bob Stewart http://www.eee.strath.ac.uk/r.w.stewart/dsp_radio.pdf
	October 9–10	1st Int. Workshop on Interactive Rich Media Content Production (WS2003)	Lausanne, Switzerland	Cooperation	I. Pitas and D. Thalmann http://www.richmedia2003.org
	December 11–12	ISCA Workshop on Multimodal User Authentication	Santa Barbara, USA	Cooperation	Jean-Luc Dugelay authentication@research.panasonic.com
2004	June 23–25	17th Int. EURASIP Conf. BIOSIGNAL	Brno, Czech Republic	Co-sponsorship	Jiri Jan http://www.feec.vutbr.cz/UBMI/bs2004.html
	July 20–22	4th CSNDSP, Int. Symposium on Communication Systems, Networks and DSP	Newcastle, UK	Cooperation	T. Boukouvalas http://www.shu.ac.uk/ocr/csndsp/

Sergios Theodoridis
Workshops/Confs Coordinator EURASIP

Report on the Fourth International Workshop on Image Analysis for Emerging Interactive Services (WIAMIS03)

The Fourth International Workshop on Image Analysis for Emerging Interactive Services WIAMIS03 was organized by the Multimedia and Vision research group at Queen Mary, University of London during the 9th, 10th, and 11th of April 2003. The workshop featured three special sessions, two keynote talks, and almost one hundred presentations of works from prominent researchers and institutions around the world. The special sessions truly reflected the way WIAMIS has been growing in parallel to the technology it intends to embrace. The first special session was dedicated to Content-Based Semantic Scene Analysis and Information Retrieval. It was organized by Michael Strintzis from Aritoteles University of Thessaloniki. This session was the largest in the conference and the proceedings with over 20 papers. It also reflects the most recent trend in multimedia processing and analysis. The second special session was dedicated to the presentation of large European projects in the Information Society Technology (IST) area. This session was strongly related to the activities of the European Network of Excellence SCHEMA and was coordinated by Yiannis Kompatsiaris from the Informatics and Telematics Institute, Thessaloniki, Greece. The IST session began with the presentation of the SCHEMA and BUSMAN projects, followed by ten additional contributions describing other IST projects. The video activities as traditional focus of WIAMIS were complemented this time by a special session on Audio Segmentation and Digital Music. This complementarity has strengthened the scope of WIAMIS so that in 2003 it covered the full range of multimedia. This session was coordinated by Mark Sandler from Queen Mary University of London.

The first keynote talk was given by Thomas Huang from the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign. In this work an approach for relevance feedback in content-based image and video retrieval was introduced. The second invited talk was given by Aggelos K. Katsaggelos from the Department of Electrical and Computer Engineering, Northwestern University. He described research on speech-to-video synthesis using MPEG-4 compliant visual features. In addition to the papers submitted to the special sessions, there were many contributions to other specific areas of research in multimedia analysis. All contributions were peer-reviewed by a panel of experts and members of the Workshop Technical Committee. From over one hundred initial submissions, approximately 85% were accepted. These papers will form the regular oral and poster sessions in WIAMIS03.

The WIAMIS steering committee met on the 10th of April. The committee was enlarged with four new members: Prof. Thomas Huang, Prof. Aggelos K. Katsaggelos, Prof. Hyoung-Joong Kim, all from the Department of Control Instrumentation Engineering, Kangwon National University, and Prof. Woontack Woo from the Kwangju Institute of Science and

Technology, Korea. It was agreed to change the workshop frequency from biannual to annual. In 2004 WIAMIS will be organized in Lisbon (subject to confirmation) or Berlin. WIAMIS05 will be held for the first time outside Europe, in Chicago, USA.

Ebroul Izquierdo
General Chairman WIAMIS '03

A Brief Introduction to Nonlinear Speech Analysis and Synthesis

Stephen McLaughlin

Signals and Systems Group, School of Engineering and Electronics, University of Edinburgh, The King's Buildings, Edinburgh, EH9 3JL, Scotland, UK
Email: sml@ee.ed.ac.uk, Tel.: (+44)-131-650-5578, Fax: (+44)-131-650-6554

1. Introduction

This paper tries to offer a brief introduction to nonlinear speech analysis and synthesis. It is far from a complete overview and there are many researchers worldwide looking into this area. For a list of European research in this area, see the website of COST 277 (http://www.see.ed.ac.uk/~cost277). The paper examines the how and the why of nonlinear speech analysis and synthesis from my personal perspective. It discusses why a nonlinear approach to speech analysis and synthesis should be considered, reviews the recent history, and describes a variety of approaches to the problem. It argues that while modern concatenative speech synthesisers produce speech which is intelligible, they are inflexible and often lack a human quality. The paper does not suggest that nonlinear speech synthesisers are ready to replace conventional approaches, but rather that they offer some potential advantages that there is a considerable amount of research still to be carried out.

Speech synthesis is a complex task that aims to produce naturally sounding speech. While working systems that produce intelligible speech have existed since the 1970s, the final aim of producing a synthesiser that is indistinguishable from a human speaker has still to be realised. There remain a number of problems at all stages of the process, including the actual generation of the speech signal itself with the required intonation. This paper is structured as follows, a brief review of conventional linear-based approaches is followed by a quick review of nonlinearities which exist in speech generation. Then an example of nonlinear techniques applied to epoch marking is presented, followed by two sections on nonlinear speech synthesis. Finally some conclusions are drawn.

2. Conventional speech synthesis approaches

Conventionally the main approaches to speech synthesis depend on the type of modelling used. This may be a model of the speech organs themselves (articulatory synthesis), a model derived from the speech signal (waveform synthesis), or alternatively the use of prerecorded segments extracted from a database and joined together (concatenative synthesis).

Modelling the actual speech organs is an attractive approach since it can be regarded as being a model of the fundamental level of speech production. An accurate articulatory model would allow all types of speech to be synthesised in a natural manner without having to make many of the assumptions required by other techniques (such as attempting to separate the source and vocal tract parts out from one signal). See [1, 2, 3].

Waveform synthesisers derive a model from the speech signal as opposed to the speech organs. This approach is derived from the linear source-filter theory of speech production [4]. The simplest form of waveform synthesis is based on linear prediction (LP) [5].

Formant synthesis uses a bank of filters, each of which represents the contribution of one of the formants. The best known formant synthesiser is the Klatt synthesiser [6], which has been exploited commercially as DECTalk.

Concatenation methods involve joining together prerecorded units of speech which are extracted from a database. It must also be possible to change the prosody of the units so as to impose the prosody required for the phrase that is being generated. The concatenation technique provides the best quality synthesised speech available at present. It is used in a large number of commercial systems, including British Telecomm's Laureate [7] and the AT&T Next-Gen system [8].

Techniques for time and pitch scaling of sounds held in a database are also extremely important. Two main techniques for time scale and pitch modification in concatenative synthesis can be identified, each of which operates on the speech signal in a different manner. The pitch synchronous overlap add (PSOLA) [9] approach is nonparametric as opposed to the harmonic method, which actually decomposes the signal into explicit source and vocal tract models.

McAulay and Quatieri developed a speech generation model that is based on a glottal excitation signal made up of a sum of sine waves [10]. They then used this model to perform time scale and pitch modification. A limitation of all these techniques is that they use the linear model of speech as a basis.

3. Nonlinearities in speech

There are known to be a number of nonlinear effects in the speech production process. Firstly, it has been accepted for some time that the vocal tract and the vocal folds do not function independently of each other, but that there is in fact some form of coupling between them when the glottis is open [11], resulting in significant changes in formant characteristics between open and closed glottis cycles [12]. More controversially, H. M. Teager and S. M. Teager [13] have claimed (based on physical measurements) that voiced sounds are characterised by highly complex air flows in the vocal tract involving jets and vortices, rather than well-behaved laminar flow. In addition, the vocal folds will themselves be responsible for further nonlinear behaviour since the muscle and cartilage which comprise the larynx have nonlinear stretching qualities. Such nonlinearities are routinely included in attempts to model the physical process of vocal fold vibration, which have focussed on two or more mass models [2, 3, 14], in which the movement of the vocal folds is modelled by masses connected by springs with nonlinear coupling. Observations of the glottal waveform have shown that this waveform can change shape at different amplitudes [15] which would not be possible in a strictly linear system, where the waveform shape is unaffected by amplitude changes.

In order to arrive at the simplified linear model, a number of major assumptions are made:

1. the vocal tract and speech source are uncoupled (thus allowing source-filter separation);
2. airflow through the vocal tract is laminar;
3. the vocal folds vibrate in an exactly periodic manner during voiced speech production;
4. the configuration of the vocal tract will only change slowly.

These imply a loss of information which means that the full speech signal dynamics can never be properly captured. These inadequacies can be seen in practice in speech synthesis where, at the waveform generation level, current systems tend to produce an output signal that lacks naturalness. This is true even of concatenation techniques which copy and modify actual speech segments.

4. Poincaré maps and epoch marking

The section discusses how nonlinear techniques can be applied to pitch marking of continuous speech. We wish to locate the instants in the time domain speech signal at which the glottis is closed. A variety of existing methods can be employed to locate the epochs. These are abrupt change detection [16], maximum likelihood epoch detection [17], and dynamic programming [18]. All of the above techniques are sound and generally provide good epoch detection. The technique presented here should not be viewed as a direct competitor to the methods outlined above. Rather it is an attempt to show the practical application of ideas from nonlinear dynamical theory to a real speech processing problem. The performance in clean speech is comparable to many of the techniques discussed above.

In nonlinear processing a d-dimensional system can be reconstructed in an m-dimensional state space from a single dimension time series by a process called embedding. Takens' theorem [19] states that $m \geq 2d + 1$ for an adequate reconstruction although in practice it is often possible to reduce m. An alternative is the singular value decomposition (SVD) embedding [20], which may be more attractive in real systems where noise is an issue.

A Poincaré map is often used in the analysis of dynamical systems. It replaces the flow of an nth order continuous system with an $(n - 1)$th order discrete time map. Considering a three-dimensional attractor, a Poincaré section slices through the flow of trajectories and the resulting crossings form the Poincaré map. Reexamining the attractor reconstructions of voiced speech shown above, it is evident that these three-dimensional attractors can also be reduced to two-dimensional maps.[1] Additionally, these reconstructions are pitch-synchronous, in that one revolution of the attractor is equivalent to one pitch period. This has previously been used for cyclostationary analysis and synchronisation [21]; here we examine its use for epoch marking.

The basic processing steps required for a waveform of N points are as follows:

1. mark y_{GCI}, a known GCI in the signal;
2. perform an SVD embedding on the signal to generate the attractor reconstruction in 3D state space;

[1] Strictly these attractor reconstructions are discrete time maps and not continuous flows. However it is possible to construct a flow vector between points and use this for the Poincaré section calculation.

3. calculate the flow vector $\mathbf{h}$ at the marked point $\mathbf{y}_{\mathrm{GCI}}$ on the attractor;
4. detect crossings of the Poincaré section Σ at this point in state space by signs changes of the scalar product between $\mathbf{h}$ and the vector $\mathbf{y}_i - \mathbf{y}_{\mathrm{GCI}}$ for all $1 \leq i \leq N$ points;
5. those points on Σ which are within the same portion of the manifold as $\mathbf{y}_{\mathrm{GCI}}$ are the epochs.

When dealing with real speech signals a number of practical issues have to be considered. The input signal must be treated on a frame-by-frame basis, within which the speech is assumed stationary. Finding the correct intersection points on the Poincaré section is also a difficult task due to the complicated structure of the attractor.

In Figure 2, which is a voiced section from the phrase "see if it's raining" spoken by a male speaker, the epochs are well located for the first part of the signal, but some slight loss of synchronisation can be seen in the latter part.

5. Nonlinear synthesis approaches

5.1. Neural network synthesis background

Kubin and Birgmeier reported an attempt made to use a RBF network approach to speech synthesis. They propose the use of a nonlinear oscillator, with no external input and global feedback, in order to perform the mapping

$$x(n) = \mathscr{A}(\mathbf{x}(n-1)), \tag{1}$$

where $\mathbf{x}(n-1)$ is the delay vector with nonunit delays and $\mathscr{A}$ is the nonlinear mapping function [22].

The initial approach [23] used a Kalman-based RBF network, which has all of the network parameters trained by the extended Kalman filter algorithm. The only parameter that must be specified is the number of centres to use. This gives good prediction results, but there are many problems with resynthesis. In particular, they report that extensive manual fine-tuning of the parameters such as dimension, embedding delay and number, and initial positions of the centres are required. Even with this tuning, synthesis of some sounds with complicated phase space reconstructions does not work [22].

In order to overcome this problem, Kubin resorted to a technique that uses all of the data points in the training data frame as centres [22]. Although this gives correct resynthesis, even allowing the resynthesis of continuous speech using a frame-adaptive approach, it is unsatisfactory due to the very large number of varying parameters, and cannot be seen as actually learning the dynamics of the speech generating system.

An alternative neural network approach was proposed by Narashimhan *et al.* [24]. This involves separating the voiced source from the vocal tract contribution, and then creating a nonlinear dynamical model of the source. This is achieved by first-inverse filtering the speech signal to obtain the (LP) residual. Next the residue waveform is lowpass filtered at 1 kHz, then normalised to give a unit amplitude envelope. This processed signal is used as the training data in a time delay neural network with global feedback. The NN structure reported is extremely complex, consisting of a 30 tap delay line input and two hidden layers of 15 and 10 sigmoid activation functions, with the network training performed using back propagation through time. Finally, the NN model is used in free-running synthesis mode to recreate the voiced source. This is applied to an LP filter in order to synthesise speech. They show that the NN model successfully preserves the jitter of the original excitation signal.

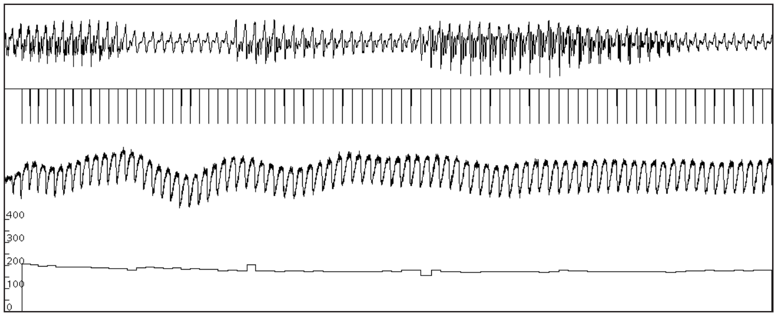

Figure 1: Results for the voiced section of "came along" from the Keele database for a female speaker. From top to bottom: the signal, the epochs as calculated by the algorithm, the laryngograph signal, and the pitch contour (Hz) resulting from the algorithm.

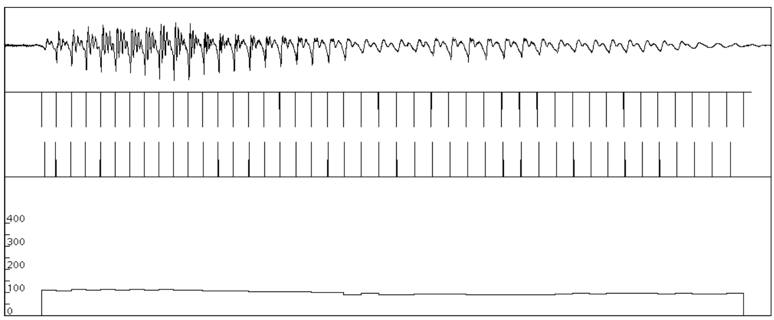

Figure 2: Results for the voiced section of "raining" from the BT Labs database for a male speaker. From top to bottom: the signal, the epochs as calculated by the algorithm, the processed laryngograph signal, and the pitch contour (Hz) resulting from the algorithm.

5.2. RBF network for synthesis

A well-known nonlinear modelling approach is the radial basis function neural network. It is generally composed of three layers, made up of an input layer of source nodes, a nonlinear hidden layer, and an output layer giving the network response. The hidden layer performs a nonlinear transformation, mapping the input space to a new space, in which the problem can be better solved. The output is the result of linearly combining the hidden space, multiplying each hidden layer output by a weight whose value is determined during the training process.

The general equation of an RBF network with an input vector $\mathbf{x}$ and a single output is

$$\mathcal{F}(\mathbf{x}(n)) = \sum_{j=1}^{P} w_j \phi(\|\mathbf{x} - \mathbf{c}_j\|), \qquad (2)$$

where there are P hidden units, each of which is weighted by w_j. The hidden units $\phi(\|\mathbf{x} - \mathbf{c}_j\|)$ are radially symmetric functions about the point $\mathbf{c}_j$, called a centre, in the hidden space, with $\|\cdot\|$ being the Euclidean vector norm [25]. The actual choice of nonlinearity does not appear to be crucial to the performance of the network. There are two distinct strategies for training an RBF network. The most common approach divides the problem into two steps. Firstly, the centre positions and bandwidths are fixed using an unsupervised approach, not dependent on the network output. Then the weights are trained in a supervised manner so as to minimise an error function.

Following from the work of Kubin *et al.*, a nonlinear oscillator structure is used. The RBF network is used to approximate the underlying nonlinear dynamics of a particular stationary voiced sound by training it to perform the prediction

$$x_{i+1} = \mathcal{F}(\mathbf{x}_i), \tag{3}$$

where $\mathbf{x}_i = \{x_i, x_{(i-\tau)}, \ldots, x_{(i-(m-1)\tau)}\}$ is a vector of previous inputs spaced by some delay τ samples, and $\mathcal{F}$ is a nonlinear mapping function. From a nonlinear dynamical theory perspective, this can be viewed as a time-delay embedding of the speech signal into an m-dimensional state space to produce a state space reconstruction of the original d-dimensional system attractor. The embedding dimension is chosen in accordance with Takens' embedding theorem [19], and the embedding delay τ is chosen as the first minimum of the average mutual information function [26]. The other parameters that must be chosen are the bandwidth, the number and position of the centres, and the length of training data to be used. With these set, the determination of the weights is linear in the parameters and is solved by minimising a sum of squares error function $E_S(\hat{\mathcal{F}})$ over the N samples of training data:

$$E_s(\hat{\mathcal{F}}) = \frac{1}{2}\sum_{i=1}^{N}(\hat{x}_i - x_i)^2, \tag{4}$$

where $\hat{x}_i$ is the network approximation of the actual speech signal x_i. Incorporating equation (2) into the above and differentiating with respect to the weights, then setting the derivative equal to zero gives the least squares problem [27], which can be written in a matrix form as

$$(\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{w}^T = \mathbf{\Phi}^T\mathbf{x}, \tag{5}$$

where $\mathbf{\Phi}$ is an $N \times P$ matrix of the outputs of the centres, $\mathbf{x}$ is the target vector of length N, and $\mathbf{w}$ is the P length vector of weights. This can be solved by standard matrix inversion techniques.

Two types of centre positioning strategy can be considered.

1. Data subset. Centres are picked as points from around the state space reconstruction. They are chosen pseudo randomly so as to give an approximately uniform spacing of centres about the state space reconstruction.
2. Hyper-lattice. An alternative data independent approach is to spread the centres uniformly over an m-dimensional hyper-lattice.

5.3. Synthesis

For each vowel in the database, the weights were learnt, with the centres either on a 7D hyper-lattice, or chosen as a subset of the training data. The global feedback loop was then

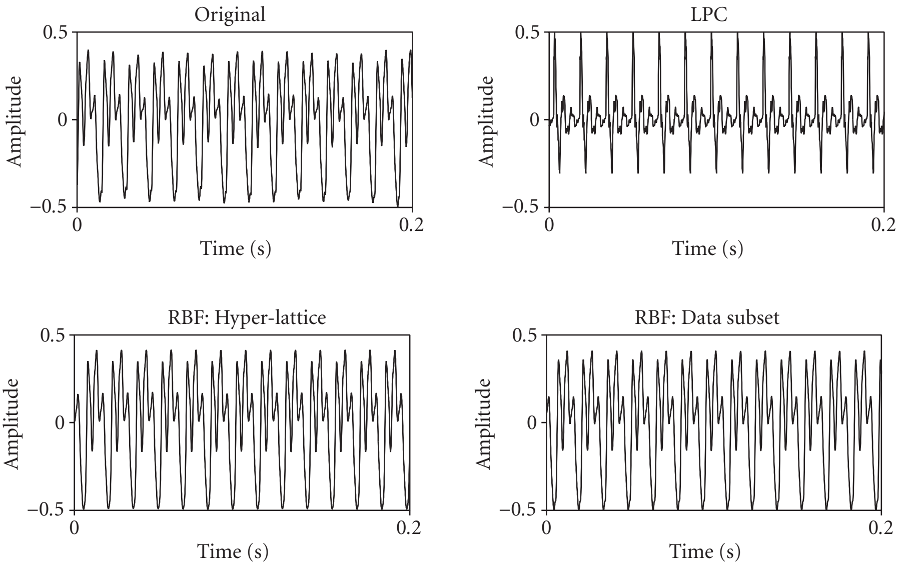

FIGURE 3: Time domain examples of the vowel /u/, speaker MC. Top row: original signal (left) and linear prediction synthesised signal (right). Bottom row: RBF network synthesised signal: hyper-lattice (left) and data subset (right).

put in place to allow free-running synthesis. The results gave varying degrees of success, from constant (sometimes zero) outputs, through periodic cycles not resembling the original speech signal and noise-like signals, to extremely large spikes at irregular intervals on otherwise correct waveforms [28].

This could have been due to a lack of smoothness in the function, in which case regularisation theory was the ideal solution. Regularisation theory applies some prior knowledge, or constraints, to the mapping function to make a well-posed problem [29].

The selection of an appropriate value for the regularisation parameter λ is done by the use of cross-validation [27]. After choosing all the other network parameters, these are held constant and λ is varied. For each value of λ, the MSE on an unseen validation set is calculated. The MSE curve should have a minimum indicating the best value of λ for generalisation. With the regularisation parameter chosen by this method, the 7D resynthesis gave correct results for all of the signals except KH /i/ and KH /u/ when using the data subset method of centre selection. However, only two signals (CA /i/ and MC /i/) were correctly resynthesised by the hyper-lattice method. It was found that λ needed to be increased significantly to ensure correct resynthesis for all the signals when the hyper-lattice was used. Achieving stable resynthesis inevitably comes at some cost. By forcing smoothness onto the approximated function, there is the risk that some of the finer detail of the state space reconstruction will be lost. Therefore, for best results, λ should be set at the smallest possible value that allows stable resynthesis.

Figure 3 shows the time domain waveforms for the original signal, the LP synthesised signal, and the two RBF synthesised signals, for the vowel /u/, speaker MC. Figure 4 shows

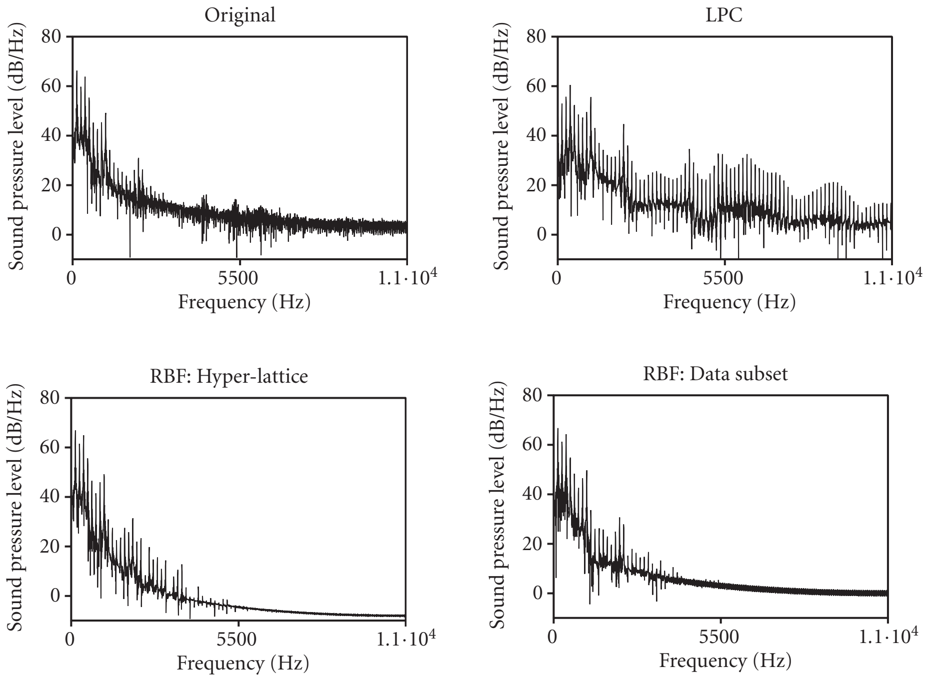

Figure 4: Spectrums for examples of the vowel /u/, corresponding to the signals in Figure 3.

the corresponding frequency domain plots of the signals. In these examples, the regularisation parameter λ was set at 0.01 for the hyper-lattice, and 0.005 for the data subset. In the linear prediction case, the technique attempts to model the spectral features of the original, although the higher frequencies have been over-emphasised. Note also that the time-domain signal bears little resemblance to the original signal. The RBF techniques, on the other hand, resemble the original in the time domain, since it is from this that the state space reconstruction is formed, although the spectral plots show the higher frequencies have not been well modelled by this method. This is because the networks have missed some of the very fine variations of the original time domain waveform, which may be due to the regularisation.

Further spectrogram examples for different vowels and speakers follow the same pattern, with the size of λ being seen to influence the quality of the signal at high frequencies.

The approach adopted here is to model the vocal tract as a forced nonlinear oscillator and to embed an observed scalar time-series of a vowel with pitch information into a higher dimensional space. This embedding, when carried out correctly, will reconstruct the data onto a higher dimensional surface which embodies the dynamics of the vocal tract (see, for example, [30, 31] for issues regarding embedding).

Previous studies, discussed above, have successfully modelled stationary (i.e. constant pitch) vowel sounds using nonlinear methods, but these have very limited use since the pitch cannot be modified to include prosody information. The new approach described

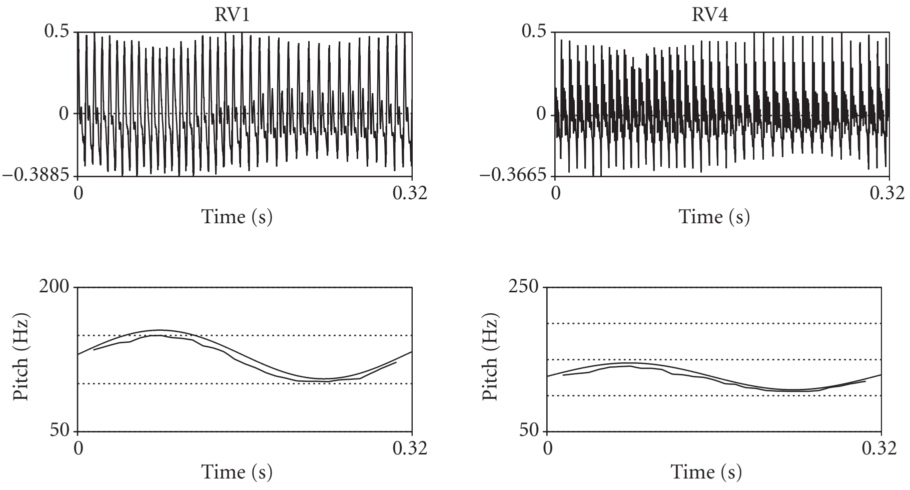

Figure 5: Synthesised vowel sounds together with desired and measured pitch profiles.

here resolves this problem by including pitch information in the embedding. Specifically, a nonstationary vowel sound is extracted from a database and, using standard pitch extraction techniques, a pitch contour is calculated for the time series so that each time domain sample has an associated pitch value. In the present study, measurements of rising pitch vowel sounds, where the pitch rises through the length of the time series, have been used as the basis for modelling.

The time series is then embedded in an m-dimensional space, along with the pitch contour, to form an $(m + 1)$-dimensional surface. A mixed embedding delay between time samples (greater than unity) is used to capture the variable time scales present in the vowel waveform. The $(m + 1)$-dimensional surface is modelled by a nearest neighbour approach, which predicts the next time series sample given a vector of previous time samples and a pitch value (it is envisaged that more sophisticated modelling techniques will be incorporated at a later date).

Synthesis is then performed by a modification of the nonlinear oscillator approach [32], whereby the input signal is removed and the delayed synthesiser output is fed back to form the next input sample. In contrast to previous techniques, the required pitch contour is also passed into the model as an external forcing input. Our results as shown in Figures 5 and 6, show that this method allows the vowel sound to be generated correctly for arbitrary specified pitch contours (within the input range of pitch values) even though the training data is only made up of the rising vowel time series and its associated pitch contour. In addition, sounds of arbitrary duration can be readily synthesised by simply running the oscillator for the required length of time. Typical synthesis results are shown. It can be seen that the sinusoidal pitch contour of the synthesised sound is quite different from the rising pitch profile of the measured data; the duration of the synthesised data is also somewhat longer than that of the measured data. The small offset evident between desired and synthesised pitch contours is attributed to minor calibration error. The initial results presented here are encouraging.

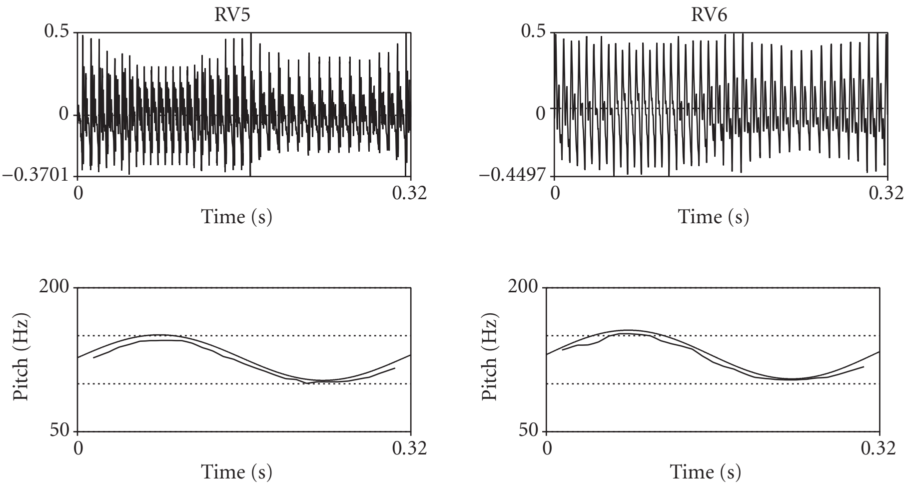

Figure 6: Synthesised vowel sounds together with desired and measured pitch profiles.

6. Conclusions

In view of these observations, it seems likely that the data-based model of the vowel dynamics possesses an important degree of structure, perhaps reflecting physiological considerations, that requires further investigation. It is also clear that whilst encouraging there is still some way to go in overcoming the limitations of the approach. It is clear that speech is a nonlinear process and that if we are to achieve the holy grail of truly natural sounding synthetic speech that this must be accounted for. It is also clear that nonlinear synthesis techniques offer some potential to achieve this although a great deal of research work remains to be done.

Not covered here in this brief tutorial are interesting applications of nonlinear theory in speech recognition [33, 34], use of fractals, modulation and power-law based methods as discussed in [35].

7. Acknowledgements

This work was supported by BT, EPSRC, and the Royal Society. The contributions of my colleague Iain Mann to this work are gratefully acknowledged.

References

[1] B. Gabioud, "Articulatory models in speech synthesis," in *Fundamentals of Speech Synthesis and Speech Recognition*, pp. 215–230, John Wiley & Sons, Chichester, UK, 1994.

[2] K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell System Technical Journal*, vol. 51, no. 6, pp. 1233–1268, July–August 1972.

[3] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Two-mass models of the vocal cords for natural sounding voice synthesis," *Journal of the Acoustical Society of America*, vol. 82, no. 4, pp. 1179–1192, 1987.

[4] G. Fant, *Acoustic Theory of Speech Production*, Mouton, The Hague, The Netherlands, 1960.
[5] J. Markel and A. Gray, *Linear Prediction of Speech*, Springer-Verlag, Berlin, Germany, 1976.
[6] D. H. Klatt, "Software for a cascade/parallel formant synthesiser," *Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 971–995, 1980.
[7] J. Page and A. Breen, "The laureate text-to-speech system—architecture and applications," *BT Technology Journal*, vol. 14, no. 1, pp. 57–67, 1996.
[8] M. Beutnagel, A. Conkie, J. Schroeter, Y. Stylianou, and A. Syrdal, "The AT&T next-gen TTS system," in *Proc. Joint Meeting of ASA, EAA, and DAGA*, pp. 18–24, Berlin, Germany, March 1999.
[9] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5/6, pp. 453–467, 1990.
[10] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.
[11] T. Koizumi, S. Taniguchi, and S. Hiromitsu, "Glottal source-vocal tract interaction," *Journal of the Acoustical Society of America*, vol. 78, no. 5, pp. 1541–1547, 1985.
[12] D. M. Brookes and P. A. Naylor, "Speech production modelling with variable glottal reflection coefficient," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 671–674, New York, NY, USA, April 1988.
[13] H. M. Teager and S. M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract," in *Proc. NATO Advanced Study Institute on Speech Production and Speech Modelling*, pp. 241–261, Bonas, France, July 1989.
[14] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *Journal of the Acoustical Society of America*, vol. 97, no. 3, pp. 1874–1884, 1995.
[15] J. Schoentgen, "Non-linear signal representation and its application to the modelling of the glottal waveform," *Speech Communication*, vol. 9, pp. 189–201, 1990.
[16] R. J. DiFrancesco and E. Moulines, "Detection of glottal closure by jumps in the statistical properties of the speech signal," *Speech Communication*, vol. 9, pp. 401–418, December 1990.
[17] Y. M. Cheng and D. O'Shaughnessy, "Automatic and reliable estimation of glottal closure instant and period," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1805–1815, 1989.
[18] D. Talkin, "Voicing epoch determination with dynamic programming," *Journal of the Acoustical Society of America*, vol. 85, no. Supplement 1, pp. S149, 1989.
[19] F. Takens, "Detecting strange attractors inturbulence," in *Proc. Symposium on Dynamical Systems and Turbulence*, A. Dold and B. Eckmann, Eds., Lecture Notes in Mathematics, pp. 366–381, Warwick, UK, 1980.
[20] D. S. Broomhead and G. P. King, "On the qualitative analysis of experimental dynamical systems," in *Nonlinear Phenomena and Chaos*, pp. 113–144, Adam Hilger, Bristol, UK, 1986.
[21] G. Kubin, "Poincaré section techniques for speech," in *Proc. 1997 IEEE Workshop on Speech Coding for Telecommunications*, pp. 7–8, Pocono Manor, Pa, USA, September 1997.
[22] G. Kubin, "Synthesis and coding of continuous speech with the nonlinear oscillator model," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 267–270, Atlanta, Ga, USA, May 1996.
[23] M. Birgmeier, *Kalman-trained neural networks for signal processing applications*, Ph.D. thesis, Technical University of Vienna, Vienna, Austria, 1996.
[24] K. Narasimhan, J. C. Principe, and D. Childers, "Nonlinear dynamic modeling of the voiced excitation for improved speech synthesis," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, pp. 389–392, Phoenix, Ariz, USA, March 1999.
[25] B. Mulgrew, "Applying radial basis functions," *IEEE Signal Processing Magazine*, vol. 13, no. 2, pp. 50–65, 1996.
[26] A. M. Fraser and H. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A (3)*, vol. 33, no. 2, pp. 1134–1140, 1986.
[27] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.

[28] I. Mann, *An investigation of nonlinear speech synthesis and pitch modification techniques*, Ph.D. thesis, Department of Electronics and Electrical Engineering, University of Edinburgh, Scotland, UK, September 1999.

[29] S. Haykin and J. Principe, "Making sense of a complex world," *IEEE Signal Processing Magazine*, vol. 15, no. 3, pp. 66–81, 1998.

[30] J. Stark, D. S. Broomhead, M. E. Davies, and J. Huke, "Takens embedding theorems for forced and stochastic systems," in *Proc. 2nd World Congress of Nonlinear Analysis*, Athens, Greece, July 1996.

[31] J. Stark, "Delay embeddings for forced systems. I. Deterministic forcing," *J. Nonlinear Sci.*, vol. 9, no. 3, pp. 255–332, 1999.

[32] H. Haas and G. Kubin, "A multi-band nonlinear oscillator model for speech," in *Proc. 32nd Asimolar Conference on Signals, Systems, and Computers*, vol. 1, pp. 338–342, Pacific Grove, Calif, USA, November 1998.

[33] D. N. Levin, "Sensor-independent stimulus representations," *Proc. National Academy of Sciences*, vol. 99, no. 11, pp. 7346–7351, 2002.

[34] D. N. Levin, "Representations of sound that are insensitive to spectral filtering and parametrization procedures," *Journal of the Acoustical Society of America*, vol. 111, no. 5 Pt 1, pp. 2257–2271, 2002.

[35] P. Maragos, A. Dimakis, and I. Kokkinos, "Some advances in nonlinear speech modeling using modulations, fractals, and chaos," in *Proc. 14th International Conference on Digital Signal Processing (DSP 2002)*, Santorini, Greece, July 2002.

Expansion of the EURASIP Directory of Theses

Each year hundreds of Ph.D. theses are published by universities around the world. Access to these documents is usually via contacting the library where the degree was awarded. EURASIP aims to grow a list and directory of theses via its website http://www.eurasip.org, that will allow a much wider dissemination for theses produced by EURASIP members or the students of current EURASIP members. Please access the website for further details on the procedure.

Jonathon Chambers

SIGNAL PROCESSING

Editorial Policy

Signal Processing is an interdisciplinary journal presenting the theory and practice of signal processing. Its primary objectives are the following:

- dissemination of research results and of engineering developments to all signal processing groups and individuals;
- presentation of practical solutions to current signal processing problems in engineering and science.

The editorial policy and the technical content of the journal are the responsibility of the Editor-in-Chief and the Editorial Board.

The journal is self-supporting from subscription income and contains a minimum amount of advertisements. Advertisements are subject to the prior approval of the Editor-in-Chief. The journal welcomes contributions from every country in the world.

Scope

Signal Processing incorporates all aspects of the theory and practice of signal processing (analogue and digital). It features original research work, tutorial and review articles, and accounts of practical developments. It is intended for a rapid dissemination of knowledge and experience to engineers and scientists working in signal processing research, development, or practical application.

Subjects

Subject areas covered by the journal include: Signal Theory; Stochastic Processes; Detection and Estimation; Spectral Analysis; Filtering; Communication Signal Processing; Biomedical Signal Processing; Geophysical and Astrophysical Signal Processing; Earth Resources Signal Processing; Acoustic and Vibration Signal Processing; Signal Processing Systems; Software Developments; Image Processing; Pattern Recognition; Optical Signal Processing; Multidimensional Signal Processing; Data Processing; Remote Sensing; Signal Processing Technology; Speech Processing; Radar Signal Processing; Sonar Signal Processing; Special Signal Processing; Industrial Applications; New Applications.

Editor-in-Chief

SIGNAL PROCESSING: IMAGE COMMUNICATION

Editorial Policy

Signal Processing: Image Communication is an international journal for the development of the theory and practice of image communication. Its primary objectives are the following:

- to present a forum for the advancement of the theory and practice of image communication;
- to simulate cross fertilization between areas similar in nature which have traditionally been separated, for example, various aspects of visual communications and information systems;
- to contribute to a rapid information exchange between the industrial and academic environments.

The editorial policy and the technical content of the journal are the responsibility of the Editor-in-Chief and the Editorial Board. The journal is self-supporting from subscription income and contains a minimum amount of advertisements.

Advertisements are subject to the prior approval of the Editor-in-Chief. The journal welcomes contributions from every country in the world.

Scope

Signal Processing: Image Communication publishes articles relating to aspects of design, implementation, and use of image communication systems. Signal Processing: Image Communication features original research work, tutorial and review articles, and accounts of practical developments.

Subjects

Subject areas covered by the journal include: TV, HDTV, and 3DTV systems; Visual Science; Image; TV and Advanced TV; Broadcasting; Image Storage and Retrieval; Graphic Arts; Electronic Printing; Image Transmission; Interactive Image Coding Communication; Imaging Technology; Display Technology; VLSI Processors for Image Communications.

Editor-in-Chief

Murat Tekalp, Electrical and Computer Engineering, University of Rochester, 204 Hopeman Bld River Campus, Rochester, NY 14627-0126, USA

SPEECH COMMUNICATION

Editorial Policy

The journal's primary objectives are the following:

- to present a forum for the advancement of human and human-machine speech communication science;
- to stimulate cross-fertilization between different fields of this domain;
- to contribute towards the rapid and wide diffusion of scientifically sound contributions in this domain.

Speech Communication is an interdisciplinary journal whose primary objective is to fulfill the need for the rapid dissemination and thorough discussion of basic and applied research results. In order to establish frameworks to inter-relate results from the various areas of the field, emphasis will be placed on viewpoints and topics of a transdisciplinary nature. The editorial policy and the technical content of the journal are the responsibility of the Editors and the Institutional Representatives. The Institutional Representatives assist the Editors in the definition and the control of editorial policy as well as in maintaining connections with scientific associations, international congresses, and regional events. The Editorial Board contributes towards the gathering of material for publication and assists the Editors in the editorial process.

Scope

Speech Communication is an interdisciplinary journal for the development and dissemination of all basic and applied aspects of speech communication processes. Speech Communication features original research work, tutorial, and review articles dealing with the theoretical, empirical, and practical aspects of this scientific field.

Subject Coverage

Subject areas covered in this journal include:

- Basics of oral communication and dialogue: modelling of production and perception processes; phonetics and phonology; syntax; semantics of speech communication; cognitive aspects.
- Models and tools for language learning: functional organisation and developmental models of human language capabilities; acquisition and rehabilitation of spoken language; speech & hearing defects and aids.
- Speech signal processing: analysis, coding, transmission, enhancement, robustness to noise.
- Models for automatic speech communication: speech recognition; language identification; speaker recognition; speech synthesis; oral dialogue.

- Development and evaluation tools: monolingual and multilingual databases; assessment methodologies; specialised hardware and software packages; field experiments; market development.
- Multimodal human computer interface: using speech I/O in combination with modalities, e.g., gesture and handwriting.

Editors-in-Chief

EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING

Scope

The overall aim of EURASIP Journal on Applied Signal Processing (EURASIP JASP) is to bring science and applications together with emphasis on practical aspects of signal processing in new and emerging technologies. It is directed as much at the practicing engineers as at the academic researchers. EURASIP JASP will highlight the diverse applications of signal processing and encourage a cross-fertilization of techniques. All papers should attempt to bring theory to life with practical simulations and examples. Tutorial articles on topics of interest are also welcomed. EURASIP JASP employs paperless, electronic review process to foster fast and speedy turnaround in review process.

There are two different issues: regular issues and special issues. The regular issues publish collections of papers without special solicitation. The special issues have specifically aimed and targeted topics of interest contributed by authors responding to a particular Call-for-Papers or by invitation, edited by invited guest editor(s). Regular papers can be submitted at any time, while special issue papers can be submitted only based on planned schedules and submission guidelines of the Call-for-Papers. Proposals for special issues can be submitted directly to the Editor-in-Chief.

Subjects

Subject areas include, but are by no means limited to:

- Signal processing theory, algorithm, architecture, design, and implementation
- Speech processing, coding, compression, and recognition
- Audio signal processing, coding, and compression
- Image/video processing, coding, compression, restoration, analysis and understanding and communications
- Multimedia signal processing and technology
- Signal processing for communications and networking
- Statistical and adaptive signal processing
- Nonlinear signal processing techniques
- Signal processing design tools
- Signal processing for security, authentication, and cryptography
- Analog signal processing
- Signal processing for smart sensor and systems

Application areas include but not limited to: communications; networking; sensors and actuators; radar and sonar; medical imaging; biomedical applications; remote sensing; consumer electronics; computer vision; pattern recognition; robotics; fiber optic sensing/transducers; industrial automation; transportation; stock market and financial analysis; seismography; avionics.

Indexed/Abstracted In

The articles of the EURASIP JASP are reviewed/indexed in Acoustics Abstracts; Computer and Communications Security Abstracts (CCSA); Current Contents Engineering, Computing & Technology; INSPEC; Science Citation Index Expanded; and Technology and Management (TEMA).

Editor-in-Chief

EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING

Scope

The overall aim of EURASIP Journal on Wireless Communications and Networking (EURASIP JWCN) is to bring science and applications together on wireless communications and networking technologies with emphasis on signal processing techniques and tools. It is directed at both practicing engineers and academic researchers. EURASIP JWCN will highlight the continued growth and new challenges in wireless technology, both for application development and basic research. Papers should emphasize original results relating to the theory and/or applications of wireless communications and networking. Tutorial papers, especially those emphasizing multidisciplinary views of communications and networking, are also welcome. EURASIP JWCN employs a paperless, electronic submission and evaluation system to promote a rapid turnaround in the peer review process.

The journal publishes two types of issues: regular issues and special issues. Regular issues publish collections of papers without special solicitation. The special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call-for-Papers or by invitation, edited by invited guest editor(s). Regular papers can be submitted at any time, while special issue papers can be submitted only based on planned schedules and submission guidelines of the Call-for-Papers. Proposals for special issues can be submitted directly to the Editor-in-Chief.

Subjects

Subject areas include, but are by no means limited to: Ad-hoc networks; Channel modeling and propagation; Detection, estimation, and synchronization; Diversity and space-time techniques; End-to-end design techniques; Error control coding; Iterative techniques for joint optimization; Modulation techniques (CDMA, OFDM, multicarrier, spread-spectrum, etc.); Multiuser, MIMO channels, and multiple access schemes; Network performance, reliability, and quality of service; Resource allocation over wireless networks Security, authentication, and cryptography; Signal processing techniques and tools; Ultra wide-band systems; Wireless network services and medium access control.

Editor-in-Chief

Phillip Regalia, Institut National des Télécommunications, 9 rue Charles Fourier, F-91011 Evry Cedex, France

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 1, 1 January 2003

Special Issue on Multimedia Signal Processing

Jean-Luc Dugelay

Institut Eurécom MultiMedia Communications Department, 2229 route des Cretes, BP 193, F-06904 Sophia Antipolis Cedex, France
Email: jld@eurecom.fr

Kenneth Rose

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, USA
Email: rose@ece.ucsb.edu

Recent years have seen the emergence of a variety of new multimedia (text, speech, music, image, graphics, and video) services, which accompanied an unprecedented explosion in the capacity and universal availability of networks. This evolution raised considerable challenges in the area of signal processing, where new algorithms are needed for efficient manipulation, analysis, interactive accessing, compression, storage, indexing, watermarking, and communication of multimedia signals, as well as the associated problems of hardware implementation, database management, understanding of human perception, and so on. Not surprisingly, the field of multimedia signal processing (MMSP) has been experiencing progress at a rapid pace. As evident from the above partial lists of signals and research problems, MMSP involves a diverse research community with complementary areas of expertise. Hence the continuing need for cross-fertilization and exchange of ideas, which also motivates the present collection of research papers.

This special issue offers a sample of current research in several areas of MMSP. It grew out of the October 2001 IEEE Workshop on Multimedia Signal Processing, and is a collection of invited papers that were selected to provide full treatment of work whose preliminary presentation at the workshop generated considerable interest. In particular, the areas of audio signal recognition and compression; human-machine interaction; image watermarking; video indexing; and video coding and streaming are covered.

The first group of three papers is dedicated to algorithms for audio signal compression, classification, and human-machine interaction. In *Musical Instrument Timbres Classification with Spectral Features*, G. Agostini, M. Longari, and E. Pollastri propose a framework for the classification and recognition of musical instruments based on monophonic music signals. In *Sinusoidal Analysis-Synthesis of Audio Using Perceptual Criteria*, T. Painter and A.

Spanias present a new method for the selection of sinusoidal components for use in compact representation of narrowband audio. Finally, in *An Acoustic Human-Machine Front-End for Multimedia Applications*, W. Herbordt, H. Buchner, and W. Kellermann address the problem of stereophonic acoustic echo cancellation.

The fourth paper entitled *Embedding Color Watermarks in Color Images*, by C.-H. Chou and T.-L. Wu, focuses on image watermarking with particular emphasis on color information, which has not been given enough consideration in the literature.

The next group of two papers is concerned with video indexing, which is crucial to the management of, navigation in, and retrieval from large databases. The paper *Retrieval by Local Motion*, by B. Erol and F. Kossentini, focuses on the important role of local motion in indexing. It proposes two new descriptors that capture the local motion of the video object within its bounding box. I. Yahiaoui, B. Huet, and B. Merialdo present a comparison of methodologies for automatic generation of video summaries in *Comparison of Multiepisode Video Summarization Algorithms*.

The last group of three papers addresses various aspects of video coding and transmission and focuses on source-channel coding optimization or compression-complexity tradeoffs. In *3D Scan-Based Wavelet Transform and Quality Control for Video Coding*, C. Parisot, M. Antonini, and M. Barlaud propose a new temporal scan-based wavelets that maintain the central advantages of wavelet coding without recourse to excessive complexity. The next paper, *Combined Wavelet Video Coding and Error Control for Internet Streaming and Multicast*, by T. Chu and Z. Xiong, is also concerned with wavelet video coding and proposes an integrated (compression and error control) approach to Internet video streaming and multicast. The last paper, by D. Comas R. Singh, A. Ortega, and F. Marqués, entitled *Unbalanced Multiple Description Video Coding Based on a Rate-Distortion Optimization* tackles the problem of robust streaming of video data over best-effort packet networks, using the multiple description paradigm.

Jean-Luc Dugelay
Kenneth Rose

Volume 2003, No. 1, 1 January 2003

Contents and Abstracts

Musical Instrument Timbres Classification with Spectral Features

Giulio Agostini, Maurizio Longari, and Emanuele Pollastri

http://dx.doi.org/10.1155/S1110865703210118

A set of features is evaluated for recognition of musical instruments out of monophonic musical signals. Aiming to achieve a compact representation, the adopted features regard only spectral characteristics of sound and are limited in number. On top of these descriptors, various classification methods are implemented and tested. Over a dataset of 1007 tones from 27 musical instruments, support vector machines and quadratic discriminant analysis show comparable results with success rates close to 70% of successful classifications. Canonical discriminant analysis never had momentous results, while nearest neighbours performed on average among the employed classifiers. Strings have been the most misclassified instrument family, while very satisfactory results have been obtained with brass and woodwinds. The most relevant features are demonstrated to be the inharmonicity, the spectral centroid, and the energy contained in the first partial.

Sinusoidal Analysis-Synthesis of Audio Using Perceptual Criteria

Ted Painter and Andreas Spanias

http://dx.doi.org/10.1155/S111086570321009X

This paper presents a new method for the selection of sinusoidal components for use in compact representations of narrowband audio. The method consists of ranking and selecting the most perceptually relevant sinusoids. The idea behind the method is to maximize the matching between the auditory excitation pattern associated with the original signal and the corresponding auditory excitation pattern associated with the modeled signal that is being represented by a small set of sinusoidal parameters. The proposed component-selection methodology is shown to outperform the maximum signal-to-mask ratio selection strategy in terms of subjective quality.

An Acoustic Human-Machine Front-End for Multimedia Applications

Wolfgang Herbordt, Herbert Buchner, and Walter Kellermann

http://dx.doi.org/10.1155/S1110865703211094

A concept of robust adaptive beamforming integrating stereophonic acoustic echo cancellation is presented which reconciles the need for low-computational complexity and efficient adaptive filtering with versatility and robustness in real-world scenarios. The synergetic combination of a robust generalized sidelobe canceller and a stereo acoustic echo canceller is designed in the frequency domain based on a general framework for multichannel adaptive filtering in the frequency domain. Theoretical analysis and real-time experiments show the superiority of this concept over comparable time-domain approaches in terms of computational complexity and adaptation behaviour. The real-time implementation confirms that the concept is robust and meets well the practical requirements of real-world scenarios, which makes it a promising candidate for commercial products.

Embedding Color Watermarks in Color Images

Chun-Hsien Chou and Tung-Lin Wu

http://dx.doi.org/10.1155/S1110865703211227

Robust watermarking with oblivious detection is essential to practical copyright protection of digital images. Effective exploitation of the characteristics of human visual perception to color stimuli helps to develop the watermarking scheme that fills the requirement. In this paper, an oblivious watermarking scheme that embeds color watermarks in color images is proposed. Through color gamut analysis and quantizer design, color watermarks are embedded by modifying quantization indices of color pixels without resulting in perceivable distortion. Only a small amount of information including the specification of color gamut, quantizer stepsize, and color tables is required to extract the watermark. Experimental results show that the proposed watermarking scheme is computationally simple and quite robust in face of various attacks such as cropping, low-pass filtering, white-noise addition, scaling, and JPEG compression with high compression ratios.

Retrieval by Local Motion

Berna Erol and Faouzi Kossentini

http://dx.doi.org/10.1155/S1110865703210076

Motion feature plays an important role in video retrieval. The current literature mostly addresses motion retrieval only by camera motion and global motion of individual video objects in a video scene. In this paper, we propose two new motion descriptors that capture the local motion of the video object within its bounding box. The proposed descriptors are rotation and scale invariant and based on the angular and circular area variances of the video object and the variances of the angular radial transform coefficients. Experiments show that ranking obtained by querying with our proposed descriptors closely match with the human ranking.

Comparison of Multiepisode Video Summarization Algorithms

Itheri Yahiaoui, Bernard Merialdo, and Benoit Huet

http://dx.doi.org/10.1155/S1110865703210052

This paper presents a comparison of some methodologies for the automatic construction of video summaries. The work is based on the simulated user principle to evaluate the quality

of a video summary in a way that is automatic, yet related to the user's perception. The method is studied for the case of multiepisode video, where we do not describe only what is important in a video but rather what distinguishes this video from the others. Experimental results are presented to support the proposed ideas.

3D Scan-Based Wavelet Transform and Quality Control for Video Coding

Christophe Parisot, Marc Antonini, and Michel Barlaud

http://dx.doi.org/10.1155/S1110865703210064

Wavelet coding has been shown to achieve better compression than DCT coding and moreover allows scalability. 2D DWT can be easily extended to 3D and thus applied to video coding. However, 3D subband coding of video suffers from two drawbacks. The first is the amount of memory required for coding large 3D blocks; the second is the lack of temporal quality due to the sequence temporal splitting. In fact, 3D block-based video coders produce jerks. They appear at blocks temporal borders during video playback. In this paper, we propose a new temporal scan-based wavelet transform method for video coding combining the advantages of wavelet coding (performance, scalability) with acceptable reduced memory requirements, no additional CPU complexity, and avoiding jerks. We also propose an efficient quality allocation procedure to ensure a constant quality over time.

Combined Wavelet Video Coding and Error Control for Internet Streaming and Multicast

Tianli Chu and Zixiang Xiong

http://dx.doi.org/10.1155/S1110865703210106

This paper proposes an integrated approach to Internet video streaming and multicast (e.g., receiver-driven layered multicast (RLM) by McCanne) based on combined wavelet video coding and error control. We design a packetized wavelet video (PWV) coder to facilitate its integration with error control. The PWV coder produces packetized layered bitstreams that are independent among layers while being embedded within each layer. Thus, a lost packet only renders the following packets in the same layer useless. Based on the PWV coder, we search for a multilayered error-control strategy that optimally trades off source and channel coding for each layer under a given transmission rate to mitigate the effects of packet loss. While both the PWV coder and the error-control strategy are new—the former incorporates embedded wavelet video coding and packetization and the latter extends the single-layered approach for RLM by Chou et al.—the main distinction of this paper lies in the seamless integration of the two parts. Theoretical analysis shows a gain of up to 1 dB on a channel with 20% packet loss using our combined approach over separate designs of the source coder and the error-control mechanism. This is also substantiated by our simulations with a gain of up to 0.6 dB. In addition, our simulations show a gain of up to 2.2 dB over previous results reported by Chou et al.

Unbalanced Multiple-Description Video Coding with Rate-Distortion Optimization

David Comas, Raghavendra Singh, Antonio Ortega, and Ferran Marqués

http://dx.doi.org/10.1155/S1110865703211215

We propose to use multiple-description coding (MDC) to protect video information against packet losses and delay, while also ensuring that it can be decoded using a standard decoder. Video data are encoded into a high-resolution stream using a standard compliant encoder. In addition, a low-resolution stream is generated by duplicating the relevant information (motion vectors, headers and some of the DCT coefficient) from the high-resolution stream while the remaining coefficients are set to zero. Both streams are independently decodable by a standard decoder. However, only in case of losses in the high resolution description, the corresponding information from the low resolution stream is decoded, else the received high resolution description is decoded. The main contribution of this paper is an optimization algorithm which, given the loss ratio, allocates bits to both descriptions and selects the right number of coefficients to duplicate in the low-resolution stream so as to minimize the expected distortion at the decoder end.

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 2, 1 February 2003

Special Issue on

Unstructured Information Management from Multimedia Data Sources

Jing Huang

IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA
Email: jghg@us.ibm.com

Mukund Padmanabhan

Renaissance Technologies Corporation, 600 Route 25A, East Setauket, NY 11733, USA
Email: mukund@rentec.com

Savitha Srinivasan

IBM Almaden Research Center, San Jose, CA 95120, USA
Email: savitha@almaden.ibm.com

The recent proliferation of the worldwide web and the low cost of storage have contributed to an explosively growing volume of information. Traditionally, in order to be usable, information needs to be in some form of structured format, such as records in relational databases, XML tagged data types, and so forth. The field of structured-information management deals with techniques to create, store, query, and mine these data types. A fundamental characteristic of accessing such a database is that a data query returns an absolute list of matches in the database.

However, the vast majority of data created and stored today does not exist in structured format. For instance, a recent analytic study reports that only about 20 percent of all corporate content exists in structured formats such as transactional data or product specifications. The rest of the data exists in unstructured, machine-generated formats such as data from medical sensors, security cameras, audio recordings of meetings, broadcasts, traffic video, and so forth. There is often very valuable information buried in such unstructured data (e.g., call-center data may contain information about customer trends); however, the information is not directly accessible, because of its unstructured nature. Although it is possible to convert such data sources to structured forms by manual processing, the high cost associated with this enables only a very small portion of the data to be processed in this fashion. Consequently, there is a great deal of research and commercial value in developing methods both to manage this data and to automatically analyze and extract semantics present in it.

The ease of managing such unstructured data depends on its complexity. One way to characterize complexity is to examine its multimedia properties such as visual, spatial, and temporal components, the ease of data entry, and the existence of well-defined semantic units by which the data can be indexed and searched. Measuring the complexity of unstructured data types along these properties leads to an increasing order of complexity from text and image to audio and video.

For text data types, the basic approach used in information management is to first "extract a sequence of features" from the data; subsequently, the data is "indexed" by the features or the features are compared to templates stored in a library, and the data is "indexed" by a list of templates. A data query of this processed unstructured data would then compute the "similarity" between the query and the indexed data, and return a "ranked list of potential matches" (as opposed to an absolute list of matches as in the case of a query on structured data). Such methods have evolved to some level of maturity in the case of text data types, and in order to capitalize on this, most current methods of dealing with multimedia data first attempt to convert the data into text format and then use text-based techniques to manage it.

We could hence think of an unstructured-information management system as having three phases. In the initial phase of converting multimedia sources into text, research in speech recognition (conversion of speech to text) plays a pivotal role in the processing of unstructured speech data, and research in video processing and content analysis play a pivotal role in the processing of image and video data. As signal processing plays a fundamental role in speech and video processing, we could think of the problem of extracting information from unstructured multimedia sources as an extended application of signal processing. In the second phase of information management, research in feature extraction, indexing, similarity matching, and ranking plays a pivotal role. The third and final phase relates to integrating querying, browsing, and the search paradigm of the complete system. The development of efficient multimedia navigation, summarization, and browsing tools is an important part of this last phase.

This special issue focuses on unstructured-information management across several different unstructured data types. The first paper deals with unstructured text data. In the remaining papers, we transit into other unstructured data types beginning with audio, move on to image, and conclude with video. Each section starts with an overview paper, which attempts to give a high-level picture of the various building blocks used in the solution. This is followed with papers that provide further details about specific building blocks. The section is then concluded with a paper that describes an example of a complete solution or a real application.

The first paper is about a novel feature selection method with applications in managing text data. The next four papers deal with audio as the raw data format (e.g., broadcast news, call-center conversations). The section starts with an overview paper by James Allan that gives a high-level view of the components of a system that starts with audio data as a source and extracts information from it. Subsequently, the papers by Wolfang Macherey et al. and Chiori Hori et al. delve into the theoretical aspects of the system. Finally, the paper by Jean-Luc Gauvain and Lori Lamel describe a system that employs all these methods to successfully process radio-broadcast news. Switching gear from temporal data (audio) to temporal-spatial data (image), the paper by Jing Huang et al. presents a scheme for hierarchical classification of images via supervised learning. The last five papers deal with images and video as the raw data format. The section starts with a paper by Yihong Gong on audio-video summarization that generates a video summary by alignment of the visual summary with the audio summary. The next paper by W. H. Adams et al. that explores semantic indexing of multimedia content building upon well-known techniques for audio, video, and text retrieval and focuses on the use of Bayesian networks for the fusion of different classifiers. The next paper by Thijs Westerveld et al. investigates the effect of language models both in text retrieval and for visual features such as shots and scenes. This is followed by a video classification and retrieval paper that takes advantage of motion patterns. The last paper in this section, by Arnon Amir et al., discusses the practical aspects of a multimedia retrieval system and emphasizes the role of browsing in multimedia retrieval systems.

It is hoped that these papers would give the readers an introduction to the vast field of unstructured-information management and its potential benefits and applications, and also acquaint them with the state-of-the-art in extracting information from various formats of unstructured multimedia data.

Jing Huang
Mukund Padmanabhan
Savitha Srinivasan

Volume 2003, No. 2, 1 February 2003

Contents and Abstracts

Discriminative Feature Selection via Multiclass Variable Memory Markov Model

Noam Slonim, Gill Bejerano, Shai Fine, and Naftali Tishby

http://dx.doi.org/10.1155/S111086570321115X

We propose a novel feature selection method based on a variable memory Markov (VMM) model. The VMM was originally proposed as a generative model trying to preserve the original source statistics from training data. We extend this technique to simultaneously handle several sources, and further apply a new criterion to prune out nondiscriminative features out of the model. This results in a multiclass discriminative VMM (DVMM), which is highly efficient, scaling linearly with data size. Moreover, we suggest a natural scheme to sort the remaining features based on their discriminative power with respect to the sources at hand. We demonstrate the utility of our method for text and protein classification tasks.

Robust Techniques for Organizing and Retrieving Spoken Documents

James Allan

http://dx.doi.org/10.1155/S1110865703211070

Information retrieval tasks such as document retrieval and topic detection and tracking (TDT) show little degradation when applied to speech recognizer output. We claim that the robustness of the process is because of inherent redundancy in the problem: not only are words repeated, but semantically related words also provide support. We show how document and query expansion can enhance that redundancy and make document retrieval robust to speech recognition errors. We show that the same effect is true for TDT's tracking task, but that recognizer errors are more of an issue for new event and story link detection.

Probabilistic Aspects in Spoken Document Retrieval

Wolfgang Macherey, Hans Jörg Viechtbauer, and Hermann Ney

http://dx.doi.org/10.1155/S1110865703210088

Accessing information in multimedia databases encompasses a wide range of applications in which spoken document retrieval (SDR) plays an important role. In SDR, a set of automatically transcribed speech documents constitutes the files for retrieval, to which a user may address a request in natural language. This paper deals with two probabilistic aspects in SDR. The first part investigates the effect of recognition errors on retrieval performance and inquires the question of why recognition errors have only a little effect on the retrieval performance. In the second part, we present a new probabilistic approach to SDR that is

based on interpolations between document representations. Experiments performed on the TREC-7 and TREC-8 SDR task show comparable or even better results for the new proposed method than other advanced heuristic and probabilistic retrieval metrics.

A Statistical Approach to Automatic Speech Summarization

Chiori Hori, Sadaoki Furui, Rob Malkin, Hua Yu, and Alex Waibel

http://dx.doi.org/10.1155/S1110865703211112

This paper proposes a statistical approach to automatic speech summarization. In our method, a set of words maximizing a summarization score indicating the appropriateness of summarization is extracted from automatically transcribed speech and then concatenated to create a summary. The extraction process is performed using a dynamic programming (DP) technique based on a target compression ratio. In this paper, we demonstrate how an English news broadcast transcribed by a speech recognizer is automatically summarized. We adapted our method, which was originally proposed for Japanese, to English by modifying the model for estimating word concatenation probabilities based on a dependency structure in the original speech given by a stochastic dependency context free grammar (SDCFG). We also propose a method of summarizing multiple utterances using a two-level DP technique. The automatically summarized sentences are evaluated by summarization accuracy based on a comparison with a manual summary of speech that has been correctly transcribed by human subjects. Our experimental results indicate that the method we propose can effectively extract relatively important information and remove redundant and irrelevant information from English news broadcasts.

Structuring Broadcast Audio for Information Access

Jean-Luc Gauvain and Lori Lamel

http://dx.doi.org/10.1155/S1110865703211033

One rapidly expanding application area for state-of-the-art speech recognition technology is the automatic processing of broadcast audiovisual data for information access. Since much of the linguistic information is found in the audio channel, speech recognition is a key enabling technology which, when combined with information retrieval techniques, can be used for searching large audiovisual document collections. Audio indexing must take into account the specificities of audio data such as needing to deal with the continuous data stream and an imperfect word transcription. Other important considerations are dealing with language specificities and facilitating language portability. At Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), broadcast news transcription systems have been developed for seven languages: English, French, German, Mandarin, Portuguese, Spanish, and Arabic. The transcription systems have been integrated into prototype demonstrators for several application areas such as audio data mining, structuring audiovisual archives, selective dissemination of information, and topic tracking for media monitoring. As examples, this paper addresses the spoken document retrieval and topic tracking tasks.

Automatic Hierarchical Color Image Classification

Jing Huang, S. Ravi Kumar, and Ramin Zabih

http://dx.doi.org/10.1155/S1110865703211161

Organizing images into semantic categories can be extremely useful for content-based image retrieval and image annotation. Grouping images into semantic classes is a difficult problem, however. Image classification attempts to solve this hard problem by using low-level image features. In this paper, we propose a method for hierarchical classification of images via supervised learning. This scheme relies on using a good low-level feature and subsequently performing feature-space reconfiguration using singular value decomposition to reduce noise and dimensionality. We use the training data to obtain a hierarchical classification tree that can be used to categorize new images. Our experimental results suggest that this scheme not only performs better than standard nearest-neighbor techniques, but also has both storage and computational advantages.

Summarizing Audiovisual Contents of a Video Program

Yihong Gong

http://dx.doi.org/10.1155/S1110865703211082

In this paper, we focus on video programs that are intended to disseminate information and knowledge such as news, documentaries, seminars, etc, and present an audiovisual summarization system that summarizes the audio and visual contents of the given video separately, and then integrating the two summaries with a partial alignment. The audio summary is created by selecting spoken sentences that best present the main content of the audio speech while the visual summary is created by eliminating duplicates/redundancies and preserving visually rich contents in the image stream. The alignment operation aims to synchronize each spoken sentence in the audio summary with its corresponding speaker's face and to preserve the rich content in the visual summary. A Bipartite Graph-based audiovisual alignment algorithm is developed to efficiently find the best alignment solution that satisfies these alignment requirements. With the proposed system, we strive to produce a video summary that: (1) provides a natural visual and audio content overview, and (2) maximizes the coverage for both audio and visual contents of the original video without having to sacrifice either of them.

Semantic Indexing of Multimedia Content Using Visual, Audio, and Text Cues

W. H. Adams, Giridharan Iyengar, Ching-Yung Lin, Milind Ramesh Naphade, Chalapathy Neti, Harriet J. Nock, and John R. Smith

http://dx.doi.org/10.1155/S1110865703211173

We present a learning-based approach to the semantic indexing of multimedia content using cues derived from audio, visual, and text features. We approach the problem by developing a set of statistical models for a predefined lexicon. Novel concepts are then mapped

in terms of the concepts in the lexicon. To achieve robust detection of concepts, we exploit features from multiple modalities, namely, audio, video, and text. Concept representations are modeled using Gaussian mixture models (GMM), hidden Markov models (HMM), and support vector machines (SVM). Models such as Bayesian networks and SVMs are used in a late-fusion approach to model concepts that are not explicitly modeled in terms of features. Our experiments indicate promise in the proposed classification and fusion methodologies: our proposed fusion scheme achieves more than 10% relative improvement over the best unimodal concept detector.

A Probabilistic Multimedia Retrieval Model and Its Evaluation

Thijs Westerveld, Arjen P. de Vries, Alex van Ballegooij, Franciska de Jong, and Djoerd Hiemstra

http://dx.doi.org/10.1155/S111086570321101X

We present a probabilistic model for the retrieval of multimodal documents. The model is based on Bayesian decision theory and combines models for text-based search with models for visual search. The textual model is based on the language modelling approach to text retrieval, and the visual information is modelled as a mixture of Gaussian densities. Both models have proved successful on various standard retrieval tasks. We evaluate the multimodal model on the search task of TREC's video track. We found that the disclosure of video material based on visual information only is still too difficult. Even with purely visual information needs, text-based retrieval still outperforms visual approaches. The probabilistic model is useful for text, visual, and multimedia retrieval. Unfortunately, simplifying assumptions that reduce its computational complexity degrade retrieval effectiveness. Regarding the question whether the model can effectively combine information from different modalities, we conclude that whenever both modalities yield reasonable scores, a combined run outperforms the individual runs.

Motion Pattern-Based Video Classification and Retrieval

Yu-Fei Ma and Hong-Jiang Zhang

http://dx.doi.org/10.1155/S1110865703211021

Today's content-based video retrieval technologies are still far from human's requirements. A fundamental reason is the lack of content representation that is able to bridge the gap between visual features and semantic conception in video. In this paper, we propose a motion pattern descriptor, *motion texture* that characterizes motion in a generic way. With this representation, we design a semantic classification scheme to effectively map video clips to semantic categories. Support vector machines (SVMs) are used as the classifiers. In addition, this scheme also improves significantly the performance of motion-based shot retrieval due to the comprehensiveness and effectiveness of motion pattern descriptor and the semantic classification capability as shown by experimental evaluations.

Search the Audio, Browse the Video—A Generic Paradigm for Video Collections

Arnon Amir, Savitha Srinivasan, and Alon Efrat

http://dx.doi.org/10.1155/S111086570321012X

The amount of digital video being shot, captured, and stored is growing at a rate faster than ever before. The large amount of stored video is not penetrable without efficient video indexing, retrieval, and browsing technology. Most prior work in the field can be roughly categorized into two classes. One class is based on image processing techniques, often called content-based image and video retrieval, in which video frames are indexed and searched for visual content. The other class is based on spoken document retrieval, which relies on automatic speech recognition and text queries. Both approaches have major limitations. In the first approach, semantic queries pose a great challenge, while the second, speech-based approach, does not support efficient video browsing. This paper describes a system where speech is used for efficient searching and visual data for efficient browsing, a combination that takes advantage of both approaches. A fully automatic indexing and retrieval system has been developed and tested. Automated speech recognition and phonetic speech indexing support text-to-speech queries. New browsable views are generated from the original video. A special synchronized browser allows instantaneous, context-preserving switching from one view to another. The system was successfully used to produce searchable-browsable video proceedings for three local conferences.

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 3, 1 March 2003

Volume 2003, No. 3, 1 March 2003

Contents and Abstracts

Removing Impulse Bursts from Images by Training-Based Filtering

Pertti Koivisto, Jaakko Astola, Vladimir Lukin, Vladimir Melnik, and Oleg Tsymbal

http://dx.doi.org/10.1155/S1110865703211045

The characteristics of impulse bursts in remote sensing images are analyzed and a model for this noise is proposed. The model also takes into consideration other noise types, for example, the multiplicative noise present in radar images. As a case study, soft morphological filters utilizing a training-based optimization scheme are used for the noise removal. Different approaches for the training are discussed. It is shown that these techniques can provide an effective removal of impulse bursts. At the same time, other noise types in images, for example, the multiplicative noise, can be suppressed without compromising good edge and detail preservation. Numerical simulation results, as well as examples of real remote sensing images, are presented.

GA-Based Image Restoration by Isophote Constraint Optimization

Jong Bae Kim and Hang Joon Kim

http://dx.doi.org/10.1155/S1110865703212142

We propose an efficient technique for image restoration based on a genetic algorithm (GA) with an isophote constraint. In our technique, the image restoration problem is modeled as an optimization problem which, in our case, is solved by a cost function with isophote constraint that is minimized using a GA. We consider that an image is decomposed into isophotes based on connected components of constant intensity. The technique creates an optimal connection of all pairs of isophotes disconnected by a caption in the frame. For connecting the disconnected isophotes, we estimate the value of the smoothness, given by the best chromosomes of the GA and project this value in the isophote direction. Experimental results show a great possibility for automatic restoration of a region in an advertisement scene.

An Adaptive Video Coding Control Scheme for Real-Time MPEG Applications

Shih-Chang Hsia

http://dx.doi.org/10.1155/S1110865703210040

This paper proposes a new rate control scheme to increase the coding efficiency for MPEG systems. Instead of using a static group of picture (GOP) structure, we present an adaptive GOP structure that uses more P- and B-frame coding, while the temporal correlation among the video frames maintains high. When there is a scene change, we immediately insert intramode coding to reduce the prediction error. Moreover, an enhanced prediction frame is used to improve the coding quality in the adaptive GOP. This rate control algorithm can both achieve better coding efficiency and solve the scene change problem. Even if the coding bit rate is over the predefined level, this coding scheme does not require re-encoding for real-time systems. Simulations demonstrate that our proposed algorithm can achieve better quality than TM5, and satisfactory reliability for detecting scene changes.

Audio Watermarking Based on HAS and Neural Networks in DCT Domain

Hung-Hsu Tsai, Ji-Shiung Cheng, and Pao-Ta Yu

http://dx.doi.org/10.1155/S1110865703208027

We propose a new intelligent audio watermarking method based on the characteristics of the HAS and the techniques of neural networks in the DCT domain. The method makes the watermark imperceptible by using the audio masking characteristics of the HAS. Moreover, the method exploits a neural network for memorizing the relationships between the original audio signals and the watermarked audio signals. Therefore, the method is capable of extracting watermarks without original audio signals. Finally, the experimental results are also included to illustrate that the method significantly possesses robustness to be immune against common attacks for the copyright protection of digital audio.

Improved Facial-Feature Detection for AVSP via Unsupervised Clustering and Discriminant Analysis

Simon Lucey, Sridha Sridharan, and Vinod Chandran

http://dx.doi.org/10.1155/S1110865703209045

An integral part of any audio-visual speech processing (AVSP) system is the front-end visual system that detects facial features (e.g., eyes and mouth) pertinent to the task of visual speech processing. The ability of this front-end system to not only locate, but also give a confidence measure that the facial feature is present in the image, directly affects the ability of any subsequent postprocessing task such as speech or speaker recognition. With these issues in mind, this paper presents a framework for a facial-feature detection system suitable for use in an AVSP system, but whose basic framework is useful for any application requiring frontal facial-feature detection. A novel approach for facial-feature detection is presented, based on an appearance paradigm. This approach, based on intraclass unsupervised clustering and discriminant analysis, displays improved detection performance over conventional techniques.

Robust Clustering of Acoustic Emission Signals Using Neural Networks and Signal Subspace Projections

Vahid Emamian, Mostafa Kaveh, Ahmed H. Tewfik, Zhiqiang Shi, Laurence J. Jacobs, and Jacek Jarzynski

http://dx.doi.org/10.1155/S1110865703210027

Acoustic emission-based techniques are being used for the nondestructive inspection of mechanical systems. For reliable automatic fault monitoring related to the generation and propagation of cracks, it is important to identify the transient crack-related signals in the presence of strong time-varying noise and other interferences. A prominent difficulty is the inability to differentiate events due to crack growth from noise of various origins. This work presents a novel algorithm for automatic clustering and separation of acoustic emission (AE) events based on multiple features extracted from the experimental data. The algorithm consists of two steps. In the first step, the noise is separated from the events of interest and subsequently removed using a combination of covariance analysis, principal component analysis (PCA), and differential time delay estimates. The second step processes the remaining data using a self-organizing map (SOM) neural network, which outputs the noise and AE signals into separate neurons. To improve the efficiency of classification, the short-time Fourier transform (STFT) is applied to retain the time-frequency features of the remaining events, reducing the dimension of the data. The algorithm is verified with two sets of data, and a correct classification ratio over 95% is achieved.

An Effective Technique for Enhancing an Intrauterine Catheter Fetal Electrocardiogram

Steven L. Horner and William M. Holls III

http://dx.doi.org/10.1155/S1110865703212026

Physicians can obtain fetal heart rate, electrophysiological information, and uterine contraction activity for determining fetal status from an intrauterine catheters electrocardiogram with the maternal electrocardiogram canceled. In addition, the intrauterine catheter would allow physicians to acquire fetal status with one noninvasive to the fetus biosensor as compared to invasive to the fetus scalp electrode and intrauterine pressure catheter used currently. A real-time maternal electrocardiogram cancellation technique of the intrauterine catheters electrocardiogram will be discussed along with an analysis for the methods effectiveness with synthesized and clinical data. The positive results from an original detailed subjective and objective analysis of synthesized and clinical data clearly indicate that the maternal electrocardiogram cancellation method was found to be effective. The resulting intrauterine catheters electrocardiogram from effectively canceling the maternal electrocardiogram could be used for determining fetal heart rate, fetal electrocardiogram electrophysiological information, and uterine contraction activity.

Chebyshev Functions-Based New Designs of Halfband Low/Highpass Quasi-Equiripple FIR Digital Filters

Ishtiaq Rasool Khan and Ryoji Ohba

http://dx.doi.org/10.1155/S1110865703211148

Chebyshev functions, which are equiripple in a certain domain, are used to generate equiripple halfband lowpass frequency responses. Inverse Fourier transformation is then used to obtain explicit formulas for the corresponding impulse responses. The halfband lowpass FIR digital filters designed in this way are quasi-equiripple, having performances very close to those of true equiripple filters, and are comparatively much simpler to design.

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 4, 15 March 2003

Special Issue on Sensor Networks

Kung Yao

Department of Electrical Engineering, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1594, USA
Email: yao@ee.ucla.edu

Deborah Estrin

Department of Computer Science, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1596, USA
Email: destrin@cs.ucla.edu

Yu Hen Hu

Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706-1691, USA
Email: hu@engr.wisc.edu

Advances in low-cost and low-power wireless communication, microsensor, and microprocessor hardware, as well as progress in ad hoc networking routing and protocols, distributed signal and array processing, pervasive computing, and embedded systems have all made sensor networking a topic of active interest. In recent years, the Internet has been able to provide a large number of users with the ability to move diverse forms of information readily and thus revolutionized business, industry, defense, science, education, research, and human interactions. Sensor networking may, in the long run, be equally significant by providing measurement of the physical phenomena around us, leading to their understanding and ultimately the utilization of this information for a wide range of applications. Potential applications of sensor networking include environmental monitoring, health care monitoring, battlefield surveillance and reconnaissance, modern highway, modern manufacturing, condition-based maintenance of complex systems, and so forth.

In order to understand and build sensor networks, diverse technology and technical disciplines are involved. However, in this special issue we deal only with various signal processing aspects of sensor networking. Of the seven papers, four of them deal with source localization, two of them with tracking, and one with sensor network decomposition and organization. *Energy-Based Collaborative Source Localization Using Acoustic Microsensor Array*, by D. Li and Y. H. Hu, uses acoustic energy measurements to perform source localization. This approach assumes the acoustic source energy decays inversely with the square of the distance. By comparing acoustic sensor energy measurements around the source, the source location can be estimated as the intersection of multiple hyperspheres. *The Fusion of Distributed Microphone Arrays for Sound Localization*, by P. Aarabi, also deals with acoustic source localization. The author proposes to use the spatial observability function (SOF), which gives an indication of how well a microphone array perceives events at different spatial position. Each microphone array also has a spatial likelihood function (SLF) which reports the likelihood of a source at each spatial location. SOF and SLF approaches are used together for sound localization. In *A Self-Localization Method for Wireless Sensor Networks*, by R. L. Moses, D. Krishnamurthy, and R. Patterson, the authors consider the problem of locating and orienting a network of unattended sensors by using a number of known source signals for calibration purposes. The maximum-likelihood (ML) estimation and Cramér-Rao Bound (CRB) techniques are used. *Acoustic Source Localization and Beamforming: Theory and Practice*, by J. C. Chen, K. Yao, and R. E. Hudson, again uses the ML method for direct localization of wideband acoustical source in the near field and uses the cross bearing of the direction-of-arrivals (DOA) for localization in the far field. For multiple sources, an alternating projection procedure is used. CRB analysis provides various insights for the localization problem. *Dynamic Agent Classification and Tracking Using an Ad Hoc Mobile Acoustic Sensor Network*, by D. Friedlander, C. Griffin, N. Jacobson, S. Phoha, and R. R. Brooks, presents methods for dynamic distributed signal processing using an ad hoc mobile network of sensors to detect, identify, and track targets. Forming dynamic clusters around events of interest allows for processing multiple events in parallel over different geographic areas along the trajectory of the targets. In *Collaborative In-Network Processing for Target Tracking*, J. Liu, J. Reich, and F. Zhao consider collaborative signal processing using acoustic-amplitude sensors for target distance estimation and DOA sensors for bearing estimation. The information-driven sensor querying framework selectively activates sensors based on their utility and cost. Issues of distributed processing for tracking and energy efficiency of the network are addressed. *Preprocessing in a Tiered Sensor Network for Habitat Monitoring*, by H. Wang, D. Estrin, and L. Girod, considers some common principles for task-decomposition and collaboration for tiered sensor networks. The system has a few powerful macronodes in the first tier and many less-powerful nodes in the second tier. Each macronode combines data collected by many micronodes for target classification and localization. Application is made to habitat monitoring and classification and localization of birds. All seven of these papers use simulations and measured data to verify the proposed methods. In the coming years, it is expected that sensor networking will become ever more important both in research and industry and that hardware and software availability will enable significant data collection and field experimentation.

Kung Yao
Deborah Estrin
Yu Hen Hu

Volume 2003, No. 4, 15 March 2003

Contents and Abstracts

Energy-Based Collaborative Source Localization Using Acoustic Microsensor Array

Dan Li and Yu Hen Hu

http://dx.doi.org/10.1155/S1110865703212075

A novel sensor network source localization method based on acoustic energy measurements is presented. This method makes use of the characteristics that the acoustic energy decays inversely with respect to the square of distance from the source. By comparing energy readings measured at surrounding acoustic sensors, the source location during that time interval can be accurately estimated as the intersection of multiple hyperspheres. Theoretical bounds on the number of sensors required to yield unique solution are derived. Extensive simulations have been conducted to characterize the performance of this method under various parameter perturbations and noise conditions. Potential advantages of this approach include low intersensor communication requirement, robustness with respect to parameter perturbations and measurement noise, and low-complexity implementation.

The Fusion of Distributed Microphone Arrays for Sound Localization

Parham Aarabi

http://dx.doi.org/10.1155/S1110865703212014

This paper presents a general method for the integration of distributed microphone arrays for localization of a sound source. The recently proposed sound localization technique, known as SRP-PHAT, is shown to be a special case of the more general microphone array integration mechanism presented here. The proposed technique utilizes spatial likelihood functions (SLFs) produced by each microphone array and integrates them using a weighted addition of the individual SLFs. This integration strategy accounts for the different levels of access that a microphone array has to different spatial positions, resulting in an intelligent integration strategy that weighs the results of reliable microphone arrays more significantly. Experimental results using 10 2-element microphone arrays show a reduction in the sound localization error from 0.9 m to 0.08 m at a signal-to-noise ratio of 0 dB. The proposed technique also has the advantage of being applicable to multimodal sensor networks.

A Self-Localization Method for Wireless Sensor Networks

Randolph L. Moses, Dushyanth Krishnamurthy, and Robert M. Patterson

http://dx.doi.org/10.1155/S1110865703212063

We consider the problem of locating and orienting a network of unattended sensor nodes that have been deployed in a scene at unknown locations and orientation angles. This self-calibration problem is solved by placing a number of source signals, also with unknown

locations, in the scene. Each source in turn emits a calibration signal, and a subset of sensor nodes in the network measures the time of arrival and direction of arrival (with respect to the sensor node's local orientation coordinates) of the signal emitted from that source. From these measurements we compute the sensor node locations and orientations, along with any unknown source locations and emission times. We develop necessary conditions for solving the self-calibration problem and provide a maximum likelihood solution and corresponding location error estimate. We also compute the Cramér-Rao bound of the sensor node location and orientation estimates, which provides a lower bound on calibration accuracy. Results using both synthetic data and field measurements are presented.

Acoustic Source Localization and Beamforming: Theory and Practice

Joe C. Chen, Kung Yao, and Ralph E. Hudson

http://dx.doi.org/10.1155/S1110865703212038

We consider the theoretical and practical aspects of locating acoustic sources using an array of microphones. A maximum-likelihood (ML) direct localization is obtained when the sound source is near the array, while in the far-field case, we demonstrate the localization via the cross bearing from several widely separated arrays. In the case of multiple sources, an alternating projection procedure is applied to determine the ML estimate of the DOAs from the observed data. The ML estimator is shown to be effective in locating sound sources of various types, for example, vehicle, music, and even white noise. From the theoretical Cramér-Rao bound analysis, we find that better source location estimates can be obtained for high-frequency signals than low-frequency signals. In addition, large range estimation error results when the source signal is unknown, but such unknown parameter does not have much impact on angle estimation. Much experimentally measured acoustic data was used to verify the proposed algorithms.

Dynamic Agent Classification and Tracking Using an Ad Hoc Mobile Acoustic Sensor Network

David Friedlander, Christopher Griffin, Noah Jacobson, Shashi Phoha, and Richard R. Brooks

http://dx.doi.org/10.1155/S1110865703212099

Autonomous networks of sensor platforms can be designed to interact in dynamic and noisy environments to determine the occurrence of specified transient events that define the dynamic process of interest. For example, a sensor network may be used for battlefield surveillance with the purpose of detecting, identifying, and tracking enemy activity. When the number of nodes is large, human oversight and control of low-level operations is not feasible. Coordination and self-organization of multiple autonomous nodes is necessary to maintain connectivity and sensor coverage and to combine information for better understanding the dynamics of the environment. Resource conservation requires adaptive clustering in the vicinity of the event. This paper presents methods for dynamic distributed signal processing using an ad hoc mobile network of microsensors to detect, identify, and track targets in noisy environments. They seamlessly integrate data from fixed and mobile

platforms and dynamically organize platforms into clusters to process local data along the trajectory of the targets. Local analysis of sensor data is used to determine a set of target attribute values and classify the target. Sensor data from a field test in the Marine base at Twentynine Palms, Calif, was analyzed using the techniques described in this paper. The results were compared to "ground truth" data obtained from GPS receivers on the vehicles.

Collaborative In-Network Processing for Target Tracking

Juan Liu, James Reich, and Feng Zhao

http://dx.doi.org/10.1155/S111086570321204X

This paper presents a class of signal processing techniques for collaborative signal processing in ad hoc sensor networks, focusing on a vehicle tracking application. In particular, we study two types of commonly used sensors—acoustic-amplitude sensors for target distance estimation and direction-of-arrival sensors for bearing estimation—and investigate how networks of such sensors can collaborate to extract useful information with minimal resource usage. The information-driven sensor collaboration has several advantages: tracking is distributed, and the network is energy-efficient, activated only on a when-needed basis. We demonstrate the effectiveness of the approach to target tracking using both simulation and field data.

Preprocessing in a Tiered Sensor Network for Habitat Monitoring

Hanbiao Wang, Deborah Estrin, and Lewis Girod

http://dx.doi.org/10.1155/S1110865703212087

We investigate task decomposition and collaboration in a two-tiered sensor network for habitat monitoring. The system recognizes and localizes a specified type of birdcalls. The system has a few powerful macronodes in the first tier, and many less powerful micronodes in the second tier. Each macronode combines data collected by multiple micronodes for target classification and localization. We describe two types of lightweight preprocessing which significantly reduce data transmission from micronodes to macronodes. Micronodes classify events according to their cross-zero rates and discard irrelevant events. Data about events of interest is reduced and compressed before being transmitted to macronodes for target localization. Preliminary experiments illustrate the effectiveness of event filtering and data reduction at micronodes.

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 5, 1 April 2003

Special Issue on
Advances in Modality-Oriented Medical Image Processing

Jiri Jan

Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 118, CZ 612 00 Brno, Czech Republic
Email: jan@feec.vutbr.cz

Milan Sonka

Department of Electrical and Computer Engineering, The University of Iowa, 4322 SC, Iowa City, IA 52242, USA
Email: milan-sonka@uiowa.edu

Ivo Provaznik

Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Purkynova 118, CZ 612 00 Brno, Czech Republic
Email: ivo@ieee.org

Modern medical imaging is perhaps the most progressive and also the most appreciated diagnostic tool in health care. Images provided by different imaging modalities correspond well to the background anatomical knowledge and are therefore well accepted and understood by the medical staff. The contribution of modern imaging to the progress of medicine and level of health care is thus widely recognised. While there is admirable progress in designing new or innovated imaging modalities often based on new physical principles, the overall success is equally due to the computational part of the imaging. All modern medical imaging modalities use image data in the digital form. Image reconstruction from incomprehensible projection data, their processing, noise and distortion removal, or various display methods matched to particular needs of diagnostics, all depend heavily on the computational aspects of medical imaging. The progress in medical imaging is thus in a great part a success of information processing, both on the side of algorithm design and implementation, as well as utilising large-scale-integration-based hardware. Without being too visible, computers are a substantial part of any modern imaging system and the specialised software forms a great deal of the system value, the corresponding algorithms being often the well guarded “family silver” of the imaging systems producing firms.

During the last two decades, the development in medical imaging was crucially conditioned by inclusion of complex digital processing of the measured raw image data. This has formed a great number of scientifically interesting problems and has lead to solutions reflecting the properties of image data provided by different medical imaging modalities. It has been recognised that utilising the knowledge of physical mechanisms behind each modality, or identifying modality-specific data properties, can significantly contribute to the efficiency of designed processing methods. The image analysis methods needed, for example, in tissue characterisation, are typically modality-dependent; at least in parameter selection but often requiring new specific approaches. On the other hand, the medical knowledge of anatomic structures can be used with an advantage during the analytic phase of processing, namely to improve the segmentation reliability and quality of visualisation. Another area typical for medical image processing concerns merging of multimodal data into consistent three-dimensional or (including time-dimension) four-dimensional data blocks enabling better diagnosis based on combining different-modality information. All in all, the list of contributions and application areas is virtually endless.

New methods or modifications of modality-oriented medical image processing are continuously designed and developed but the research reports are often rather scattered in the literature. This was the philosophy behind the decision to prepare a special issue of the EURASIP Journal on Applied Signal Processing devoted to this area. The invitation to submit papers describing advances in acquisition, restoration, reconstruction, segmentation, and visualisation of medical image data was thus formulated. Contributions were also considered dealing with the impact of the new data-processing methods to imaging process itself, and to the possibilities of clinical applications. Naturally, only a limited contribution in this direction can be expected from a single issue. Nevertheless, the contributors' response shows that the concept found its audience.

Out of twenty submitted papers, nine have been finally selected by the Guest Editors, taking into account the evaluations via standard international peer-review process. The selected papers cover a wide range of imaging modalities: primarily the magnetic resonance imaging (papers by Lethmate et al. and Positano et al.), X-ray CT (Púčik et al.), X-ray projection (Öktem et al. and Liang et al.), γ-ray SPECT imaging (Lundqvist et al.), and ultrasonic imaging (Argenti et al. and Mischi et al.); each dealing with a particular problem in data processing or interpreting. The paper by Yang et al. represents a method applied to different modality images. The classification of papers could be based on other viewpoints as well (e.g., mathematical background, medical application area, etc.) but the selected one is the most natural with respect to the special issue characterisation.

The Editors were primarily seeking high-quality research papers presenting methods evaluated against state-of-the-art solutions. Whether the goal was reached is for the reader to assess. If the conclusion is positive, the objective followed by the Guest Editors is fulfilled.

Jiri Jan
Milan Sonka
Ivo Provaznik

Volume 2003, No. 5, 1 April 2003

Contents and Abstracts

Dynamic MR-Imaging with Radial Scanning, a Post-Acquisition Keyhole Approach

Ralf Lethmate, Frank T. A. W. Wajer, Yannick Crémillieux, Dirk van Ormondt, and Danielle Graveron-Demilly

http://dx.doi.org/10.1155/S1110865703211197

A new method for 2D/3D dynamic MR-imaging with radial scanning is proposed. It exploits the inherent strong oversampling in the centre of *k*-space, which holds crucial temporal information of the contrast evolution. It is based on (1) a rearrangement of (novel 3D) *isotropic* distributions of trajectories during the scan according to the desired time resolution and (2) a post-acquisition *keyhole* approach. The 2D/3D dynamic images are reconstructed using 2D/3D-gridding and 2D/3D-IFFT. The scan time is not increased with respect to a *conventional* 2D/3D radial scan of the same image resolution, in addition one benefits from the dynamic information. An application to in vivo ventilation of rat lungs using hyperpolarized helium is demonstrated.

Automatic Characterization of Myocardial Perfusion in Contrast Enhanced MRI

Vincenzo Positano, Maria Filomena Santarelli, and Luigi Landini

http://dx.doi.org/10.1155/S1110865703211057

The use of contrast medium in cardiac MRI allows joining the high-resolution anatomical information provided by standard magnetic resonance with functional information obtained by means of the perfusion of contrast agent in myocardial tissues. The current approach to perfusion MRI characterization is the qualitative one, based on visual inspection of images. Moving to quantitative analysis requires extraction of numerical indices of myocardium perfusion by analysis of time/intensity curves related to the area of interest. The main problem in quantitative image sequence analysis is the heart movement, mainly due to patient respiration. We propose an automatic procedure based on image registration, segmentation of the myocardium, and extraction and analysis of time/intensity curves. The procedure requires a minimal user interaction, is robust with respect to the user input, and allows effective characterization of myocardial perfusion. The algorithm was tested on cardiac MR images acquired from voluntaries and in clinical routine.

CT Image Reconstruction Approaches Applied to Time-Frequency Representation of Signals

Jozef Púčik and Rami Oweis

http://dx.doi.org/10.1155/S1110865703211100

The mathematical formulation used in tomography has been successfully applied to time-frequency analysis, which represents an important "imaging modality" of the structure of

signals. Based on the interrelation between CT and time-frequency analysis, new methods have been developed for the latter. In this paper, an original method for constructing the time-frequency representation of signals from the squared magnitudes of their fractional Fourier transforms is presented. The method uses α-norm minimization with $\alpha \to 1$ which is motivated by Rényi entropy maximization. An iterative optimization method with adaptive estimation of the convergence parameter is elaborated. The proposed method exhibits advantages in the suppression of interference terms for signals with simple time-frequency configurations.

An Approach to Adaptive Enhancement of Diagnostic X-Ray Images

Hakan Öktem, Karen Egiazarian, Jarkko Niittylahti, and Juha Lemmetti

http://dx.doi.org/10.1155/S1110865703211069

Digital radiography is a popular diagnostic imaging method. Denoising and enhancement have an important potential in obtaining as much easily interpretable diagnostic information as possible with reasonable absorbed doses of ionising radiation. Due to the increasing usage of high resolution and high precision images with a limited number of human experts, the computational efficiency of the denoising and enhancement becomes important. In this paper, a local adaptive image enhancement and simultaneous denoising algorithm for fulfilling the requirements of digital X-ray image enhancement is introduced. The algorithm is based on modification of the wavelet transform coefficients by a pointwise nonlinear transformation and reconstructing the enhanced image from the modified wavelet transform coefficients. The implementation of algorithm in software is simple, quick, and universal.

Dynamic Chest Image Analysis: Model-Based Perfusion Analysis in Dynamic Pulmonary Imaging

Jianming Liang, Timo Järvi, Aaro Kiuru, Martti Kormano, and Erkki Svedström

http://dx.doi.org/10.1155/S1110865703212117

The "Dynamic Chest Image Analysis" project aims to develop model-based computer analysis and visualization methods for showing focal and general abnormalities of lung ventilation and perfusion based on a sequence of digital chest fluoroscopy frames collected with the dynamic pulmonary imaging technique. We have proposed and evaluated a multiresolutional method with an explicit ventilation model for ventilation analysis. This paper presents a new model-based method for pulmonary perfusion analysis. According to perfusion properties, we first devise a novel mathematical function to form a perfusion model. A simple yet accurate approach is further introduced to extract cardiac systolic and diastolic phases from the heart, so that this cardiac information may be utilized to accelerate the perfusion analysis and improve its sensitivity in detecting pulmonary perfusion abnormalities. This makes perfusion analysis not only fast but also robust in computation; consequently, perfusion analysis becomes computationally feasible without using contrast media. Our clinical case studies with 52 patients show that this technique is effective for pulmonary embolism even without using contrast media, demonstrating consistent correlations with

computed tomography (CT) and nuclear medicine (NM) studies. This fluoroscopical examination takes only about 2 seconds for perfusion study with only *low* radiation dose to patient, involving *no* preparation, *no* radioactive isotopes, and *no* contrast media.

Multilevel Wavelet Feature Statistics for Efficient Retrieval, Transmission, and Display of Medical Images by Hybrid Encoding

Shuyu Yang, Sunanda Mitra, Enrique Corona, Brian Nutter, and D. J. Lee

http://dx.doi.org/10.1155/S1110865703211203

Many common modalities of medical images acquire high-resolution and multispectral images, which are subsequently processed, visualized, and transmitted by subsampling. These subsampled images compromise resolution for processing ability, thus risking loss of significant diagnostic information. A hybrid multiresolution vector quantizer (HMVQ) has been developed exploiting the statistical characteristics of the features in a multiresolution wavelet-transformed domain. The global codebook generated by HMVQ, using a combination of multiresolution vector quantization and residual scalar encoding, retains edge information better and avoids significant blurring observed in reconstructed medical images by other well-known encoding schemes at low bit rates. Two specific image modalities, namely, X-ray radiographic and magnetic resonance imaging (MRI), have been considered as examples. The ability of HMVQ in reconstructing high-fidelity images at low bit rates makes it particularly desirable for medical image encoding and fast transmission of 3D medical images generated from multiview stereo pairs for visual communications.

A Combined Intensity and Gradient-Based Similarity Criterion for Interindividual SPECT Brain Scan Registration

Roger Lundqvist, Ewert Bengtsson, and Lennart Thurfjell

http://dx.doi.org/10.1155/S1110865703211239

An evaluation of a new similarity criterion for interindividual image registration is presented. The proposed criterion combines intensity and gradient information from the images to achieve a more robust and accurate registration. It builds on a combination of the normalised mutual information (NMI) cost function and a gradient-weighting function, calculated from gradient magnitude and relative gradient angle values from the images. An investigation was made to determine the best settings for the number of bins in the NMI joint histograms, subsampling, and smoothing of the images prior to the registration. The new method was compared with the NMI and correlation-coefficient (CC) criterions for interindividual SPECT image registration. Two different validation tests were performed, based on the displacement of voxels inside the brain relative to their estimated true positions after registration. The results show that the registration quality was improved when compared with the NMI and CC measures. The actual improvements, in one of the tests, were in the order of 30–40% for the mean voxel displacement error measured within 20 different SPECT images. A conclusion from the studies is that the new similarity measure significantly improves the registration quality, compared with the NMI and CC similarity measures.

Speckle Suppression in Ultrasonic Images Based on Undecimated Wavelets

Fabrizio Argenti and Gionatan Torricelli

http://dx.doi.org/10.1155/S1110865703211136

An original method to denoise ultrasonic images affected by speckle is presented. Speckle is modeled as a signal-dependent noise corrupting the image. Noise reduction is approached as a Wiener-like filtering performed in a shift-invariant wavelet domain by means of an adaptive rescaling of the coefficients of an undecimated octave decomposition. The scaling factor of each coefficient is calculated from local statistics of the degraded image, the parameters of the noise model, and the wavelet filters. Experimental results demonstrate that excellent background smoothing as well as preservation of edge sharpness and fine details can be obtained.

Videodensitometric Methods for Cardiac Output Measurements

Massimo Mischi, Ton Kalker, and Erik Korsten

http://dx.doi.org/10.1155/S1110865703211185

Cardiac output is often measured by indicator dilution techniques, usually based on dye or cold saline injections. Developments of more stable ultrasound contrast agents (UCA) are leading to new noninvasive indicator dilution methods. However, several problems concerning the interpretation of dilution curves as detected by ultrasound transducers have arisen. This paper presents a method for blood flow measurements based on UCA dilution. Dilution curves are determined by real-time densitometric analysis of the video output of an ultrasound scanner and are automatically fitted by the Local Density Random Walk model. A new fitting algorithm based on multiple linear regression is developed. Calibration, that is, the relation between videodensity and UCA concentration, is modelled by in vitro experimentation. The flow measurement system is validated by in vitro perfusion of SonoVue contrast agent. The results show an accurate dilution curve fit and flow estimation with determination coefficient larger than 0.95 and 0.99, respectively.

EURASIP Journal on Applied Signal Processing, Volume 2003, No. 6, 1 May 2003

Special Issue on Rapid Prototyping of DSP Systems

Magdy Bayoumi

The Center for Advanced Computer Studies, University of Louisiana at Lafayette, Lafayette, LA 70504-4330, USA
Email: mbayoumi@cacs.louisiana.edu

Shuvra S. Bhattacharyya

Department of Electrical and Computer Engineering and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA
Email: ssb@eng.umd.edu

Rudy Lauwereins

Interuniversity Micro Electronics Center (IMEC), Kapeldreef 75, B-3001 Leuven, Belgium
Email: rudy.lauwereins@imec.be

With the increasing complexity of applications, rapid evolution of technology, and intense market competition in DSP consumer markets, the ability to quickly take a product concept to a working hardware/software demonstration is critical to the DSP industry. Key technologies required to meet this challenge include new types of programmable components that offer novel trade-offs between flexibility and efficiency, models for exchange of intellectual property, and computer aided design techniques for simulation, synthesis, verification, and integration of complex systems. The prototyping of modern DSP systems is especially complicated by increasing levels of application dynamics, and complex physical constraints. The papers in this special issue span a broad range of topics related to the rapid prototyping of DSP systems.

The first paper, by Tanougast et al., develops an approach for dynamic reconfiguration of FPGA implementations. The authors apply a temporal partitioning approach to the application dataflow graph with the objective of minimizing the FPGA resources required to meet a given performance constraint. Significant improvements in efficiency are demonstrated on a number of image processing applications.

The next paper, by Kuusilinna et al., describes the design of a large-scale emulation engine and an application example from the field of low-power wireless devices. The primary goal of the emulator is to support design space exploration of real-time algorithms. The emulator is customized for dataflow dominant architectures especially focusing on telecommunication related applications. The paper proves that real-time emulation of a low-power TDMA receiver is feasible at a clock speed of 25 MHz.

The third paper, by Oh and Ha, presents a software synthesis technique for minimizing buffer memory requirements of multimedia applications. This technique starts with a given schedule of the application dataflow graph, and is based on decomposing buffering requirements into global buffers and local pointers into these buffers. Methods for sharing global buffers and efficiently managing the local pointer buffers are developed in this framework. Large reductions in memory requirements are reported when applying these techniques to a JPEG encoder and an H.263 encoder.

The next paper, by Zhang and Parhi, focuses on an FPGA implementation of a (3,6)-regular low-density parity-check code (LDPC) decoder. In the past few years, the recently rediscovered LDPC codes have received a lot of attention and have been widely considered as next-generation error-correcting codes for telecommunication and magnetic storage. Unfortunately, the direct fully parallel decoder implementation usually incurs too high hardware complexity for many real applications. Hence, partly parallel decoder design approaches that can achieve appropriate trade-offs between hardware complexity and decoding throughput are highly desirable. Applying a joint code and decoder design methodology, this paper develops a high-speed (3,6)-regular LDPC code partly parallel decoder architecture. This implementation supports a maximum symbol throughput of 54 Mbps and achieves a BER of 10^{-6} at 2 dB.

In the paper by Fox and Turner, FPGA is used as rapid development platform for DCT approximations. The approximations are used to control the coding gain, MSE, quantization noise, hardware cost, and power consumption by optimizing the coefficient values and the datapath world lengths. Lagrangian local search method is used to optimize the coefficients for the desired controlled parameters. Xilinix FPGA is used to rapidly prototype the DCT architecture with near optimal coding gain. The developed design methodology allows trade-offs among coding gain, hardware cost, and power consumption.

The next paper, by Carreira et al., describes an approach for FPGA-based design of FIR filters using the method of peak-constrained least squares for controlling the frequency response. The approach allows one to trade off passband-to-stopband energy ratio with FPGA resource requirements without altering the minimum stopband attenuation. Implementation of the approach utilizes JBits, which is a Java interface for controlling the configuration bitstream of Xilinx FPGAs.

In the next paper, by Spivey et al., a framework for rapid prototyping FPGA-based systems is presented. The framework “logic foundry” targets the integration of various DSP architecture design levels (or modules) and components. Four areas of integration: design flow integration, component integration, platform integration, and software integration are presented. Experimental results on Xilinix FPGA of an incrementer design and turbo decoder design show the use of the logic foundry for rapid prototyping. The framework is very flexible; it can be used as an integrated design environment or different modules can be used as stand-alone tools that can be integrated with other environments.

The paper by Madahar et al. presents case studies of adaptive beamformer applications for radar and sonar using a design environment and rapid prototyping methodology developed under the ESPADON (Environment for Signal Processing and Rapid Prototyping) project. ESPADON builds on existing tools, including Ptolemy Classic, GEDAE, and Handel-C, to provide an integrated process for reducing the cost and development time involved in implementing military signal processing applications. The paper focuses on demonstrating the productivity gains achieved through the ESPADON approach.

The next paper, by Bednara et al., develops a design method for mapping digital linear controllers with large signal processing requirements into efficient FPGA implementations. A case study of an inverse pendulum controller is used to illustrate the ideas and demonstrate the advantages of the proposed design techniques compared to software implementation.

The last paper, by Jones and Cavallaro, addresses one of the main challenges in developing design methodology for rapid prototyping which is the transition from simulation to a working prototype of a system. The developed design methodology is based on using an appropriate design language to bridge the gap between simulation and prototyping. Such an approach combines the strengths of simulation and prototyping, allowing the designer to develop and evaluate the target system partly in simulation on a host computer and partly as a prototype on embedded hardware. Several software tools have been developed for implementing the proposed design methodology. It has been successfully used in the development of a next-generation code division multiple access (CDMA) cellular wireless communication system.

We would like to thank all of the people who have contributed to this special issue, including all of the authors who submitted papers; the Editorial Board for their encouragement of the special issue; and the reviewers for all of their efforts, without which this special issue would not be possible.

Magdy Bayoumi
Shuvra S. Bhattacharyya
Rudy Lauwereins

Volume 2003, No. 6, 1 May 2003

Contents and Abstracts

A Partitioning Methodology That Optimises the Area on Reconfigurable Real-Time Embedded Systems

Camel Tanougast, Yves Berviller, Serge Weber, and Philippe Brunet

http://dx.doi.org/10.1155/S1110865703212051

We provide a methodology used for the temporal partitioning of the data-path part of an algorithm for a reconfigurable embedded system. Temporal partitioning of applications for reconfigurable computing systems is a very active research field and some methods and tools have already been proposed. But all these methodologies target the domain of existing reconfigurable accelerators or reconfigurable processors. In this case, the number of cells in the reconfigurable array is an implementation constraint and the goal of an optimised partitioning is to minimise the processing time and/or the memory bandwidth requirement. Here, we present a strategy for partitioning and optimising designs. The originality of our method is that we use the dynamic reconfiguration in order to minimise the number of cells needed to implement the data path of an application under a time constraint. This approach can be useful for the design of an embedded system. Our approach is illustrated by a reconfigurable implementation of a real-time image processing data path.

Designing BEE: A Hardware Emulation Engine for Signal Processing in Low-Power Wireless Applications

Kimmo Kuusilinna, Chen Chang, M. Josephine Ammer, Brian C. Richards, and Robert W. Brodersen

http://dx.doi.org/10.1155/S1110865703212154

This paper describes the design of a large-scale emulation engine and an application example from the field of low-power wireless devices. The primary goal of the emulator is to support design space exploration of real-time algorithms. The emulator is customized for dataflow dominant architectures, especially focusing on telecommunication-related applications. Due to its novel routing architecture and application-specific nature, the emulator is capable of real-time execution of a class of algorithms in its application space. Moreover, the dataflow structure facilitates the development of a highly abstracted design flow for the emulator. Simulations and practical measurements on commercial development boards are used to verify that real-time emulation of a low-power TDMA receiver is feasible at a clock speed of 25 MHz.

Memory-Optimized Software Synthesis from Dataflow Program Graphs with Large Size Data Samples

Hyunok Oh and Soonhoi Ha

http://dx.doi.org/10.1155/S1110865703212130

In multimedia and graphics applications, data samples of nonprimitive type require significant amount of buffer memory. This paper addresses the problem of minimizing the buffer memory requirement for such applications in embedded software synthesis from graphical dataflow programs based on the synchronous dataflow (SDF) model with the given execution order of nodes. We propose a memory minimization technique that separates global memory buffers from local pointer buffers: the global buffers store live data samples and the local buffers store the pointers to the global buffer entries. The proposed algorithm reduces 67% memory for a JPEG encoder, 40% for an H.263 encoder compared with unshared versions, and 22% compared with the previous sharing algorithm for the H.263 encoder. Through extensive buffer sharing optimization, we believe that automatic software synthesis from dataflow program graphs achieves the comparable code quality with the manually optimized code in terms of memory requirement.

An FPGA Implementation of $(3,6)$-Regular Low-Density Parity-Check Code Decoder

Tong Zhang and Keshab K. Parhi

http://dx.doi.org/10.1155/S1110865703212105

Because of their excellent error-correcting performance, low-density parity-check (LDPC) codes have recently attracted a lot of attention. In this paper, we are interested in the practical LDPC code decoder hardware implementations. The direct fully parallel decoder implementation usually incurs too high hardware complexity for many real applications, thus partly parallel decoder design approaches that can achieve appropriate trade-offs between hardware complexity and decoding throughput are highly desirable. Applying a joint code and decoder design methodology, we develop a high-speed $(3,k)$-regular LDPC code partly parallel decoder architecture based on which we implement a 9216-bit, rate-1/2 $(3,6)$-regular LDPC code decoder on Xilinx FPGA device. This partly parallel decoder supports a maximum symbol throughput of 54 Mbps and achieves BER 10^{-6} at 2 dB over AWGN channel while performing maximum 18 decoding iterations.

Rapid Prototyping of Field Programmable Gate Array-Based Discrete Cosine Transform Approximations

Trevor W. Fox and Laurence E. Turner

http://dx.doi.org/10.1155/S1110865703301027

A method for the rapid design of field programmable gate array (FPGA)-based discrete cosine transform (DCT) approximations is presented that can be used to control the coding gain, mean square error (MSE), quantization noise, hardware cost, and power consumption by optimizing the coefficient values and datapath wordlengths. Previous DCT design meth-

ods can only control the quality of the DCT approximation and estimates of the hardware cost by optimizing the coefficient values. It is shown that it is possible to rapidly prototype FPGA-based DCT approximations with near optimal coding gains that satisfy the MSE, hardware cost, quantization noise, and power consumption specifications.

A Methodology for Rapid Prototyping Peak-Constrained Least-Squares Bit-Serial Finite Impulse Response Filters in FPGAs

Alex Carreira, Trevor W. Fox, and Laurence E. Turner

http://dx.doi.org/10.1155/S1110865703301015

Area-efficient peak-constrained least-squares (PCLS) bit-serial finite impulse response (FIR) filter implementations can be rapidly prototyped in field programmable gate arrays (FPGA) with the methodology presented in this paper. Faster generation of the FPGA configuration bitstream is possible with a new application-specific mapping and placement method that uses JBits to avoid conventional general-purpose mapping and placement tools. JBits is a set of Java classes that provide an interface into the Xilinx Virtex FPGA configuration bitstream, allowing the user to generate new configuration bitstreams. PCLS coefficient generation allows passband-to-stopband energy ratio (PSR) performance to be traded for a reduction in the filter's hardware cost without altering the minimum stopband attenuation. Fixed-point coefficients that meet the frequency response and hardware cost specifications can be generated with the PCLS method. It is not possible to meet these specifications solely by the quantization of floating-point coefficients generated in other methods.

Logic Foundry: Rapid Prototyping for FPGA-Based DSP Systems

Gary Spivey, Shuvra S. Bhattacharyya, and Kazuo Nakajima

http://dx.doi.org/10.1155/S1110865703301039

We introduce the Logic Foundry, a system for the rapid creation and integration of FPGA-based digital signal processing systems. Recognizing that some of the greatest challenges in creating FPGA-based systems occur in the integration of the various components, we have proposed a system that targets the following four areas of integration: design flow integration, component integration, platform integration, and software integration. Using the Logic Foundry, a system can be easily specified, and then automatically constructed and integrated with system level software.

How Rapid Is Rapid Prototyping? Analysis of ESPADON Programme Results

Bob K. Madahar, Ian D. Alston, Denis Aulagnier, Hans Schurer, Mark Thomas, and Brigitte Saget

http://dx.doi.org/10.1155/S1110865703212129

New methodologies, engineering processes, and support environments are beginning to emerge for embedded signal processing systems. The main objectives are to enable defence

industry to field state-of-the-art products in less time and with lower costs, including retrofits and upgrades, based predominately on commercial off the shelf (COTS) components and the model-year concept. One of the cornerstones of the new methodologies is the concept of rapid prototyping. This is the ability to rapidly and seamlessly move from functional design to the architectural design to the implementation, through automatic code generation tools, onto real-time COTS test beds. In this paper, we try to quantify the term "rapid" and provide results, the metrics, from two independent benchmarks, a radar and sonar beamforming application subset. The metrics show that the rapid prototyping process may be sixteen times faster than a conventional process.

Design and Implementation of Digital Linear Control Systems on Reconfigurable Hardware

Marcus Bednara, Klaus Danne, Markus Deppe, Oliver Oberschelp, Frank Slomka, and Jürgen Teich

http://dx.doi.org/10.1155/S1110865703301040

The implementation of large linear control systems requires a high amount of digital signal processing. Here, we show that reconfigurable hardware allows the design of fast yet flexible control systems. After discussing the basic concepts for the design and implementation of digital controllers for mechatronic systems, a new general and automated design flow starting from a system of differential equations to application-specific hardware implementation is presented. The advances of reconfigurable hardware as a target technology for linear controllers is discussed. In a case study, we compare the new hardware approach for implementing linear controllers with a software implementation.

A Rapid Prototyping Environment for Wireless Communication Embedded Systems

Bryan A. Jones and Joseph R. Cavallaro

http://dx.doi.org/10.1155/S111086570330304X

This paper introduces a rapid prototyping methodology which overcomes important barriers in the design and implementation of digital signal processing (DSP) algorithms and systems on embedded hardware platforms, such as cellular phones. This paper describes rapid prototyping in terms of a simulation/prototype bridge and in terms of appropriate language design. The simulation/prototype bridge combines the strengths of simulation and of prototyping, allowing the designer to develop and evaluate next-generation communications systems, partly in simulation on a host computer and partly as a prototype on embedded hardware. Appropriate language design allows designers to express a communications system as a block diagram, in which each block represents an algorithm specified by a set of equations. Software tools developed for this paper implement both concepts, and have been successfully used in the development of a next-generation code division multiple access (CDMA) cellular wireless communications system.

Special Issue on

Cross Layer Design for Communications and Signal Processing Systems

CALL FOR PAPERS

An important aspect of wireless networks is dynamic behavior. The conventional protocol structure is inflexible as various protocol layers can only communicate in a strict manner. In such a case, the layers are designed to operate under the worst conditions, rather than adapting to changing conditions. This leads to inefficient use of spectrum and energy.

Adaptation represents the ability of network protocols and applications to observe and respond to the channel variation. Central to adaptation is the concept of cross layer design. In general, cross layer design involves four key layers in the overall protocol stack (i.e., application-layer, transport-layer, network-layer and link-layer). The application can adjust its behavior, e.g., its flow rate or the amount of overhead devoted to error resilience, according to the changing network and channel conditions. The adaptation can also take place in the underlying layers such as TCP and UDP, so that the application originally developed for different networks remain unchanged. Information derived from the application, such as its QoS requirements and the priorities of the packets it produces, could be used in coordinating the behavior of the lower layers for resource efficiency. For example, the persistence level of the link layer ARQ mechanism should be varied according to each application's latency and reliability requirements, as well as the traffic load. Another essential factor to adaptation is each layer's ability to estimate the current and even predict the future network and channel conditions and exchange the information across different layers.

This special issue will be devoted to the latest developments in the field of cross layer design, where the emphasis is on interactions among different network layers, so as to improve the performance of communication and signal processing systems. We seek original, previously unpublished, and completed contributions, not currently under review by another journal. Contributions emphasizing recent progress and new research directions are strongly encouraged. Papers are invited from, but not limited to, the following topics:

- Architectures and methodologies for cross layer design in wireless and wireline communications
- Energy efficiency in wireless ad hoc networks
- Dynamic resource allocation and quality of service
- Scheduling algorithms and link adaptation

- Signal processing and medium access control
- Channel adaptive routing protocols
- Interaction between TCP, UDP, and RLP layers
- Channel-aware applications
- Joint source and channel design in wireless multimedia communications
- Signaling design for cross layer interaction

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system at http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	July 31, 2003
Acceptance Notification	February 28, 2004
Final Manuscript Due	April 30, 2004
Publication Date	4th Quarter, 2004

GUEST EDITORS:

Antonio Ortega, Department of Electrical Engineering, University of South California, Los Angeles, CA 90036, USA; ortega@sipi.usc.edu

Lang Tong, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA; ltong@ece.cornell.edu

Haitao Zheng, Bell Laboratory, Lucent Technologies, Holmdel, NJ 07733, USA; haitaoz@lucent.com

Michele Zorzi, Dipartimento di Ingegneria, Universita'di Ferrara, Ferrara, Italy; mzorzi@ing.unife.it

EDITORIAL BOARD REPRESENTATIVE:

Bastiaan Kleijn, Royal Institute of Technology, Sweden

Special Issue on

Improved CDMA Detection Techniques for Future Wireless Systems

CALL FOR PAPERS

The past few years have been marked by a worldwide standardization activity for third generation (3G) wireless systems, which are intended to deliver high data rates and are expected to handle multimedia applications in addition to voice. The key multiple access technique that has been chosen for 3G wireless systems is CDMA.

The performance of 3G wireless systems might not be sufficient to meet the needs of future high-performance multimedia applications such as full-motion video and teleconferencing. Hence, there will be a need for systems that extend the capabilities of 3G wireless systems, sometimes referred to as fourth generation (4G) wireless systems, whatever they might be. It is likely that these wireless systems will retain a CDMA component, but compared to 3G wireless systems, the changed operating conditions will present a new set of challenges in the development of CDMA detection techniques.

The aim of this special issue is to cover present research in the development of improved CDMA detection techniques for future wireless systems. Prospective papers should be unpublished and present novel, fundamental research offering innovative contributions to the wireless communications community.

Topics of interest include (but are not limited to):

- Spreading sequence design
- Quasi-synchronous CDMA systems
- Multicarrier CDMA systems
- CDMA performance evaluation
- Acquisition and tracking
- Channel estimation
- Multiuser detection
- Interference reduction techniques
- Smart antennas for CDMA
- Space-time coding for CDMA
- Power control
- (Soft/softer) handoff issues

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system at http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	August 1, 2003
Acceptance Notification	December 1, 2003
Final Manuscript Due	March 15, 2004
Publication Date	3rd Quarter, 2004

GUEST EDITORS:

Geert Leus, Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium; leus@esat.kuleuven.ac.be

Philippe Loubaton, Equipe Signal pour les Communications et URA820 CNRS/ENST, Institut Gaspard Monge/Electronique, Université de Marne la Vallee, Cite Descartes, 5 boulevard Descartes, Champs sur Marne, 77454 Marne la Vallee Cedex 2, France; loubaton@univ-mlv.fr

Dirk Slock, Eurécom Institute, 2229 routes des Cretes, BP 193, 06904 Sophia Antipolis Cedex, France; slock@eurecom.fr

Michael D. Zoltowski, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285, USA; mikedz@ecn.purdue.edu

EDITORIAL BOARD REPRESENTATIVE:

Dirk Slock, Eurécom Institute, 2229 routes des Cretes, BP 193, 06904 Sophia Antipolis Cedex, France; slock@eurecom.fr

Special Issue on

Nonlinear Signal and Image Processing

CALL FOR PAPERS

While the field of signal and image processing has matured within the framework of linear systems, novel areas of nonlinear signal processing continue to appear. This, due to the fact that the physics of image formation are inherently nonlinear and that the stochastic components of signals and images are inherently nonstationary and non-Gaussian, for which linear processing is not optimum.

The goal of this special issue is to bring together the latest advances in the areas of nonlinear signal and image processing as represented in the 2003 NSIP Workshop to be held in Grado, Italy. We encourage participants of the workshop to extend their conference paper contributions into full pages for submission to this special issue. Prospective manuscripts should be unpublished in journals and should present innovative contributions either from a methodological or applications point of view.

Areas of interest are targeted to advanced nonlinear signal and image processing (but not limited to):

- Lower and higher order statistics
- Median and order-statistics signal processing
- Mathematical morphology
- Multichannel and array nonlinear signal processing
- Nonlinear time-frequency methods
- Nonlinear methods for communications and networking
- Signal processing in bio-computing
- Polynomial signal processing structures
- Image and video signal processing applications
- Halftoning and digital printing
- Image, video, audio, and multimedia applications
- Signal processing for security, authentication, and cryptography

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	September 1, 2003
Acceptance Notification	November 1, 2003
Final Manuscript Due	December 1, 2003
Publication Date	2nd Quarter, 2004

GUEST EDITORS:

Gian Luca Foresti, Department of Computer Science, University of Udine, 33100 Udine, Italy; foresti@dimi.uniud.it

Gianni Ramponi, Signal and Image Processing, University of Trieste, 34127 Trieste, Italy; ramponi@univ.trieste.it

Carlo S. Regazzoni, Department of Biophysical and Electronic Engineering, University of Genoa, 16145 Genoa, Italy; carlo@dibe.unige.it

Giovanni Sicuranza, Signal and Image Processing, University of Trieste, 34127 Trieste, Italy; sicuranza@univ.trieste.it

Gianni Vernazza, Department of Biophysical and Electronic Engineering, University of Genoa, 16145 Genoa, Italy; vernazza@dibe.unige.it

Special Issue on Turbo Processing

CALL FOR PAPERS

Turbo codes first appeared in 1993. Although limited at the beginning to coding, the idea of exchanging soft information between SISO (soft-input/soft-output) modules has subsequently been applied to other elements of a digital communications receiver, leading to the nowadays famous turbo principle. Due to their excellent performance, turbo codes are already used in a number of standards, and have become a very active area of research.

The goal of this special issue is therefore to present recent research results on the theory and applications of the "turbo-principle."

Topics of interest include (but are not limited to):

- Design of turbo codes
- Performance and bounds of turbo codes
- Design of SISO algorithms
- Modelling of turbo/iterative processing
- Turbo detection/equalization for time dispersive channels
- Turbo joint detection for multiuser communications
- Turbo space-time coding
- Turbo reception for MIMO systems
- Turbo synchronization
- Turbo demodulation
- Joint source-channel decoding based on soft information
- Implementation issues
- Applications and standards

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system at http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	October 1, 2003
Acceptance Notification	March 1, 2004
Final Manuscript Due	July 1, 2004
Publication Date	4th Quarter, 2004

GUEST EDITORS:

Alex M. Haimovich, Professor of Electrical and Computer Engineering, Director, New Jersey Center for Wireless Telecommunications, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA; haimovic@njit.edu

Ramesh Pyndiah, Signal & Communications Department, ENST BRETAGNE Technopole Brest IROISE, BP832-29285 BREST Cedex, France; ramesh.pyndiah@enst-bretagne.fr

Luc Vandendorpe, Communications Laboratory, Université Catholique de Louvain, B-1348 Louvain la Neuve, Belgium; vandendorpe@tele.ucl.ac.be

EDITORIAL BOARD REPRESENTATIVE:

Marc Moonen, Department of Electrical Engineering, Katholieke Universiteit Leuven, ESAT-SISTA, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium; marc.moonen@esat.kuleuven.ac.be

Special Issue on UWB—State of the Art

CALL FOR PAPERS

Recent standardisation outcomes substantiate the potential impact of Ultra-Wideband (UWB) systems, and motivate the development of UWB products for the mass-market. In this light the moment seems appropriate for summarising the state of the art in UWB methods and technologies, and for setting an overview covering open issues in this broad research and application area. In particular, coexistence with other wireless standards plays a crucial role. Moreover, broadband antenna and receiver design is a real challenge, as well as the development of sophisticated algorithms for dense multipath environment, synchronization and several other topics such as channel and interference modelling. A low-complexity, low-power, and low-cost system solution may be far away.

The aim of this special issue is to present recent research in UWB systems and technology from multiple points of view with emphasis on future applications in the area of wireless communications. Prospective papers should be unpublished and present novel, fundamental research offering innovative contributions either from a methodological or an application perspective.

Topics of interest include (but are not limited to):

- Channel modeling and measurement
- Channel estimation and equalisation
- Synchronisation
- Modulation and multiple access
- Interference and coexistence issues
- Pulse shaping and filtering
- Antenna and receiver design
- Information theory
- Ad hoc networks and sensor networks
- Routing and MAC design
- Standardisation
- Applications
- UWBMIMO
- Multiband UWB

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system at http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	October 1, 2003
Acceptance Notification	March 1, 2004
Final Manuscript Due	July 1, 2004
Publication Date	4th Quarter, 2004

GUEST EDITORS:

Gabriella di Benedetto, INFOCOM Department, Engineering School, University of Rome La Sapienza, Via Eudossiana 18, 00184 Rome, Italy; gaby@acts.ing.uniroma1.it

Thomas Kaiser, Department of Comm. Systems, Faculty of Engineering, University of Duisburg-Essen, Bismarckstr. 81, 47057 Duisburg, Germany; thomas.kaiser@uni-duisburg.de

Norbert Schmidt, IMST GmbH, Information and Communications Systems, Mobilfunksysteme, Carl-Friedrich-Gauß Str. 2, 47475 Kamp-Lintfort, Germany; schmidt@imst.de

Armin Wittneben, ETH, Swiss Federal Institute of Technology, Zurich, Communication Technology Laboratory, Sternwartstrasse 7, CH-8092 Zurich, Switzerland, wittneben@nari.ee.ethz.ch

EDITORIAL BOARD REPRESENTATIVE:

Mos Kaveh, Department of Electrical and Computer Engineering, University of Minnesota, 200 Union Street, SE Minneapolis, MN 55455, USA; kaveh@ece.umn.edu

Special Issue on

Anthropomorphic Processing of Audio and Speech

CALL FOR PAPERS

Anthropomorphic systems process signals "at the image of man." They are designed to solve a problem in signal processing by imitation of the processes that accomplish the same task in humans. In the area of audio and speech processing, remarkable successes have been obtained by anthropomorphic systems: perceptual audio coding even caused an MP3 hype.

At first sight, it could seem obvious that the performance of audio processing systems should benefit from taking into account the perceptual properties of human audition. For example, front-ends that extract perceptually meaningful features currently show the best results in speech recognizers. However, their features are typically used for a stochastic optimization that is itself not anthropomorphic at all. Thus, it is not obvious why they should perform best, and perhaps the truly optimal features have not yet been found because, after all, "airplanes do not flap their wings."

In general, we believe that there are several situations when an anthropomorphic approach may not be the best solution. First, its combination with nonanthropomorphic systems could result in a suboptimal overall performance (the quantization noise that was cleverly concealed by a perceptual audio coder could become unmasked by subsequent linear or nonlinear processing). Second, other than anthropomorphic approaches might be better adapted to the technology that is chosen for the implementation (airplanes do not flap their wings because it is technically much more efficient to use jet engines for propulsion). Nevertheless, a lot can be learned from imitating natural systems. As such anthropomorphic and, by extension, biomorphic systems can be considered to play an important role in the process of developing new technologies.

The aim of this special issue is to bring together papers from different areas of audio and speech processing that deal with aspects of anthropomorphic processing or in which an anthropomorphic or perceptual approach was taken. Papers with a research nature, review papers, and tutorial papers will be considered, provided that they are unpublished.

Topics of interest include (but are not limited to):

- Speech and audio coding
- Audio measurements and speech analysis
- Objective quality measures for audio and speech
- Speech synthesis (rule-based, articulatory, ...)

- Audio virtual reality
- Content-based music search
- Music and instrument recognition
- Audio classification and retrieval
- Speech and speaker recognition

Authors should follow the EURASIP JASP manuscript format described at the journal site http://asp.hindawi.com/. Prospective authors should submit an electronic copy of their complete manuscript through the EURASIP JASP's web submission system at http://mts.hindawi.com/asp/, according to the following timetable.

Manuscript Due	November 1, 2003
Acceptance Notification	April 1, 2004
Final Manuscript Due	August 1, 2004
Publication Date	4th Quarter, 2004

GUEST EDITORS:

Werner Verhelst, Vrije Universiteit Brussel, Belgium; wverhels@etro.vub.ac.be
Jürgen Herre, Fraunhofer IIS-A, Germany; hrr@iis.fhg.de
Gernot Kubin, Technical University Graz, Austria; g.kubin@ieee.org
Hynek Hermansky, Oregon Health & Science University, USA; hynek@ece.ogi.edu

EDITORIAL BOARD REPRESENTATIVE:

Soeren Hold Jensen, Aalborg University, Fredrik Bajers Vej 7, A3, DK-9220 Aalborg Oest, Denmark; shj@cpk.auc.dk

How to Become a Member of EURASIP

EURASIP membership is open to all persons and institutions active or interested in signal processing within Europe or outside. Membership benefits include: free subscription to EURASIP Newsletter, reduced subscription to the journals "Signal Processing," "Speech Communication," "Signal Processing: Image Communication," "EURASIP Journal on Applied Signal Processing," and "EURASIP Journal on Wireless Communications and Networking", reduced fees for conferences and workshops that are sponsored or co sponsored by the Association. reduced fees for EURASIP's short courses.

Personal (individual) membership must be paid by personal funds. To validate their category, students should provide an endorsement from school of ricials stating that they are enrolled in regular academic programs. Please send an application letter containing your name, title, position, company/institution, full mailing address, country, phone number, telefax number and the desired journals to: EURASIP, European Association for Signal, Speech, and Image Processing, P.O. Box 134, CH-1000 Lausanne 13, Switzerland, together with your remittance.

Your payments to EURASIP can be made in any of the following methods. In all cases Euro is the only acceptable currency:

1. Using credit cards: MasterCard, American Express or Visa only. Please fill in your Card No., Expiring date and signature.
2. Using cheques drawn in Euro mailed to EURASIP.
3. Using international money orders drawn in Euro mailed to EURASIP.
4. Requesting your bank to write transfer funds including your name to EURASIP at the Swiss Credit Bank Lausanne, Switzerland, account no. 322.294-31.
5. Using international postal money transfer to EURASIP postal account no. 10-3279-8 in Lausanne, Switzerland.

For membership inquiries: Fabienne Vionnet, sigpro@epfl.ch

2003 Membership Fees in Euros for EURASIP Members

Institution: 70€; Individual: 30€; Student/retired: 20€

2003 Print Journal Prices for EURASIP Members

	Institution	Member
Signal Processing	1935€	95€
Speech Communication	755€	70€
Signal Processing: Image Communication	625€	70€
EURASIP Journal on Applied Signal Processing	495€	60€ (Student: 30€)

The 495 Euro Institution subscription for EURASIP Journal on Applied Signal Processing includes electronic access.

2003 Electronic Journal Access for EURASIP Members

	Institution	Member
Signal Processing	—	70€
Speech Communication	—	70€
Signal Processing: Image Communication	—	70€
EURASIP Journal on Applied Signal Processing	396€	60€ (Student: 30€)

Please remember to mention your name and EURASIP membership number on all payment forms.

EURASIP MEMBERSHIP APPLICATION

Please check:

- ☐ I wish to become a EURASIP member from now on.
- ☐ This institution wishes to become an institutional EURASIP member from now on.
- ☐ I enclose a copy of my student/retirement certificate to be entitled to pay reduced fees.

I/We wish to subscribe to the journal(s):

- ☐ Signal Processing
- ☐ Speech Communication
- ☐ Image Communication
- ☐ EURASIP Journal on Applied Signal Processing

Name: . . . .

Institution (if applicable): . . . .

Address: . . . .

. . . .

. . . .

Please mail to:

EURASIP, European Association for Signal, Speech, and Image Processing,
P.O. Box 134; CH-1000 Lausanne 13, Switzerland

The AdCom invites EURASIP members to send in to the Newsletter Editor appropriate announcements that they wish to be communicated to the Signal, Speech, and Image Processing community.

The Editor reserves the right to edit these to fit into the available space as appropriate for the next newsletter publication.