21. Digital Signal Processing

Academic and Research Staff

Prof. A.V. Oppenheim, Prof. A.B. Baggeroer, Prof. J.S. Lim, Prof. B.R. Musicus, Dr. D.R. Mook, Dr. G.L. Duckworth

Graduate Students

T.E. Bordley, P. Chan, S.R. Curtis, D.S. Deadrick, W.P. Dove, F.U. Dowla, D.W. Griffin, W.A. Harrison, D. Izraelevitz, D.M. Martinez, E.E. Milios, C. Myers, T.N. Pappas, M.D. Richard, H. Sekiguchi, R. Sundaram, P.L. Van Hove, M.S. Wengrovitz

21.1 Introduction

The Digital Signal Processing Group is carrying out research in the general area of digital signal processing. While a major part of our activities focus on the development of new algorithms, there is a strong conviction that theoretical developments must be closely tied to applications and to issues of implementation. The application areas which we deal with principally are speech, image and geophysical data processing. In addition to specific projects being carried out on campus, there is close interaction with Lincoln Laboratory and with the Woods Hole Oceanographic Institution.

In the area of speech processing, we have over the past several years worked on the development of systems for bandwidth compression of speech, parametric modeling of speech using pole–zero models, time–scale modification of speech and enhancement of degraded speech. Recently we have obtained some important new results on time–scale modification of speech, growing out of a more general set of issues involving the estimation of a signal after its short–time Fourier transform has been modified. We are also exploring new techniques for speech enhancement using adaptive noise cancelling when multiple microphones are available.

There are also a number of projects related to image processing that we are currently pursuing. One project is restoration of images degraded by additive noise, multiplicative noise, and convolutional noise. Out of this project, we have developed a new image restoration system which is applicable to restoring images degraded by various different types of degradation. Our current work in this project involves development of new image restoration systems by exploiting additional available information such as the range map in infrared radar images. Another project is development of new image coding techniques by reducing quantization noise in PCM image coding or by reducing blocking effect in cosine transform image coding. Our approach to first decorrelate the quantization noise, and then reduce the quantization noise by a noise reduction system, led to a noticeable improvement in the performance of a simple PCM image coding system. We are currently working on the extension of these results to a more complex PCM image coding system. To reduce the blocking

effect in cosine transform image coding, we have studied two approaches. In one approach, the coder is modified to account for the blocking effect and in the second approach, the coded image with blocking effect is processed to reduce the blocking effect. In both approaches, we have developed specific algorithms that significantly reduce the blocking effect in cosine transform coding. Another project that we are currently exploring is the development of a very low bit rate (below 50 kbits/sec) video-conferencing system. The specific approach we are currently studying is to model a human face, which is a regular feature in typical video-conferencing applications, with a set of parameters and then synthesize the image at the receiver from the coded parameters. This approach is analogous to modeling human speech for speech coding, which led to significant bit rate reduction without seriously degrading the speech intelligibility.

In the area of geophysical data processing, there are a variety of on-going and new projects. During March–May 1980, we led a large acoustics and geophysics experiment, FRAM II, in the eastern Arctic. This was followed by an even more extensive program in March–May 1982, FRAM IV. Both of these experiments implemented an array of hydrophones and geophones with multichannel digital data recording. Work has been carried out on applying adaptive array processing to the measurement of the reverberation associated with the resulting acoustic signals, as well as the phase and group velocities of the seismic paths within the seabed and water column for refraction and bottom interaction studies. Work is also currently under way to examine the properties of several velocity function inversion techniques for multi-channel seismic ocean-bottom interaction data. The array data at FRAM II and FRAM IV are also being used to measure the scattering function of the channel at low frequencies and the directional spectra of the ambient noise in the Arctic. Associated with the acoustics experiment is a project aimed at extending the parabolic wave equation approximation for modeling underwater acoustics.

The summer of 1983 saw the successful beginning of a series of geophysical and acoustic experiments in the marginal ice zone of the Arctic. In this first MIZEX experiment, a large multichannel telemetered array was used to receive acoustic signals propagated across ocean and crustal paths marking the transition between the open waters of the Atlantic, and the ice covered regions of the Arctic Ocean. The goal of this experiment, and a more extensive one to be carried out in the same region in 1984, is to study the acoustic transmission properties of the laterally inhomogeneous and time varying water column, and to characterize the crustal velocity-depth function in this region. To carry out this work, much emphasis has been placed on the use of digital signal processing algorithms to decompose the multichannel data into domains that are easily related to the model parameters of interest. This is the beginning of the chain of analysis that leads to inversion of the observations for the model parameters.

Two additional projects related to geophysical signal processing in the context of ocean acoustics are the development of an algorithm for data processing to measure the acoustic reflection coefficient from the ocean bottom both for the deep water and shallow water cases. Out of this work

has come a Hankel transform algorithm as well as a new method for generating synthetic data.

In both the context of image processing and geophysical data processing we have obtained some significant results in the multi-dimensional high resolution spectral estimation problem. Specifically, we have developed new algorithms for maximum entropy power spectrum estimation which are computationally simple relative to previous approaches and applicable to both equally spaced and non-equally spaced data for both one-dimensional and two-dimensional signals. This algorithm has been applied to investigate the characteristics of multi-dimensional maximum entropy spectral estimates. In addition, we are investigating several approaches to improve the performance of the maximum likelihood method for spectral estimation.

Recently, we have proposed a new approach to the problem of estimating multiple signal and/or parameter unknowns using incomplete and noisy data. Our Minimum Cross-Entropy Method applies an information theoretic criterion to optimally estimate a separable probability density for the signal model. Not only does this new approach include all the various Maximum Likelihood and Maximum A Posteriori methods as degenerate cases, but it also directly leads to a simple iterative method of solution in which we alternate between estimating the various unknowns, one at a time. We are now exploring applications to statistical problems, iterative signal reconstruction, short-time analysis/synthesis, and noisy pole/zero estimation.

Another interesting area of research is the connection between signal processing algorithms and computer architectures. The "speed" of an algorithm depends not only on how many operations it requires, but also on how suitable it is for the computer architecture it runs on. With the advent of VLSI technology, it is now possible to build customized computer systems of astonishing complexity for very low cost. Exploiting this capability, however, requires designing algorithms which not only use few operations, but also have a high degree of regularity and parallelism, or can be easily pipelined. Directions we are exploring include systematic methods for designing multi-processor arrays for signal processing, isolating signal processing primitives for hardware implementation, and searching for algorithms for multi-dimensional processing which exhibit a high degree of parallelism.

There also are a number of projects directed at the development of new algorithms with broad potential applications. For some time we have had considerable interest in the broad question of signal reconstruction from partial information such as Fourier transform phase or magnitude. We have shown theoretically how under very mild conditions signals can be reconstructed from Fourier transform phase information alone. We have also developed a variety of theories and algorithms relating to signal reconstruction from Fourier transform magnitude and from partial short-time Fourier transform information. We are also exploring the application of some of these theoretical results to problems such as speech and image coding.

A recent and growing emphasis in our group is the combination of signal processing and artificial intelligence techniques. There are a variety of problems in signal analysis that can be approached

either from the analytical viewpoint characteristic of signal processing or the symbolic viewpoint characteristic of knowledge-based systems and artificial intelligence. We believe there is considerable potential for combining these two viewpoints into what we refer to as knowledge-based signal processing. There are currently two projects under way directed at developing this approach in the context of specific signal processing problems. One attempts to exploit artificial intelligence concepts to develop a knowledge-based pitch detector and the second, to explore knowledge-based signal processing in the context of signal enhancement. We have also coupled our work on knowledge-based signal processing into a project at Lincoln Laboratory on distributed sensor nets.

21.2 Improved Paraxial Methods for Modeling Underwater Acoustic Propagation

Arthur B. Baggeroer, Thomas E. Bordley

In modeling long-range sound propagation through the world's oceans, the severe computational and informational demands of the processing render straightforward solution of the hyperbolic wave equation governing the fields prohibitively expensive. For this reason, much effort has been expended to develop approximate solutions which effectively exploit the near stationarity of the channel with range. In mode methods, for example, the inhomogeneities in range are totally or partially ignored. The more general of these techniques, invoking the quasi-adiabatic approximation, assume that the primary effect of slow change in the channel is slow change in the shape of the modes, so that coupling between the modes is negligible. Other approaches such as ray methods simplify the problem by considering only the high frequency asymptotes of the field. In addition, a more general solution, the method of Gaussian beams, has been developed which treats the diffusive effects ignored by conventional ray tracing.

Of particular importance when the full field must be modeled and the inhomogeneities cannot be neglected are paraxial methods. These techniques solve a simplified form of the wave equation, obtained by splitting the field into transmitted and reflected components, and then neglecting the reflected field. The justification for this approach is that for a stationary media, no energy is reflected back to the source, so for a media which varies slowly with range, the reflected field should be small. In practice, current techniques perform well except when bottom interaction is important — for example, in areas such as the Arctic Ocean where no SOFAR channel exists. Even when these methods fail, however, their underlying assumption remains true: the energy transmitted forward greatly dominates the energy reflected back. Therefore, effective modeling within the framework of this approximation should be possible.

The concern of this research is to develop paraxial methods less restricted in angle, yet retaining the numerical stability and efficiency which have made this class of solutions so valuable. Two

methods are considered. The first, a multiple beam approach, finds a series of local solutions for the field, accurate in the neighborhood of the ray paths, and patches these solutions together to generate the total field. In the second, a dynamic splitting algorithm, the notion of a split is generalized to allow variation in range and depth, and the solution is determined iteratively with the current estimate of the field used to generate a better split for the next step. The effectiveness of these techniques is examined in the context of modeling sound propagation in the Arctic Marginal Ice Zone.

21.3 Adaptive Image Restoration

Jae S. Lim, Philip Chan

There have been many mathematically optimal linear noise smoothing techniques developed for noisy images with the assumption of stationarity over the entire image. The use of space invariant filters results in blurring of edges in the processed images. Space–variant filters offer an attractive alternative to space–invariant filter. This class of filters does not assume global stationarity nor require detailed *a priori* knowledge of the image and noise processes. These methods are adaptive to varying details over the image. In addition, these algorithms are simple to implement.

One effective adaptive noise smoothing algorithm estimates the *a priori* mean and variance of each pixel from its local mean and variance in a window. Then, the minimum mean–square error estimator in its simplest form is applied. Essentially, the algorithm multiplies the difference between the pixel value and the pixel mean by a gain factor and adds it to the local mean to obtain the output pixel. The gain factor k is a function of the local image and noise variances. For images with low SNR, this algorithm does not result in sufficient noise removal, especially near the high–contrast regions.

To achieve better control on the trade–off between the degree of noise smoothing and the degree of edge preserving, we introduce a contrast expansion factor β and modify the gain factor to k^β. For β = 0, the image remains unprocessed and for β = ∞, the filter reduces to simple local averaging. In addition, we have made a modification in the algorithm implementation, which leads to significant improvement both in performance and computations. Specifically, we use up to four one–dimensional filters oriented at 0, 45, 90, and 135 degrees respectively. Because each filter is adaptive in the same manner as the basic algorithm, a sharp edge inclined at a large angle to the filter direction remains intact, while the noise in the edge area is removed by one of the filters that lies closest to the direction of the edge. The same factor β is used to control the degree of smoothing as before. The result is the simultaneous reduction of noise and preservation of edges.

The algorithm is being developed and tested on images corrupted by additive Gaussian noise at a SNR as low as 3 dB. Preliminary results show remarkable improvement over other methods in its

ability to preserve edges and remove noise in all regions including the edge areas.

21.4 Signal Reconstruction from Partial Fourier Domain Information

Alan V. Oppenheim, Susan R. Curtis

In a variety of practical problems, only the magnitude or the phase of the Fourier transform of a signal is available, and it is desired either to reconstruct the signal exactly or to synthesize a signal which retains many of the important characteristics of the original signal. We are working on several distinct problems in this area. One set of problems is to identify those portions of the FT which contain most of the "intelligibility" information, and to develop conditions under which a signal can be exactly reconstructed from only this partial information. Another set of problems involves finding practical algorithms for doing the reconstruction and finding applications where such a reconstruction is desirable.

On the intelligibility problem, past work has shown that a signal synthesized with the correct phase or signed-magnitude maintains many of the important characteristics of the original signal, whereas a signal synthesized from the correct magnitude and zero or random phase does not. We have also found that a signal synthesized from one bit of phase alone is intelligible. Since this one bit of phase is contained in both the signed-magnitude and the phase but not in the magnitude alone, this result helps to explain the earlier results.

Past work on exact reconstruction involved the development of conditions under which signal reconstruction is possible from only the magnitude, phase, or signed-magnitude of the FT. We have developed a number of new extensions and generalizations of these results. Recently, we have also developed some surprising new results on signal reconstruction from one bit of phase alone, without any magnitude information.

We have also applied algorithms often used for reconstruction from magnitude or phase towards the problem of reconstruction from one bit of phase. An iterative reconstruction procedure, which involves imposing constraints in both time and frequency domains, only yields good results if the FT is significantly oversampled. Another procedure, involving the solution of linear equations, yields excellent results for small signals, but is impractical for large signals. We hope to develop an algorithm which works well for larger signals.

Despite the potential applicability of results on reconstruction from partial information to a wide

range of problems, so far the results have been mostly of theoretical value. Thus, we are also interested in finding applications of these results. One application we have been exploring is the design of FIR filters to match a given magnitude and phase specification. Our approach is to first find an FIR filter to match the phase specification using phase–only reconstruction techniques, and then use the Parks–McClellan algorithm to compensate for the resulting magnitude. We are also interested in exploring applications of our results on reconstruction from one bit of phase.

21.5 Helium Speech Enhancement from the Short–Time Fourier Transform Magnitude

Jae S. Lim, Douglas S. Deadrick

Since sound travels faster in a hyperbaric helium–oxygen atmosphere than in air at normal pressure, speech uttered in this type of environment suffers from certain severe degradations. This effect handicaps communication systems for deep–sea divers and others who must work in such an atmosphere.

The effects of the helium can be easily identified using short–time Fourier analysis. Specifically, the frequencies of the speech formants are increased non–linearly and the formant bandwidths are increased. These phenomena take place while the pitch information is left relatively undisturbed. Models exist for translating the formants of helium speech back to their normal frequencies, and these models are suitable for use with the short–time Fourier transform (STFT).

Previous attempts to enhance such speech using the STFT have introduced noise into the pitch magnitude and phase information. This work will apply pitch detection algorithms to the enhancement scheme in order to reduce the amount of noise. Moreover, since the required modifications to the STFT are valid only for its magnitude, recent techniques for estimating a signal from its STFT magnitude will be used to construct the enhanced speech.

21.6 Knowledge–Based Pitch Detection

Alan V. Oppenheim, Randall Davis, Webster P. Dove

Knowledge–based signal processing is an effort to design signal processing programs that go beyond purely numerical processing of the data and try to symbolically reason about the problem in

order to better solve it. Problems appropriate for this area are those whose model is either too complex to be solved directly with a numerical algorithm, and those for which the model is not well understood.

Pitch detection falls into this category both because the speech signal model is not well specified, and because the model for the generation of pitch is not fully understood.

This project is developing a program called the Pitch Detector's Assistant (PDA) which will both serve to reduce the effort involved in generating hand edited pitch and provide a laboratory for studying and programming the knowledge that makes humans better pitch trackers than existing automatic algorithms.

Existing methods of semi-automatic pitch detection[1] require the user to make a voicing decision and select a pitch individually for every frame. The PDA program is intended to analyze as much of the utterance as it is sure of and then help the user with the remaining difficult portions. Thus we expect a dramatic reduction in the time from the current 30 minutes per second of speech analyzed.

Although there have been projects which combine signal processing and AI technology for particular problems such as speech understanding[2,3] and underwater acoustic signal recognition,[4] the actual signal processing present in these systems has only been used as a means for generating symbolic objects. These objects are then manipulated by the AI portions of the program until an interpretation of the data is complete. The symbols do not provide information to assist subsequent numerical processing, and thus the information flows one way from numeric to symbolic form. The pitch detection problem choice is motivated by the observation that these other problems are ones of recognition (i.e. signals in, symbols out) and naturally lead to solutions which process numerically first and symbolically later.

By choosing a problem which involves signal output we assure the use of numerical processing in later portions of the program. The creation and study of programs which emphasize the interaction between symbolic and numerical processing, is the primary purpose of the knowledge-based signal processing effort at MIT.

References

1. C.A. McGonegal, L.R. Rabiner, and A.E. Rosenberg, "A Semiautomatic Pitch Detector (SAPD)," IEEE Trans. ASSP 23. 570-574 (December 1975).
2. L. Erman, R. Hayes-Roth, V.R. Lesser, and D.R. Reddy, "The Hearsay-II Speech Understanding System: Integrating Knowledge to Resolve Uncertainty," Computing Surveys 12, 213-254 (June 1980).
3. B. Lowerre and R. Reddy, "The HARPY Speech Understanding System," in W. Lea (Ed.), Trends in Speech Recognition (Prentice-Hall, 1980) pp. 340-360.
4. H.P. Nii, E.A. Feigenbaum, J.J. Anton, and A.J. Rockmore, "Signal-to-Symbol Transformation: HASP/SIAP Case Study," AI Magazine 3, 23-35 (Spring 1982).

21.7 Multi–Dimensional High–Resolution Spectral Analysis and Improved Maximum Likelihood Method

Jae S. Lim, Farid U. Dowla

Although MEM, MLM, and AR–modeling spectral estimation are high resolution spectral estimation algorithms, in multidimensional situations there are problems with each of these methods. The MEM algorithms are iterative and their applications to real world problems is prohibitive from numerical viewpoints. MLM does not quite achieve the resolution which one would like it to have. The AR–modeling has problems as the shape of the spectrum is distorted when the filter mask does not conform to certain symmetry.

We are developing a closed–form high resolution spectral estimation algorithm based on the concepts of MLM, MEM, and AR spectral estimation. We have found a useful relationship between the MLM and AR signal modeling in multidimensions. By exploiting this relationship and by studying the problem of array design in multidimensions for these algorithms, we propose to present an algorithm with computational properties like the MLM but whose resolution property is better than the MLM.

The performance of the algorithm is being evaluated on synthetic data and on real data in a multichannel radar tracking problem.

21.8 Speech Synthesis from Short–Time Fourier Transform Magnitude Derived from Speech Model Parameters

Jae S. Lim, Daniel W. Griffin

Previous approaches to the problem of synthesis from estimated speech model parameters have primarily employed time–domain techniques. Most of these methods generate an excitation signal from the estimated pitch track. Then, this excitation signal is filtered with a time–varying filter obtained from estimates of the vocal tract response. This approach tends to produce poor quality "synthetic" sounding speech.

In this research, a new approach to speech synthesis from the same estimated speech model parameters is investigated. In this approach, a desired modified short–time Fourier transform magnitude (MSTFTM) is derived from the estimated speech model parameters. The speech model

parameters used in this approach are the pitch estimate, voiced/unvoiced decision, and the spectral envelope estimate. The desired MSTFTM is the product of a MSTFTM derived from the estimated excitation parameters and a MSTFTM derived from spectral envelope estimates. (The word "modified" is included in MSTFTM in these cases to emphasize that no signal exists. in general, with this MSTFTM). Then, a signal with short–time Fourier transform magnitude (STFTM) close to this MSTFTM is estimated using the recently developed LSEE–MSTFTM algorithm.[1,2] Preliminary results indicate that this method is capable of synthesizing very high quality speech, very close to the original speech.

This method has applications in a number of areas including speech coding and speech enhancement. In speech coding applications. the excitation parameters and spectral envelope can be coded separately to reduce transmission bandwidth. Then, these coded parameters are transmitted and then decoded and recombined at the receiver. In speech enhancement applications, the excitation parameters and the spectral envelope can be separated, processed separately, and then recombined.

References

1. D. W. Griffin and J. S. Lim, "Signal Estimation from Modified Short–Time Fourier Transform," Proceedings of the 1983 Int. Conf. on Acoustics, Speech and Signal Processing, Boston, Massachusetts, April 14–16, 1983, pp. 804–807.
2. D. W. Griffin and J. S. Lim, "Signal Estimation from Modified Short–Time Fourier Transform," IEEE Trans. ASSP 32, 2, 236–243 (1984).

21.9 Speech Enhancement Using Adaptive Noise Cancelling Algorithms

Jae S. Lim, William A. Harrison

This research is directed towards evaluating the performance of some general adaptive noise cancelling algorithms in a distributed noise environment. General adaptive noise cancelling algorithms utilize one or more reference microphones that record a correlated version of noise that is additively corrupting a desired signal in the primary microphone. In practice, the desired signal is often recorded by the reference microphones. Under these conditions. the algorithm will attempt to cancel part or all of the desired signal. Under certain conditions, modifications in an adaptive noise cancelling algorithm such as the Widrow–Hoff least mean square algorithm. allows one to still reduce the background noise without severely distorting the desired signal. One application of this work is in reducing the background noise in a jet fighter pilot's speech. Test cases conducted with simulated data have shown some promise that the ANC algorithm can be used to improve the SNR of the pilot's speech.

21.10 Overspecified Normal Equations for Autoregressive Spectral Estimation

Jae S. Lim, David Izraelevitz

There is a one-to-one relationship between a set of P normalized positive definite correlation estimates and the P predictor coefficients derived using autoregressive modeling. Several researchers have proposed the use of M > P correlation estimates to provide a better Pth order model. Specifically, the normal equations are augmented to provide M linear equations between the correlation estimates and the predictor coefficients. Since the system of equations is now overspecified, a least squares solution is required.

In this research a study is presented of some of the properties of the method of overspecified normal equations as applied to the problem of spectral estimation. The main contribution of this research is the derivation of the relationships between the number of correlations used, the model order and the signal-to-noise ratio of the signal, to the characteristics of the resulting spectral estimate. The characteristics studied are the spectral height, bandwidth and area. The method is shown to be a spectral density estimator like the ME method, where spectral areas rather than spectral values should be interpreted as estimates of power.

The relationships derived point to the number of correlations used over the minimum, i.e., model order, as a signal-to-noise enhancer. The resulting spectrum is equivalent to the ME spectrum under higher signal-to-noise conditions. Another result is the requirement of a proportionality constant dependent on the number of correlations and the model order which is necessary for unbiased signal-to-noise measurements. This constant is not required however, for measurements of relative power within the same spectral estimate, as in the power ratio of two sinusoids in noise.

The second part of the research presents some empirical studies using computer simulations which verify the theoretical predictions and provide the region of validity of the analysis. Further experiments study the interfering effect of several closely spaced sinusoids. The method of overspecified normal equations is shown to be much more sensitive to this interference than the ME method. Finally, some further empirical studies are made of the resolution capabilities of the method. Using the data derived, an empirical model is derived which seems to agree to some extent with the data.

This work was completed in June 1983.

21.11 Restoration of Image Sequences with Motion

Jae S. Lim, Dennis M. Martinez

The problem of image restoration for static, single frame images has received much attention in signal processing research. Most of the effort has been devoted to the problem of restoring single image frames which have been degraded in some way (additive noise, quantization noise, etc.). However, in many situations one is dealing with sequences of image frames, which when viewed in temporal succession constitute a scene with motion. Either in communicating or processing of such signals, various degradations can also occur.

In developing algorithms for restoring sequences of image frames, there are a number of issues which are not present in the single frame case. For example, there is the possibility of utilizing inter-frame as well as intra-frame information in performing restoration. Inter-frame methods take advantage of the fact aht most motion video frames have a high degree of temporal correlation. However, because the scenes are changing with time, there is the added problem of motion compensation to be contended with.

This research is concerned with developing methods for using motion information in performing restoration of scenes with motion. There are two important questions presently under consideration. Even before attempting to use motion information in restoration algorithms, a fundamental question concerns how accurate motion estimates have to be in order to be useful in performing restoration. One goal of the presen work is to determine how the performance of several different restoration algorithms vary as the accuracy of a motion estimate deteriorates. Once this has been determined, attention will be focused on developing restoration algorithms which can combine motion estimation with restoration.

21.12 Knowledge–Based Array Processing

Alan V. Oppenheim, Evangelos E. Milios

The objective of this research is to explore the integration of signal processing knowledge with symbolic knowledge about the physical objects that generated the signal in the context of acoustic sensor array processing. A taxonomy of array processing techniques is being formed on the basis of their assumptions and performance in the real world of acoustic sensor array processing. This knowledge will be matched with information about the motion of the sources of sound and the system

will possibly be coupled with another system that performs low–level interpretation of signal processing results. For now, the interpretation system will be simulated by a human interpreter.

In order to explore the issues involved in planning, diagnosis, monitoring, design and prediction within the array processing context, a simpler problem is being considered, that of performing spectral analysis of a single acoustic waveform with the purpose of classifying the source that generated it. The signal processing issues in one–dimensional spectral analysis are much better understood than in the multidimensional case to which the array processing problem can be reduced. For this reason, it is felt that the one–dimensional problem is an appropriate first step toward exploring the signal–symbol integration. It is believed that research in the one–dimensional case will generate the necessary ideas to enable the design of the more general knowledge–based array processing system.

21.13 The Use of Speech Knowledge in Speech Analysis

Alan V. Oppenheim, Cory Myers

The problem of speech modeling is one which has been of great interest in the signal processing community. As part of our efforts to explore methods by which signal processing and symbolic processing techniques can be made to interact we are building a system for symbolic reasoning about speech analysis. The system produces a time–varying parametric model of the speech signal where, not only do the parameters change over time, but the models used may also change over time. For example, the system may perform all–pole modeling of the signal but may change the number of poles which are to be used in different portions of the signal. The system attempts to develop a parametric model of a speech signal given both the speech signal (or a noisy version of it) and a symbolic description of how the signal was produced. The symbolic description of the signal includes information about the speaker, such as gender and age, and information about the recording environment, such as noise characterization, sampling rate and speech bandwidth. The system is also given a symbolic description of the content of the signal in the form of a time–aligned phonetic transcription. The output from the system is a time–varying parametric representation of the speech signal (excluding excitation information).

The system generates its parametric representation in three stages. First a symbolic reasoning system takes the description of the signal, breaks the signal into sections, and recommends a signal analysis method for each section of the signal. These recommendations include choice of model and choice of analysis parameters, e.g. model order, window size, analysis rate, etc. The second stage of

the system runs the signal analysis methods recommended by the first stage. Estimates of modeling accuracy are also generated during the signal analysis. The third stage of the system examines the results from the signal analysis and the accuracy measures. It uses these and the original symbolic information to determine new models to try or modifications to make to the values from the models. Modifications are suggested according to both modeling error criteria and model smoothness criteria. The choice of appropriate modification criteria varies with the symbolic information available. The third stage continues to make modifications until a satisfactory set of models has been found.

21.14 Estimation of the Degree of Coronary Stenosis Using Digital Image Processing Techniques

Jae S. Lim, Thrasyvoulos N. Pappas

The aim of this research is the development of an algorithm for evaluating the degree of coronary artery stenosis from coronary cine–angiograms. A cine–angiogram is a sequence of x–ray pictures of the coronary arteries in which a contrast agent has been injected via a catheter. The precise measurement of the stenosis of the coronary arteries is important in the treatment of patients with ischemic heart disease.

The first step will be the determination of the percentage diameter reduction from a single frame of the cine–angiogram. This will require the detection of the boundaries of the coronary arteries and the analysis of the variation of their diameter. Preprocessing of the image for noise reduction will hopefully lead to improved boundary detection.

Subsequent steps will involve the analysis of multiple frames of the cine–films, and the investigation of densitometric procedures, which use the brightness information within the artery to obtain estimates of the cross–sectional percentage area reduction.

21.15 Automatic Target Detection in Aerial Reconnaissance Photographs

Jae S. Lim, Michael D. Richard

The detecting of small anomalous regions in images has aroused much interest in such areas as optical aerial reconnaissance, radar analysis, terrain classification, and medical diagnosis through imagery. A recently developed algorithm[1] has proven highly successful in detecting small objects or

targets in natural terrain such as trees, grass, and fields of aerial photographs. The algorithm uses a significance test to distinguish each image pixel as either background or non–background (i.e., target). Specifically, the background is assumed to be characterized by a nonstationary Gaussian probability distribution. The algorithm further represents the background by a two dimensional (2–D) autoregressive model. The resulting significance test is expressed as the error residuals of 2–D linear prediction.

This research will explore several new areas to either develop a superior detection algorithm or to significantly improve the existing one. First, the issue of target modelling will be explored. The current algorithm models only the background and treats targets simply as anomalies in the background. The question arises as to how a suitable model for targets can be incorporated in a detection algorithm. Second, methods to detect and to fully determine the boundaries of larger objects will be considered. The present algorithm can detect only small point objects representing statistical irregularities in the background random process. The issue of detecting larger targets poses significant questions regarding object detection, image segmentation, and boundary extraction. Additional research in these two areas should improve the somewhat favorable results obtained by using linear predictive techniques to detect anomalous regions in images.

References

1. T.F. Quatieri, "Object Detection by Two–Dimensional Linear Prediction," M.I.T. Lincoln Laboratory Technical Report 632, 28 January 1983.

21.16 Separation of Desired Speech from Interference Speech Reverberating in a Room

Bruce R. Musicus, Hiroshi Sekiguchi

An important problem in the area of speech processing is speech enhancement. The objective of speech enhancement is ultimately to improve the intelligibility of the desired speech or to separate the desired speech from interfering signals, such as noise, other speech, or echoes.

A specific problem in speech enhancement, which this research will deal with, is the separation of an acoustically added speech signal simultaneously uttered in a confined room from the desired signal. This type of problem is often observed in the real world. One example is a broadcast news room environment. When a news anchor person is on the air in a broadcast studio, studio engineers running the broadcast often transmit to the news anchor person specific cues or messages to inform him/her of what is to be done next. If these messages are sent out over an audio monitor placed near the news anchor person, the microphone into which the newscaster speaks would also pick up the

messages played out by the audio monitor. It would be then necessary to subtract out the monitor signal together with any acoustic reflections from the microphone signal. This research will seek a practical and effective algorithm for canceling interfering speech in a real news room environment.

One class of enhancement methods attempts to subtract the interfering speech signal by exploiting the harmonic nature of speech. This class is called the one–data channel method, since only one microphone is used. One of these methods utilizes the difference in the fundamental frequencies, or pitch periods, between the interfering speech signal and the desired speech signal to selectively eliminate the interfering speech from the microphone signal using a comb filter. The two–data channel method primarily employs adaptive noise canceling schemes. Related work has been done by Mark Paulik, in which he utilized the noise cancellation scheme of Benard Widrow because both the microphone signal and the monitor signal are available to the algorithm. Although his work was restricted to exploring the theoretical feasibility of the adaptive noise cancellation approach, based on modeling the reverberant impulse response with a 21 tap filter with known coefficients, his results suggest that the approach may be quite practical.

The aim of this research is to investigate various types of modeling and estimation techniques for a reverberant environment. These include the periodogram technique, the pole–zero modeling technique, and the FIR technique. Facts which complicate the problem include: 1) The room reverberant impulse response may last more than 200 msec. This would produce more than 2000 parameters which would need to be estimated in the FIR technique at a 10 KHz sampling rate. It also indicates that at least 5 Hz frequency resolution may be required in the periodogram technique. 2) The news anchor person may move around in the studio with the microphone. Therefore, the estimation algorithm must adjust itself to the change of the system function adaptively. This requires a real–time algorithm fast enough to keep track of the change. The results of this algorithm will be combined with the noise cancellation scheme previously proposed by Mark Paulik, and the applicability of this new approach to a real studio environment will be studied. Hopefully, it will present an algorithm effective enough to display good performance in canceling the interfering signal in an actual studio.

21.17 Low Bit Rate Video Conferencing

Jae S. Lim, Ramakrishnan Sundaram

The attempt of this research is to achieve a drastic reduction in the channel capacity requirements for video transmissions. Currently, such transmissions require bandwidth of the order of 100 Mbps. An algorithm based on local averaging and adaptive thresholding has been developed which reduces

an 8–bit picture to a 1–bit binary picture without significant loss in quality. The resulting image is then smoothed by adaptive median filtering to remove random bit transitions. This is then suitable for run length coding. Compression factors of up to 60 are expected to result. Other forms of data encryption are also being examined. Currently the emphasis is only on intraframe redundancy reduction. The work is expected to lead on to interframe analysis.

21.18 Improved Techniques for Migrating Acoustic Fields

Alan V. Oppenheim, George V. Frisk[10] Michael S. Wengrovitz

The problem of inverting an acoustic field from a continuous–wave point source in the ocean to determine information about the bottom is being studied. Although there has been recent progress in this area, a successful inversion using real experimental data has not yet been performed. One apparent source of degradation in the inversion is the variation of sensor height off the ocean floor as a function of range. Numerical experiments with synthetic data have shown the importance of compensating for this variation.

Several algorithms which migrate the field to a fixed height off the bottom have improved the inversion results. These algorithms are based on the principle of a single dominant ray arriving at each range point. This concept suggests that improved migration schemes might account for multiple ray arrivals.

This research will explore new migration algorithms which exploit simple, robust ray approximations to acoustic data. These techniques may also be applicable to direct profile inversion using offset/ray–parameter data, mode determination in an acoustic waveguide, and reflected/direct field separation in acoustic experiments.

---

[10]Woods Hole Oceanographic Institution