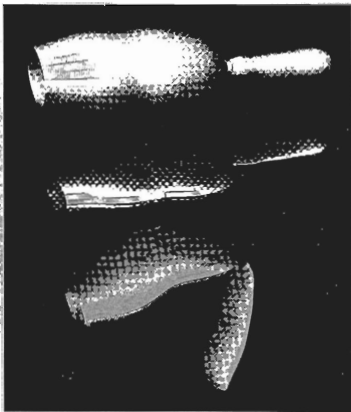


Acoustics Australia



SPECIAL ISSUE ON SPEECH SCIENCE AND TECHNOLOGY

- **The Australian "Okay"**
- **Wideband speech coding**
- **3D vocal-tract models**
- **Prospects for speech technology**
- **Forensic speaker identification**
- **Speaker recognition evaluation**
- **Hearing aid dynamic optimisation**
- **Cochlear implants and hearing aids**



Step into the frequency domain

SignalCalc[®] Mobilyzer

Up to 16 channels in a compact network peripheral with high speed DSP and DC power



FFT
SRS
MIMO
Modal
Waterfall
Spectrogram
Order Tracking

Time Averaging
Sine Reduction
Realtime Octave
Correlation and Histogram
Frequency Response Function
Disk Throughput and Playback

SignalCalc Mobilyzer is a modern architecture Dynamic Signal Analyzer that distributes the processing load over multiple DSP's and CPU's to achieve high speed, high quality signal analysis. The Mobilyzer chassis is a network peripheral that interfaces to a notebook or desktop computer running Windows 95, 98, ME 2000 or NT via a standard Ethernet network cable.



DATA PHYSICS CORPORATION SOLUTIONS IN
SIGNAL PROCESSING



Supplied by
KINGDOM PTY LTD

PHONE 02 9975 3272



Acoustics Australia

EDITORIAL COMMITTEE:

Neville Fletcher
Marion Burgess
Joseph Lai

BUSINESS MANAGER:

Mrs Leigh Wallbank

Acoustics Australia General Business

(subscriptions, extra copies, back issues, advertising, etc.)

Mrs Leigh Wallbank
P O Box 579
CRONULLA NSW 2230
Tel (02) 9528 4362
Fax (02) 9523 9637
wallbank@zipworld.com.au

Acoustics Australia All Editorial Matters

(articles, reports, news, book reviews, new products, etc)

The Editor, Acoustics Australia
Acoustics & Vibration Unit
Australian Defence Force Academy
CANBERRA ACT 2600
Tel (02) 6268 8241
Fax (02) 6268 8276
email: acoust-aust@adfa.edu.au
www.acoustics.asn.au

Australian Acoustical Society Enquiries see page 52

Acoustics Australia is published by the Australian Acoustical Society (A.B.N. 28 000 712 658)

Responsibility for the contents of articles and advertisements rests upon the contributors and not the Australian Acoustical Society. Articles are copyright, by the Australian Acoustical Society. All articles are sent to referees for peer review before acceptance. Acoustics Australia is abstracted and indexed in Engineering Index, Physics Abstracts, Acoustics Abstracts & Noise: Abstracts and Reviews.

Printed by
Cronulla Printing Co Pty Ltd,
16 Cronulla Plaza,
CRONULLA 2230
Tel (02) 9523 5954,
Fax (02) 9523 9637
email: print@cronullaprint.com.au
ISSN 0814-6039

Vol 29 No 1

CONTENTS

April 2001

• A Methodology for Modelling and Interactively Visualising the Human Vocal-tract in 3D Space M. Barlow, F. Clermont & P. Mokhtari	Page 5
• Sound Separation with a Cochlear Implant and a Hearing Aid in Opposite Ears P. Blamey, C. James & L. Martin	Page 9
• Coding Wideband Speech at Narrowband Bit Rates J. Epps & W. Holmes	Page 13
• Rapid Channel Compensation for Speaker Verification in the NIST 2000 Speaker Recognition Evaluation J. Pelecanos & S. Sridharan	Page 17
• Adaptive Dynamic Range Optimisation of Hearing Aids L. Martin, P. Blamey, C. James, K. Galvin & D. Mactariane	Page 21
• Prospects for Speech Technology in the Oceania Region J. B. Miller	Page 25
• A Comparison of Two Acoustic Methods for Forensic Speaker Discrimination P. Rose & F. Clermont	Page 31
• Auditory and F-Pattern Variations in Australian Okay: A Forensic Investigation J. Elliott	Page 37
Letters	42
New Members	42
Future Meeting	43
Meeting Reports	44
Standards	45
FASTS	45
News	45
New Products	46
People	47
Book Review	48
Diary	51
Acoustics Australia Information	52
Australian Acoustical Society Information	52
Advertiser Index	52

Cover illustration: Human vocal tract model: See paper by Barlow, Clermont & Mokhtari.

AUSTRALIAN ACOUSTICAL SOCIETY - SUSTAINING MEMBERS

ACOUSTIC RESEARCH LABORATORIES

LEVEL 7 BUILDING 2
423 PENNANT HILLS ROAD
PENNANT HILLS 2120

ACRAN

P O BOX 34
RICHLANDS 4077

ADAMSSON ENGINEERING PTY LTD

P O BOX 1294
OSBORNE PARK 6916

ARMSTRONG WORLD INDUSTRIES

P O BOX 109
MORDIALLOC 3195

ASSOCIATION OF AUSTRALIAN ACOUSTICAL CONSULTANTS

96 PETRIE TERRACE
BRISBANE 4000

BARCLAY ENGINEERING

31 FELSPAR STREET
WELSHPOOL 6106

BORAL PLASTERBOARD

676 LORIMER STREET
PORT MELBOURNE 3207

BRUEL & KJAER AUSTRALIA

24 TEPKO ROAD
TERREY HILLS 2084

CHADWICK TECHNOLOGY

9 COOK STREET
FORESTVILLE 2087

CSR BRADFORD INSULATION

55 STENNETT ROAD
INGLEBURN 2565

ENCO NOISE CONTROL PTY LTD

50 RIVERSIDE ROAD
CHIPPING NORTON 2170

G P EMBELTON & CO PTY LTD

P O BOX 207
COBURG 3058

INC CORPORATION PTY LTD

22 CLEELAND ROAD
OAKLEIGH SOUTH 3167

KELL & RIGBY

8 DUNLOP STREET
STRATHFIELD 2136

NAL CONSULTING

126 GREVILLE STREET
CHATSWOOD 2067

NOISE CONTROL AUSTRALIA PTY LTD

70 TENNYSON ROAD
MORTLAKE 2137

NSW ENVIRONMENT PROTECTION AUTHORITY

P O BOX A290
SYDNEY SOUTH 1232

NUTEK AUSTRALIA

UNIT 3, 10 SALISBURY ROAD
CASTLE HILL 2154

PEACE ENGINEERING PTY LTD

P O BOX 4160
MILPERRA 1891

PYROTEK SOUNDGUARD

149 MAGOWAR ROAD
GIRRAWEEEN 2145

SOUND CONTROL PTY LTD

61 LINKS AVENUE NTH
EAGLE FARM 4009

VIPAC ENGINEERS AND SCIENTISTS LTD

279 NORMANBY ROAD
PORT MELBOURNE 3207

WARSASH SCIENTIFIC PTY LTD

UNIT 7, 1-9 MARIAN STREET
REDFERN 2016

WORKCOVER

COMPLIANCE COORDINATION TEAM
LEVEL 3, 400 KENT STREET
SYDNEY 2000

COUNCIL OF THE AUSTRALIAN ACOUSTICAL SOCIETY 2000-2001

President:	Geoff Barnes	General Secretary:	David Watkins
Vice-President:	Charles Don	Federal Registrar:	Gillian Adams
Treasurer:	Ken Miki	Archivist:	Louis Fouvy

Councillors

(Effective from November 2000)

New South Wales

David Eager

Ken Miki

Queensland

Gillian Adams

Ian Hillcock

South Australia

Byron Martin

Pete Teague

Victoria

Geoff Barnes

Charles Don

Western Australia

Daniel Lloyd

Terry McMinn

AUSTRALIAN ACOUSTICAL SOCIETY ON THE WEB:

<http://www.acoustics.asn.au>

Guest Editorial

This special issue is dedicated to speech science and speech technology research within Australia, and in particular is inspired by the recent SST-2000 (8th Australian International Conference on Speech Science and Technology) conference held in Canberra in December 2000. The SST conference series is biennial and hosted by ASSTA (the Australian Speech Science and Technology Association Inc.), attracting national and international researchers within the speech field.

Since the mid-1980s ASSTA has served as a professional association representing the interests of members working in all aspects of the multidisciplinary speech field. That includes fundamental speech science such as language acquisition, phonetics (acoustic, linguistic, and forensic), speech perception and physiology; together with the more applied areas of speech

technology such as speech recognition and understanding, speech coding, speech synthesis, speaker recognition, signal processing, and multi-modal speech.

This issue contains a total of eight papers which, while it cannot represent the diversity and depth of speech research within Australia, at least goes some way towards conveying its multifaceted quality. Two papers (Elliott, and Rose & Clermont) concern forensic phonetics, the questing of reliable and legal identification of speakers from their voice. This is a long-standing research area that lies at the interface between science and society. A further two papers (Martin et. al., and Blamey et. al.) concern cochlear implants and hearing aids; areas in which Australia is an acknowledged world leader. Barlow et. al.'s paper deals with plausible 3D modelling of the human vocal-tract, a challenging area with potential for application in many

fields. Millar's paper highlights the prospects for, and challenges faced, in applying speech technology within the Oceania region with its varied linguistic, social and economic groups. Epps & Holmes paper addresses speech coding; the compression of speech for transmission or storage, and the issue of achieving high perceptual quality while keeping bit-rates low. Pelecanos & Sridharan's paper deals with the difficult task of speaker verification (the acceptance or rejection of an identity claim using speech as the sole criterion) over telephone lines when more than person may be speaking. Taken together these eight papers form a snapshot glimpse of just some of the speech research activity within Australia today, and hint at the richness of the field. I can only hope that it has whetted your interest.

Michael Barlow

From the President

The 'Historical Issue' of Acoustics Australia (vol 28, no 3, 2000) was surely an impressive overview of the history of the Australian Acoustical Society from a variety of perspectives. The issue also chronicled the past works of notable acoustical researchers in specific acoustic fields such as architectural and musical acoustics, community noise, audiology and others. I expect this edition will become in itself a reference document in any future historical treatise on the subject of Acoustics in Australia.

On behalf of all members of the Australian Acoustical Society I commend the Editorial Committee for your massive undertaking in the presentation of such an excellent publication. I also commend all

who contributed to the edition, whether representatives of the various divisions or authors of the papers and articles.

It was also encouraging to see new advertisers along with new acoustic products being promoted in Acoustics Australia by our Sustaining Members. There also seems to be a significant increase in 'flier insert' advertising material including details of seminars, conferences and opportunities for positions in the acoustic field. With this in mind, there is at least one Australian University who is encouraging their physics graduates to look towards a career in what they consider is the growing field of acoustics...so watch out you Mechanical Engineering graduates, there may be some competition!

My experience over the past several years, along with that of other acoustical consultants throughout Australia who seem to be extremely busy or expanding their practices, or both busy and expanding, appears to confirm the University's prediction. This augurs well for the acoustic fraternity of the future.

Amongst the flier material was the notice for Acoustics 2001 "Noise and Vibration Policy - The Way Forward" to be held in Canberra, 21st to 23rd November. Diary the date to attend and participate in what promises to be a vital conference. I look forward to meeting you there.

Geoff Barnes



ACOUSTICS 2001

Noise and Vibration Policy — The Way Forward?

21 to 23 November 2001 Canberra, ACT

See Page 49 of this issue or www.acoustics.asn.au

YOUR ACOUSTIC DESIGNS WORK WITH ALLIANCE

Our job starts where your job ends

At Alliance, we are dedicated to making your acoustical designs work for your clients.

We provide the products and services to meet all of your design requirements.

Experience counts

18 years in the business of building, retrofit and five years as the leading contractor to the Sydney Aircraft Noise Insulation Project, offers you the experience and expertise to take your acoustic solutions to successful completion.

Among our clients are:

- Sydney Aircraft Noise Insulation Project
- NSW Roads and Traffic Authority
- Warringah Shire Council
- Vodafone

ALLIANCE

CONSTRUCTIONS

72 Parramatta Road, Summer Hill 2130 PO Box 368 Summer Hill NSW 2130

Phone 02 9716 9799 Fax 02 9716 9800

Email info@nationalalliance.net Web: <http://www.nationalalliance.net>

NEW PRODUCTS - Lowest cost meters on the market 1/1 & 1/3 Oct loggers, Dual purpose sound level meters



SVAN 945

Main Features

SVAN 945

- Type 1 meter
- One measurement range
24dB ~ 139dB
- 1/1 & 1/3 octave real time
analysis
- Acoustic loudness measurement
- 3 Megabyte of memory
- RS232 interface
- Built-in rechargeable battery
- Light weight 600 grams
- Advanced PC download software



SVAN 943

Main Features

SVAN 943

- Type 2 meter
- One measurement range
26dB ~ 133dB
- 1/1 & 1/3 octave real time
analysis
- Acoustic dose meter function
- 3 Megabyte of memory
- RS232 interface
- Built-in rechargeable battery
- Light weight 500 grams
- Advanced PC download software

SALES, CALIBRATION, HIRE & REPAIRS

NUTEK AUSTRALIA

Ph: (02) 9894-2377

Fax: (02) 9894-2386

Email: contact@nutek-australia.com.au Website: www.nutec-australia.com.au



Reg. Lab. No. 9262
Acoustic & Vibration
Measurements

A METHODOLOGY FOR MODELLING AND INTERACTIVELY VISUALISING THE HUMAN VOCAL-TRACT IN 3D SPACE

Michael Barlow¹, Frantz Clermont¹ and Parham Mokhtari²

¹School of Computer Science, University College, University of NSW

²Electrotechnical Laboratory, Tsukuba, Japan

ABSTRACT. A system is described for constructing and visualising three-dimensional (3D) images of the human vocal-tract (VT), either from directly-measured articulatory data or from acoustic measurements of the speech waveform. The system comprises the following three major components: (1) a method of inversion for mapping acoustic parameters of speech into VT area-functions, (2) a suite of algorithms which transform the VT area-function into a 3D model of the VT airway, and (3) solutions for immersing the 3D model in an interactive visual environment. The emphasis in all stages of modelling is to achieve a balance between computational simplicity as imposed by the constraint of real-time operation, and visual plausibility of the reconstructed 3D images of the human vocal-tract.

1. INTRODUCTION

Vocal-tract (VT) modelling has long received considerable research efforts, often because it is seen as a step-above the acoustic signal in the communication chain; being closer to the original communicative intent and hence being a more compact as well as a richer (or at least less convolved) source of knowledge. Application areas of VT modelling include speech synthesis, speech coding, speech recognition and speaker characterisation. Whilst previous approaches to VT-modelling have ranged from acoustically- and/or physiologically-motivated mathematical models (e.g., Mermelstein and Schroeder, 1965; Lindblom and Sundberg, 1971; Mermelstein, 1973; Coker, 1976) to parameterisation of direct measurements made by magnetic resonance imaging (MRI) or ultrasound imaging techniques (e.g., Harshman et al., 1977; Yehia et al., 1996; Story and Titze, 1998), the so-called "inverse problem" of estimating the VT geometry from the acoustic speech signal has long been regarded as a potentially revolutionary approach with far-reaching applications. However, it appears that theoretical and practical problems such as non-uniqueness and articulatory compensation have limited the use of estimated VT-shape (or area-function) data mainly to theoretical investigations of the inverse problem itself.

Our paper describes an application-driven approach to the inverse problem, with the long-term goal of plausibly realistic 3D vocal-tract models, constructed in real-time (i.e., as the user phonates). Key to our methodology is the real-time constraint with its implications for the complexity of the model that can be supported, coupled with a relaxation of the exactness/uniqueness criteria. Our system has applications in areas such as foreign-language acquisition or rehabilitation of speech pathologies, where it would be a distinct advantage for users to receive real-time visual feedback on the difference between their own VT configuration and the ideal or standard one being attempted.

The system proposed herein consists of three major components: (1) a computationally tractable VT model and method of acoustic-to-articulatory mapping (or inversion) which is used to estimate VT area-functions from the acoustics of speech; (2) a suite of 3D-modelling software used to transform an estimated area-function into a 3D polygon mesh (surface of a straight tube with varying cross-sectional areas), and then to apply such spatial transformations as to make the model conform more closely to human vocal-tract anatomy, thereby yielding a more *plausible* 3D reconstruction of the VT-shape; (3) an environment in which to present these models to the user so that they may be interactively explored. These three major components are elaborated in the remainder of this paper, and followed by a concluding discussion.

2. ACOUSTICS TO AREA-FUNCTION

The first major component of our system comprises a computationally tractable model of the VT and a method of mapping from acoustic to model parameters. As an initial step, the method selected involves a VT-shape parameterisation first introduced by Mermelstein and Schroeder (1965). In particular, the logarithmic area-function is parameterised in terms of the first few odd-indexed terms of the cosine-series, as follows:

$$\ln A(x) = \ln A_0 + \sum_{n=1}^M \alpha_{2n-1} \cos((2n-1)\pi x/L), \quad (1)$$

where x is the distance along the VT airway from the glottis to the lips, L is the total length of that distance, is an area scaling factor whose value is computed to retain an overall mean logarithmic-area equal to zero, and M is the number of terms retained in the series. The elegant simplicity of this model lies in the following, quasi-linear relation which maps the n^{th} formant frequency (n^{th} resonant frequency of the vocal tract) to the corresponding model parameter α_{2n-1} :

$$a_{2n-1} = -2 \frac{(F_n - F_{n0})}{F_{n0}}, \quad (2)$$



Figure 1: An example of the first stage in the 3D modelling process. Sets of vertices are constructed from a series of points which define circular VT segments. These are then connected in triangle strips. The model shown here was constructed from an MRI-measured VT area-function (Yang & Kasuya, 1994) of a young Japanese male phonating /a/, and consists of 340 vertices composing 640 triangles.

where $F_n = (2n-1)c/4L$ is the n^{th} formant frequency of a uniform area-function of the same length L , and $c=35300$ cm/sec is the speed of sound in the VT airway.

The approximate equality in Equation (2) becomes increasingly inexact as the formants are more distant from their neutral values. The iterative method proposed by Mermelstein (1967) is therefore used to ensure that resynthesis of the formants using a completely lossless VT acoustic model will exactly reproduce the target formant frequencies. Furthermore, the VT-length L is optimised such that the inversion procedure yields a VT-shape with minimal eccentricity from that of a uniform tube. In particular a number of potentially realistic VT-lengths are considered and the corresponding VT-shapes derived. These shapes are contrasted, as an RMS difference, with the uniform neutral tube, with the VT-shape/VT-length that yielded the minimum difference then being selected.

While the above method is preferred owing to its computational simplicity and the fact that the model parameters are directly and uniquely related to acoustic (formant) parameters, our modular approach to the system design can, in principle, accommodate any reasonable method of acoustic-to-articulatory mapping that yields a plausible estimate of the speaker's VT area-function without undue computational complexity.

3. 3D MODELLING

The second major component of our system is responsible for generating a plausible 3D model based on the area-functions estimated by the first component. Notably, this second component has been deliberately designed to be independent of any particular algorithm for estimating the area-functions, and indeed directly-measured area-functions could in principle be substituted as input. In line with this modular approach, the system therefore assumes that the area-functions correspond to a simple tube of varying cross-sectional areas, which may subsequently be transformed to a more complex shape more closely matching that of the human vocal apparatus. The process (an example of which is shown later in Figures 2 and 3) involves the three subsystems of piecewise-tube construction, imposition of VT structural scaling, and VT path of airway transformation, as described respectively in the following three subsections.

Piecewise Tube

The first subsystem of the 3D modelling component generates, for each input area-function, a 3D linear tube of concatenated, circular sections of varying diameter. As shown in the left panel in Figure 1, a number of 2D circles in the x - y plane are defined, where each circle's radius is calculated directly from the corresponding section of the area-function, and where the number of (equi-spaced) points defining the perimeter of the circle (as x - y pairs) can be adjusted by the user (the more points, the smoother the image). At this stage of modelling, the entire tube is a straightened-out version of

the vocal tract, such that all points on a given circle have a constant z -coordinate which depends only on the distance of the corresponding cross-sectional area along the VT airway from the lips.

The user may also specify a different number of segments (circles) along the VT length, compared with the number of sections provided by the input area-function. In that case, cubic spline interpolation is used to provide the intermediate area values. This approach has the distinct advantage of yielding smoothed VT-shapes, thereby transcending the artificial, step-wise values generated by some methods of inversion (e.g., by the linear-prediction method), or indeed by direct measurement methods such as MRI. However, in that context it is important to note that the Mermelstein-Schroeder model summarised in Equations (1) and (2) already provides an area-function which is smooth, and which can be sampled at any desired number of points along its length.

As shown in the right panel in Figure 1, the points defining the 2D circles then form the vertices of a set of triangles that connect successive points around each circle as well as those points on adjacent circles, in a triangle strip configuration. For a model with s segments (circles) along the length of the vocal tract and p points around each segment (circle), the entire, 3D surface of the VT-airway is defined by a total of $s \cdot p$ vertices and $2p(s-1)$ triangles (polygons). Typical ranges of the parameter-values yielding visually very smooth shapes are 40-60 for s and 20-40 for p , yielding a model with 1500-4500 polygons.

Structural Scaling

The second subsystem of the 3D modelling component applies transformations to the circular cross-sections, with the aim of obtaining cross-sectional shapes that more closely match human anatomy. As a first approximation (see Figures 2 and 3), Lindblom and Sundberg's (1971, p.1173) numerical values for the 2-constants power-model are used. That model assumes somewhat more realistic, *elliptical* cross-sectional shapes, and defines three regions along the length of the VT within which the cross-dimensions of the airway in the midsagittal plane can be computed from the cross-sectional area values by a power-relation involving only two constants. Similarly to the other subsystems, spline interpolation of those constants is employed to extend Lindblom and Sundberg's original 3-region model to

a smoothed set of constant-pairs defined at each VT segment. Once the cross-dimension in the midsagittal plane is thereby determined at each section, the cross-sectional area can be used to compute the transverse dimension of the ellipse.

Path of Airway Transformation

The first two stages of the 3D modelling component yield a tube which is horizontally straight (in the z-dimension), and which therefore is visually still considerably different from the human vocal-tract. In particular, the human vocal-tract does not remain horizontal throughout its length but follows a path from lips to glottis in which it rotates through more than 90 degrees as well as experiencing varying horizontal and vertical offsets in the sagittal plane.

In order to model the VT centreline, and hence the gross shape of the vocal-tract itself, a set of direct measurements is required, which should correspond to the orientation and displacement (from the lips) of the VT at regular intervals. These were obtained by hand measurement of published midsagittal profiles for neutral vowels in an MRI-based study (Story et al., 1996, p.542, Fig.2). Measurements of the VT centreline were encoded as a set of value triplets taken at equi-spaced intervals along the length of the vocal-tract. For each such point the horizontal offset from the lips, vertical offset from the lips, and angle of rotation relative to the horizontal were recorded.

These values were then used to define a set of transformations to the 3D model's vertices, such that the new set of vertices are morphed (moved and rotated) to follow the centreline. As a first step, spline interpolation was again employed to ensure that there were as many centreline triplet values as there were segments in the VT model. All vertices corresponding to a particular segment of the VT were then transformed as per the corresponding centreline triplet, as follows:

$$v' = T_{z\text{-offset}} T_{y\text{-offset}} T_{x\text{-rotation}} T_{z\text{-zero}} v \quad (3)$$

Equation 3 specifies the transformation as a chained set of homogeneous matrix operations (transformations) on the original vertex to derive the new vertex v' . In particular, each vertex is first translated in the z-direction onto the same plane as the lips ($T_{z\text{-zero}}$), then rotated about the horizontal axis ($T_{x\text{-rotation}}$), and finally translated (or offset) in the vertical ($T_{y\text{-offset}}$) and horizontal ($T_{z\text{-offset}}$) directions (in the sagittal plane). The bottom panel in Figures 2 and 3 illustrates the result of these transformations, which yield a visually more realistic 3D image of the supralaryngeal VT airway.

4. MODEL VIEWING AND INTERACTION

In order to make full use of the 3D models constructed by the processes described above, it was felt that users should also have the opportunity to dynamically interact with and explore them. The purpose of the third major component of our system is therefore to enable visual interaction, whereby the user can view, interact with, move through, and manipulate the 3D model created by the first two components. To achieve this goal, two complementary approaches were taken: one employing an existing Web 3D standard and emphasising

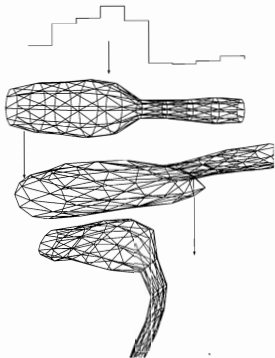


Figure 2: The three stages of the 3D modelling system, shown in wireframe form, from top to bottom. A VT area-function is first transformed into a piecewise tube with circular cross-sections. VT-structural scaling is then imposed which transforms the circular cross-sections to more plausible ones while retaining the area value at each section. Finally, a model of the VT airway is used to transform the straightened tube into a more natural, bent VT-shape.

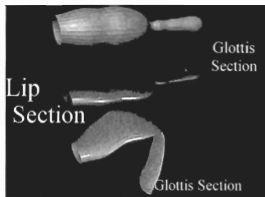


Figure 3: The three stages of the 3D modelling system, shown in solid-shaded mode, taken as a snapshot of a VRML world. The three models are the result of all three components of the system: acoustics-to-area function, 3D modelling (at various levels of complexity), and model viewing. The input data for the models consisted of the first 4 formant frequencies [669, 1241, 2736, 3356] Hz of an /a/ vowel from an adult Japanese speaker.

accessibility, the other employing immersive virtual reality technology.

The first, widely accessible solution, generates the model as VRML (Virtual Reality Modelling Language), a World Wide Web standard for 3D (ISO/IEC, 1997). Hence the models are made available on the Web (ref. URL-1) and can be viewed via any one of a number of free plug-ins for browsers such as Netscape Navigator or Microsoft's Internet Explorer. This provides access to anyone with a modern PC and internet connection, and, combined with the features of VRML, provides new opportunities for research and education in this area of speech processing (Barlow & Clermont, 2000). Figure 3 is an example of one such VRML model.

The second solution employs an immersive projection theatre known as the WEDGE (Gardner et al., 1999). Images are back-projected in stereo onto twin-screens (each 2.7 metres wide by 1.5 metres high) meeting at right-angles, providing a semi-immersive sense of presence. Within the Vee formed by the screens, the images, viewed through stereo goggles, are perceived as having true depth and are seen to "float" off the screens. As for the VRML solution, users may manipulate and interact with the WEDGE-projected 3D-VT model. However, the WEDGE system does not currently support viewing of VRML data, and hence the 3D models are saved in LightWave .obj format, for which a viewer is available.

5. DISCUSSION

We have proposed a methodology for modelling and interactively displaying a 3D representation of the human vocal-tract. While more sophisticated models of the vocal-tract which include the nasal cavities, sinus piriformis, and other physiological and articulatory structures have previously been proposed in the literature, the aim of our simplified modelling approach is to present a 3D representation which is at once visually plausible and computationally inexpensive to construct. Specific real-time applications envisaged for the model include foreign-language learning and training of individuals with speech pathologies, both of which would certainly benefit from a real-time computer display of the 3D VT-shapes produced by the user during phonation of certain speech sounds.

A number of avenues for further research remain and are currently being explored. Within the 3D modelling component a structural scaling system is being developed that employs MRI data to accurately model the fixed structure of the upper-palate. This will further improve the plausibility of the 3D model, which at times appears stretched and narrow due to the elliptical transformation currently employed to perform structural scaling. Efforts are also underway to reduce the complexity, in terms of number of polygons, of the 3D model by reducing the number of vertices in areas of near-linear shape. For the acoustic to area-function subsystem, alternates to formant frequency inputs are being explored. Perhaps most importantly, FEM (Finite Element Method) and perceptual experiments are planned to validate the models generated, with particular emphasis in the perceptual trials on the utility of the model for language acquisition.

Finally, a MATLAB implementation of the software for the 3D modelling and visualisation of the vocal tracts based on area-function data is freely available at URL-2. It consists of a suite of functions for carrying out the various stages in modelling outlined above, i.e., transforming area-functions to a variable-width model with circular cross-sections, morphing on the basis of VT-structure and -centreline, and viewing.

REFERENCES

- Barlow M, and Clermont F. (2000) "Seeing is Believing: Beyond a Static 2D-View of Formant Space for Research and Education", *Proc. Eighth Australian Int. Conf. on Speech Science and Tech.*, Canberra, Australia, 118-123.
- Coker, C. H. (1976). "A Model of Articulatory Dynamics and Control", *Proc. IEEE* **64**, 452-460.
- Gardner, H.J., Besiwell, R.W., and Whitehouse, D. (1999). "The WEDGE Immersive Projection Theatre", *Proc. 4th International SimTecT Conf.* 383-385.
- Harshman, R., Ladefoged, P. and Goldstein, L. (1977). "Factor analysis of tongue shapes", *J. Acoust. Soc. Am.* **62**, 693-707.
- ISO/IEC (1997) "VRML 97", International Specification ISO/IEC IS 14772-1, www.vrml.org.
- Lindblom, B. E. F. and Sundberg, J. E. F. (1971). "Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement", *J. Acoust. Soc. Am.* **50**, 1166-1179.
- Mermelstein, P. (1967) "Determination of vocal-tract shape from measured formant frequencies", *J. Acoust. Soc. Am.* **41**, 1283-1294.
- Mermelstein, P. (1973). "Articulatory model for the study of speech production", *J. Acoust. Soc. Am.* **53**, 1070-1082.
- Mermelstein, P., and Schroeder, M. R. (1965). "Determination of smoothed cross-sectional area functions of the vocal tract from formant frequencies", *Proc. 5th Int. Cong. Ac.*, Liège, Paper A24.
- Story, B.H., Titze, I.R., Hoffman, E.A. (1996). "Vocal Tract Area Functions from Magnetic Resonance Imaging", *J. Acoust. Soc. Am.* **100**, 735-554.
- Story, B. H. and Titze, I. R. (1998). "Parameterization of vocal tract area functions by empirical orthogonal modes", *J. Phonetics* **26**, 223-260.
- URL-1: <http://www.cs.adfa.edu.au/~spike/RFC/wch/VisualSpeech/index.html#VT>
- URL-2: <ftp://ftp.cs.adfa.edu.au/pub/users/spike/matlabVT.zip>.
- Yang, C.-S. and Kasuya, H. (1994). "Accurate measurement of vocal-tract shapes from magnetic resonance images of child, female and male subjects", *Proc. Int. Conf. Spoken Lang. Process.* Yokohama, Japan, 623-626.
- Yehia, H. C., Takoda, K. and Itakura, F. (1996). "An Acoustically Oriented Vocal-Tract Model", *IEICE Trans. Inf. & Syst.* **E79-D(8)**, 1198-1208.



SOUND SEPARATION WITH A COCHLEAR IMPLANT AND A HEARING AID IN OPPOSITE EARS

Peter J. Blamey¹, Christopher J. James¹ and Lois F.A. Martin²

¹Department of Otolaryngology, University of Melbourne

²Bionic Ear Institute

ABSTRACT - Two experiments were conducted to investigate the perception of speech and noise presented simultaneously to three subjects with impaired hearing in five monaural and binaural conditions. A broadband noise was found to have no effect on speech perception when the two signals were presented to opposite ears. When speech and noise were presented to the same ear(s), speech perception scores on a closed-set test fell from above 95% at high signal-to-noise ratios (SNR) to 71% at an SNR of about -5 dB. When two speech signals were presented simultaneously at equal intensities (0 dB SNR) speech perception scores fell to 75% or lower, regardless of the ear(s) to which the signals were presented. Thus dichotic presentation helped these listeners to separate speech from a broadband noise, but not to separate two simultaneous speech signals produced by different speakers.

1. INTRODUCTION

Under normal conditions, listeners hear sounds from more than one source at a time, such as a voice and environmental noise, or several voices speaking at the same time. At least three distinct physical characteristics of the signals are used to separate these sounds perceptually (spatial, spectral, and temporal). Spatially separated sources can be localised and distinguished from one another using interaural timing differences at low frequencies and interaural intensity differences at high frequencies (Rayleigh, 1907). There are also perceptual mechanisms that can be used to distinguish sounds that are heard simultaneously with only one ear. For example, these mechanisms make use of simultaneous onsets and comodulation of different frequency components to form separable streams of auditory information (Bregman, 1990). Another example is the separation of complex sounds that have different fundamental frequencies as in the case of human voices (Darwin & Gardner, 1986).

The focus of the present study is the use of temporal and spectral cues for the separation of sounds presented binaurally or monaurally to listeners who use a cochlear implant in one ear and a hearing aid in the other. With impaired hearing, both of these cues may be degraded by poor temporal and/or spectral resolution in the monaural condition. In the binaural condition there may be additional complications if the hearing loss is asymmetrical and/or there are different temporal delays in the processing in the two ears. The cases to be considered in this report are particularly interesting because the acoustic presentation of sound via a hearing aid to one ear and the electric presentation of the same sound via a cochlear implant to the other ear is almost certain to introduce both temporal and spectral differences between the ears. It is also unlikely that cochlear implant users will have access to fine spectral cues such as those needed to group together harmonically

related frequency components.

Without entering into a detailed consideration of proposed models and mechanisms of binaural and monaural separation of sounds, two hypotheses seem plausible: a) that two simultaneous sounds presented dichotically (both sounds to both ears) will be more difficult to separate than the same two sounds presented dichotically (one sound to each ear), b) that two simultaneous sounds with generally similar spectral and temporal characteristics (such as two voices uttering a word) will be more difficult to separate than grossly different sounds (such as a voice and a broadband noise).

2. METHOD

Participants and processors

The three participants in this pilot study were post-linguistically deafened adults who used a multiple-electrode (Cochlear Limited) cochlear implant in one ear, and who had residual hearing in the non-implanted ear. Table 1 summarises the relevant audiological information for the non-implanted ear of each participant, together with some of the factors known to have an effect on speech perception scores for individual cochlear implant users (Blamey et al, 1992, 1996). Two of the participants normally used a hearing aid and cochlear implant together, while the other wore only the cochlear implant. For the purposes of this experiment, all participants were fitted with a benchtop hearing aid based on a Motorola DSP 56303 evaluation module with additional microphone and amplifier circuits to drive an Oticon AN270 button receiver in the non-implanted ear. If the participant normally wore a hearing aid, the fixed linear gain of the benchtop hearing aid was set to equal the gain of the participant's own aid (within 2 dB) across the frequency range from 125 Hz to 4 kHz as measured in a hearing aid test box. If the participant did not normally wear a hearing aid, the fixed linear gain of the benchtop hearing aid was set according to

the NAL prescription with correction for severe and profound hearing losses (Byrne & Dillon, 1986). The implant signals for this study were presented via the participants' own Sprint cochlear implant speech processors with their usual speech processing strategies as listed in Table 1.

Table 1. Summary of audiological details for the participants.

Participant	S1	S2	S3
Hearing Loss at 250 Hz	65 dB HL	75 dB HL	30 dB HL
500 Hz	105 dB HL	75 dB HL	60 dB HL
1 kHz	110 dB HL	85 dB HL	120 dB HL
2 kHz	115 dB HL	120	No response
4 kHz	120 dB HL	No response	No response
Hearing Aid Use	Yes	No	Yes
Hearing Aid Ear	Left	Left	Right
Age	66 years	48 years	65 years
Aetiology	Progressive	Genetic	Progressive
Duration of hearing loss	22 years	25 years	16 years
Implant experience	7y 5m	5y 3m	1y 1m
No. of electrodes in use	16	19	6
Mean dynamic range	40.6 SL	38.4 SL	31.9 SL
Speech processor strategy	Mpeak	ACE	CIS
Implant Model	CI-22	CI-24M	

Separation of two voices

In the first experiment, speech recognition scores for the two voices were used as a comparative measure of sound separation in the five experimental conditions, on the assumption that better sound separation would result in higher word recognition scores. Stimuli were spondee (bisyllabic words with equal stress on each syllable) spoken by two different speakers: a male adult and a female adult who are also the second and third authors of this report. These words were chosen from a set of 40 spondee words that had already been recorded for each of the two speakers, and were selected to be easy to distinguish from one another in both of the monaural conditions to be tested ie hearing aid alone and cochlear implant alone. This was necessary because two of the participants had very poor speech recognition in the hearing aid ear. The RMS levels of the digitally recorded speech tokens were equalised. Each word was stored in a single channel of a wave file with the onset of the word within 10 ms of the start of the file.

Computer software and hardware were set up so that the stimuli could be mixed and presented via a line input directly to the benchtop hearing aid and the cochlear implant speech processor to produce the monaural, diotic, and dichotic signals required for this study. Five conditions were tested: HA in which the two voices were mixed and presented to the hearing aid only; CI in which the two voices were mixed and presented to the cochlear implant only; DIOTIC in which the two voices were mixed and presented to both the cochlear implant and the hearing aid simultaneously; MCFIHA in which the male voice was presented to the cochlear implant and the female voice to the hearing aid dichotically; and FCIMHA in which the female voice was presented to the cochlear implant and the male to the hearing aid dichotically.

Prior to starting the experiment, the individual spondee spoken by each speaker were presented in the HA, CI and

DIOTIC conditions in a practice procedure. The input levels to the hearing aid and the cochlear implant speech processor were adjusted individually so that the loudness of the stimuli was equal in each ear, with the overall loudness at a comfortable listening level. It was also checked that each participant could recognise which speaker presented each of the spondee, and that they could recognise the individual spondee presented. This was done by presenting a block of stimuli (4 of each spondee) in a random order and asking the participant to select the spondee spoken from a list. If the participant scored over 95% correct in all three of the HA, CI and DIOTIC conditions, the participant was judged to be ready for the more difficult experiment involving simultaneous presentations of the two voices. Participant S2 was able to perform this task with a set of 10 spondee (5 for each speaker). S1 and S3 performed the task with 6 spondee (3 for each speaker).

The combined stimuli were presented to S1 and S3 in blocks of 18 (2 x 9 combinations) and to S2 in blocks of 50 (2 x 25 combinations) in a random order. Two seconds before each presentation trial, either the male words or the female words were displayed on a computer screen, together with a heading "MAN'S VOICE" or "WOMAN'S VOICE," respectively. After the trial, the participant was asked to respond with the word from the list on the screen that had been spoken by the indicated speaker. For each of the combinations within a block, the participant was asked once for the male speaker's word and once for the female speaker's word. Two to five blocks of trials were presented to each participant in each of the five conditions listed above.

Results for the first participant indicated that the diotic score was slightly higher than the two dichotic scores, contrary to hypothesis a). It seemed possible that the participant may have had difficulty in switching his attention rapidly from one ear to the other or from one voice to the other, so the task was repeated with a different blocking structure. Blocks of 27 stimuli were presented in each condition (3 x 9 combinations). Within each block, the participant was always asked to respond with the same speaker's word ie always the male speaker, or in another block always the female speaker. Thus the participant did not need to switch his attention between ears or between speakers within a block of stimuli.

Separation of a voice and a noise

In this experiment, the relative levels of speech and noise were varied adaptively to find the signal-to-noise ratio (SNR) where 71% of the words were recognised correctly. It is assumed that separation of the speech (and recognition of the speech) becomes more difficult as the SNR decreases. The stimuli were three spondee spoken by the female speaker with the carrier phrase "The next word is ...". The same three spondee were used as in the first experiment ie "teapot", "drawbridge" and "football", but these were different tokens recorded with the carrier phrase. Each recorded stimulus was set to the same RMS amplitude. The stimuli were presented in a persistent

background of speech-shaped broadband noise under 7 different conditions. As in the first experiment HA, CI and DIOTIC conditions were used, where the speech and noise were mixed and presented to one or both ears. In the dichotic NCIFHA condition, noise was presented to the implanted ear and the female voice to the hearing aid. In the dichotic FCINHA condition, the noise was presented to the hearing aid and the woman's voice to the implanted ear. The remaining conditions (HA0 and CI0) were carried out with no noise and the voice presented to the hearing aid or cochlear implant, respectively.

The SNR was varied in each condition using an adaptive procedure in which the SNR was increased by 2 dB every time the listener responded incorrectly, and decreased by 2 dB after two correct responses in a row (Levitt, 1971). This up-down procedure oscillates about the SNR where the listener scores 71% correct. The chance score in this 3-alternative-forced-choice task is 33%. The numerical values of SNR refer to the ratios of RMS amplitude for the speech and the speech-shaped noise. The adaptive procedure was terminated after 6 turning points had been found, and the average of the last four turning points was taken as the asymptotic SNR. Because it was expected that the noise might have little effect in the dichotic conditions and we did not wish to present uncomfortably loud sounds, the SNR was reduced by reducing the speech level when the SNR was negative, and by increasing the noise level when the SNR was positive. Thus the procedure would typically start with the speech at the comfortable level. The adaptive procedure would increase the noise until it was at the same RMS level as the speech (0 dB SNR) and then the level of the speech would start to decrease. In the HA0 and CI0 conditions, the level of the speech was decreased to find a speech reception threshold with no noise.

3. RESULTS

Separation of two voices

The two different blocking methods produced no significant differences in the results obtained in any condition or for any subject, so the results were combined. Figure 1 shows the percentage of correct responses for each subject in each of the 5 conditions. It is clear that the listeners were unable to separate the two voices completely as the scores drop significantly below 95% in all conditions. On the other hand, the scores are all significantly above the chance score of 33%. Analysis of Variance (ANOVA) indicated that there were no significant differences between any of the five conditions shown in Figure 1. The ANOVA used subject, condition, and blocking method as independent variables.

Separation of a voice and a noise

Figure 2 shows the asymptotic SNR values for each subject in each of the 7 conditions tested. Each value shown is the mean of two or more adaptive procedures. For positive SNRs, the speech is at a comfortable loudness and the RMS level of the

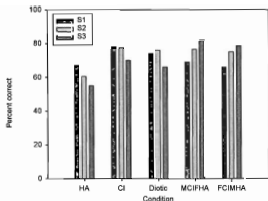


Figure 1. Percentages of correct responses by subject and condition in the separation of voices experiment.

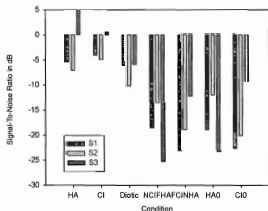


Figure 2. Mean asymptotic signal-to-noise ratio in each condition of the separation of speech and noise experiment for each subject.

noise is lower than the RMS level of the speech by the indicated number of dB. For negative SNRs, the noise is at a comfortable loudness, and the RMS level of the speech is lower than the RMS level of the noise by the indicated number of dB. High SNR indicates poor separation of speech and noise. Low SNR indicates good separation of speech from noise.

ANOVA with subject and condition as independent variables was followed by post-hoc t-tests using the Bonferroni method to compare the SNRs in the different conditions. The mean SNRs for the CI, Diotic, and HA conditions were not significantly different from one another ($p > 0.05$). The mean SNRs for the FCINHA, NCIFHA, CI0, and HA0 conditions were not significantly different from one another ($p > 0.05$). However all the SNRs for the first 3 conditions were significantly different from all the SNRs in the second group ($p < 0.001$).

4. DISCUSSION

Hypothesis a) that it is more difficult to separate sounds diotically than dichotically.

Hypothesis a) predicts that the speech perception scores in the voice separation experiment should be higher for the dichotic conditions than the diotic condition, and the SNRs for the dichotic conditions should be lower than for the diotic condition in the speech and noise experiment. The hypothesis was supported by the results when one of the sounds was a broadband noise and the other was speech. The hypothesis was rejected when the sounds were both speech signals. The result for speech and noise is consistent with masking experiments in which a masker has a much greater effect on a probe in the same ear than on a probe in the opposite ear (Zwislocki, 1972). When two speech signals are presented together, interference takes place regardless of the ear(s) of presentation. This result is not consistent with known monaural and binaural masking effects. It is more consistent with the binaural experiments of Studdert-Kennedy and Shankweiler (1970), which suggest that speech features from the two ears are processed independently and then recombined at one location in the left hemisphere (of almost all right-handed listeners and most left-handed listeners). The interference between the two signals probably occurs at the recombination stage where there is usually a small right-ear advantage, but otherwise the speech features derived from the two ears are treated equivalently. In the dichotic conditions of the two-voice experiment, S1 and S3 had higher scores for the words presented to the hearing aid, and S2 scored higher for words presented to the implant. This was a right ear advantage for S2 and S3, and a left ear advantage for S1.

Hypothesis b) that it is more difficult to separate two voices than to separate a voice and a broadband noise.

This hypothesis was supported by the results. In fact, the equality of the SNRs in the NCIFHA and FCINHA conditions with the HA0 and CI0 conditions, respectively, demonstrates that the broadband noise had no measurable effect on the perception of the speech signal when the voice and the noise were presented to opposite ears. In the HA, CI, and Diotic conditions, the mean SNRs were negative for a level of 71% correct. Thus the scores at zero dB would have been greater than 71% at zero dB SNR. The percentages correct were 59%, 75%, and 72%, respectively in the two-voice experiment (at zero dB signal to noise ratio).

It is interesting to note that Armstrong et al (1997) found a significant advantage for diotic listening with an implant and hearing aid together over monaural listening with a cochlear implant for open-set sentence perception in quiet and in 8-talker babble at an SNR of 10 dB. In the present study, there were no statistically significant differences between the diotic and monaural conditions in either experiment. The discrepancy between the studies may be due to the different materials and noise used, or the low number of subjects in the present study.

5. CONCLUSIONS

The separation of sounds into streams by the auditory system involves different mechanisms for dynamic speech signals and stationary broadband noises. These effects which are observed in normally hearing listeners are also present in subjects who have a hearing aid in one ear and a cochlear implant in the other. The separation and fusion of sounds presented to different ears, and the potential advantage of one ear or one device over the other may have consequences for the development of binaural processors for people with impaired hearing.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the National Health and Medical Research Council of Australia (project grant 970257), the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and the Bionic Ear Institute. The research was conducted under the auspices of the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (project 96/289H). The authors also gratefully acknowledge the valuable contribution of the participants who volunteered their time for these experiments.

REFERENCES

- Armstrong, M., Pegg, P., James, C. & Blamey, P.J. (1997) "Speech perception in noise with implant and hearing aid," *Am. J. Otolaryngol.* **18**, S140-S141.
- Blamey, P.J., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Faer, G., Larky, J., Lindstrom, B., Nedzelski, J., Peterson, A., Shipp, D., Staller, S. & Whitford, L. (1996) "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants," *Audiology & Neuro-Otology* **1**, 293-306.
- Blamey, P.J., Pyman, B.C., Gordon, M., Clark, G.M., Dowell, R.C. & Hollow, R.D. (1992) "Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients," *Annals Otol. Rhinol. Laryngol.* **101**, 342-348.
- Bregman, A.S. (1990) *Auditory scene analysis*, (MIT Press: Cambridge, MA).
- Byrne, D. & Dillon, H. (1986) "The National Acoustics Laboratory (NAL) new procedure for selecting the gain and frequency response of a hearing aid," *Ear & Hearing* **7**, 257-265.
- Darwin, C.J. & Gardner, R.B. (1986) "Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality," *J. Acoust. Soc. Am.* **79**, 838-845.
- Levitt, H. (1971) "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467-477.
- Lord Rayleigh (1907) "Our perception of sound duration," *Phil. Mag.* **13**, 214-232.
- Studdert-Kennedy, M. & Shankweiler, D. (1970) "Hemispheric specialization for speech perception," *J. Acoust. Soc. Am.* **48**, 579-594.
- Zwislocki, J.J. (1972) "A theory of central auditory masking and its partial validation," *J. Acoust. Soc. Am.* **52**, 644-659.



CODING WIDEBAND SPEECH AT NARROW-BAND BIT RATES

J.R. Epps and W.H. Holmes

School of Electrical Engineering and Telecommunications,
The University of New South Wales

ABSTRACT. The 'muffled' quality of coded speech, which arises from the bandlimiting of speech to 4 kHz, can be reduced either by coding speech with a wider bandwidth or by wideband enhancement of the narrowband coded speech. This paper investigates the limitations of wideband enhancement and possibilities for its improvement. A new wideband coding scheme is proposed that is based on any narrowband coder, but augmented by wideband enhancement plus a few bits per frame of highband information. The scheme thus has a bit rate only slightly greater than that of the narrowband coder. Subjective listening tests show that this scheme can produce wideband speech of significantly higher quality than the narrowband coded speech.

1. INTRODUCTION

The need for wideband speech transmission arises both from an ongoing requirement for improved speech quality in all types of services, and from the specific needs of applications such as hands-free and Internet telephony. One solution is to code the parameters of wider bandwidth speech, which leads to a substantial increase in the bit rate relative to narrowband coders (Schnitzler, 1998).

An alternative is to employ wideband enhancement (Makhoul and Berouti, 1979; Carl and Heute, 1994; Cheng et al., 1994; Yoshida and Abe, 1994; Chan and Hui, 1996; Enbom, 1998; Epps and Holmes, 1998 and 1999; Epps, 2000; Jax and Vary, 2000), a technique which synthesizes wideband speech based on pitch, voicing and spectral envelope information in the narrowband speech. Wideband enhancement requires no increase in bit rate, but the quality of the output wideband speech is poorer than that resulting from wideband coding due to less accurate highband spectral envelope estimates. In this paper, an assessment is made of the limits to the accuracy of highband envelope estimation under realistic test criteria.

A new technique for wideband coding is also proposed which is based upon a combination of the wideband enhancement paradigm with any narrowband coder. Wideband speech of higher quality than that produced by wideband enhancement alone is obtained by allocating just a few bits per frame for highband spectral coding. This means that the new wideband coder has a bit rate only slightly greater than that of the narrowband coder.

Section 2 examines the potential performance of wideband enhancement envelope estimation, section 3 reviews selected literature on very low bit rate wideband spectral coding, section 4 presents a new technique for coding wideband speech at near narrowband bit rates, and section 5 details subjective test results of various schemes.

2. WIDEBAND ENHANCEMENT

Wideband enhancement is a scheme which adds a synthesized highband signal to narrowband speech to produce a wideband speech signal, as shown in Fig. 1. The synthesis of the highband signal is based entirely on the information available

in the narrowband speech. Note that the narrowband speech is not re-synthesized, since it is assumed to be of sufficiently high quality.

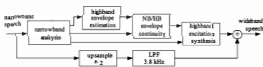


Figure 1. Overall scheme of wideband enhancement

Highband Excitation Synthesis

Previous research (Epps and Holmes, 1998) has shown that a combination of sinusoidal and random excitation can be used to produce high quality highband excitation estimates. In this technique, based on the sinusoidal transform coding (STC) harmonic model (McAulay and Quatieri, 1995), sinusoidal highband excitation is synthesized from the narrowband speech using pitch, voicing and highband spectral envelope estimates. This technique gives perfectly harmonic periodic excitation, with the amplitudes of the sinusoidal components determined directly by the spectral envelope. Random excitation, at an amplitude controlled by the narrowband degree of voicing, is employed to model the highband unvoiced components. This approach was found to accurately represent the voiced/unvoiced mix of a wide variety of different speech frames. The use of STC-derived parameter interpolation methods produced smooth variation of the sinusoid frequencies and phases between frames. These features contributed to a better perceptual performance of the novel STC-based excitation than other methods such as spectral folding (Makhoul and Berouti, 1979).

Highband Envelope Estimation

Different techniques for estimating the shape of the highband spectral envelope have also been considered, with codebook mapping (Gersho, 1990; Carl and Heute, 1994) performing well under a spectral distortion comparison (Epps and Holmes, 1999). As seen in Fig. 2, codebook mapping estimates the highband spectral envelope by selecting the

highband code vector whose corresponding narrowband code vector has the most similar envelope shape (in a spectral distortion sense) to the input narrowband envelope. Details of the codebook design from training data can be found in the work of Carl and Heute (1994) and Epps (2000). Other methods of highband envelope estimation include statistical recovery (Cheng et al., 1994), codebook mapping with codebooks split by voicing (Epps and Holmes, 1999), and codebook mapping based upon hidden Markov models (Jax and Vary, 2000).

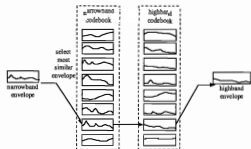


Figure 2. Codebook mapping for highband envelope estimation

Narrowband-Highband Envelope Continuity

Preserving the continuity of the spectral envelope between the narrowband envelope and the estimated highband envelope is an important perceptual requirement, however in instances where the accuracy of the highband envelope estimation is poor, the resulting highband spectral distortion can be quite large. Typically the estimated highband envelope is matched to the narrowband envelope either at a single frequency or over a range of frequencies, depending on the size of the overlap between the two envelopes, but there are alternative techniques (Epps, 2000).

Limits to Wideband Enhancement

Highband envelope estimation is based upon the assumption that two wideband spectral envelopes with similar narrowband envelope shapes will also have similar highband envelope shapes. One method for testing the validity of this assumption is to select two pairs of narrowband-highband spectral envelopes from independent speech databases. Their narrowband spectral distortion is then calculated and their highband spectral distortion is also computed, after ensuring that the second highband envelope is properly matched to (i.e. continuous with) the first narrowband envelope. This procedure is then repeated for all combinations of envelope pairs from the two speech databases.

The resulting data can be used to gain an idea of the distribution of continuity-adjusted highband spectral distortion results which could be expected from any codebook mapping scheme with a given maximum narrowband spectral distortion. Figure 3 illustrates these distributions, showing that highband spectral distortion is weakly correlated with narrowband spectral distortion. Present wideband enhancement techniques produce average continuity-adjusted

highband spectral distortions of around 6.4 dB (Epps, 2000) using narrowband codebooks with a maximum narrowband distortion of around 5.8 dB, and are thus slightly better than the median expected performance.

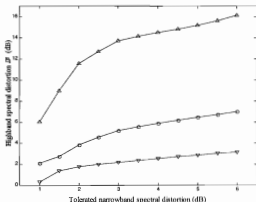


Figure 3. Distribution of continuity-adjusted highband spectral distortion as a function of tolerated (maximum) narrowband spectral distortion. Results are based on around 5(106 data points. The percentile contours shown are the 10% (-), 50% or median (o), and 90% (□). Note that there are relatively few data points for small values of tolerated narrowband spectral distortions.

Good performance (median highband spectral distortion 2 dB) with codebook mapping schemes is therefore possible only if the narrowband spectral distortion can be contained to around 1 dB. This would require codebooks consisting of around 2^{10} vectors, a size which is not feasible to implement under present storage constraints. It is concluded that the practical performance of highband envelope estimation methods is only likely to be improved, compared to existing methods, by allowing some knowledge of the original highband speech, rather than relying entirely on narrowband information. This is the subject of the following sections.

3. EXISTING TECHNIQUES FOR VERY LOW BIT RATE WIDEBAND CODING

Low Order Highband LP Coding

In some previous coder implementations (McElroy et al.; 1993, Seymour and Robinson, 1997), the highband (4-8 kHz) spectral envelope was coded using a 2nd order LP envelope. The LP parameters and highband gain are quantized using 440-500 bit/s (plus quantized highband excitation), which is all additional to the narrowband coding scheme.

Flat Highband Envelope Coding

If the narrowband is defined to be the 0-6 kHz frequency range, a 6-7 kHz highband envelope can be sufficiently well represented by a flat spectrum at the correct gain. In (Schnitzler, 1998) this gain is encoded on a sub-frame basis using 3 bit MA prediction (an additional 1.2 kb/s). Random excitation was used in the highband.



Figure 4. Proposed wideband encoder and decoder based on wideband enhancement

Wideband Enhancement and Gain Coding

The use of highband gain coding in conjunction with wideband enhancement was suggested in (Enbom, 1998). Here the (4-8 kHz) highband envelope was estimated from the narrowband envelope using codebook mapping. The gain was coded as the difference $G_{hb} - G_{nb}$ (in dB) between the highband and narrowband gains G_{hb} and G_{nb} , which were quantized on a linear scale and coded using 4 bits per frame (an additional 700 bit/s). In this case, highband excitation was provided by spectral folding (Makhoul and Berouti, 1979).

4. A NEW WIDEBAND CODER

Overview

The proposed structure of the new wideband speech coder is based around any narrowband coder, as seen in Fig. 4. The decoded narrowband information, in addition to some coded highband spectral envelope and gain information, is used to synthesize the wideband signal at the decoder. In the highband envelope and gain coding technique presented in section 4.3, narrowband-highband mapping is combined with coding. Highband excitation synthesis is discussed in section 4.2. This configuration is well suited to coders which must primarily satisfy a set of narrowband performance criteria, but which can accommodate a few bits per frame of highband information. It also allows bit allocation trade-offs to be made between the narrowband and highband in a similar fashion to sub-band or split band coding.

Highband Excitation Synthesis using a Sinusoidal Model

Earlier research has shown that high quality mixed voiced/unvoiced highband excitation may be synthesized using the STC-based approach discussed in section 2.1. In informal listening tests this approach was considered superior in quality to spectral folding (Makhoul and Berouti, 1979) or purely random excitation in the highband. If a sinusoidal narrowband codec is employed, the highband excitation could be efficiently synthesized at the decoder concurrently with the narrowband excitation.

Wideband Enhancement with Classified Highband Codebooks

A new method for highband spectral coding employs vector quantization of the highband envelope and gain using a small partition of a full highband codebook, where the selection of vectors comprising the partition is based upon the shape of the narrowband envelope. The index i of the narrowband code vector most similar to the input narrowband spectral envelope is first determined. Each narrowband code vector contains 2^n indices to the highband codebook, where n is the number of highband bits employed. The input highband envelope and

highband gain are compared with all 2^n vectors in the highband codebook whose indices are contained in the narrowband code vector with index i , and the most similar highband code vector is chosen. The index of the chosen highband code vector is then coded using n bits. This method is illustrated in Fig. 5. The use of as few as one, two or three highband bits to select between many highband candidate envelopes substantially reduces the highband spectral distortion resulting from the codebook mapping scheme discussed in sections 2.2 to 2.4.

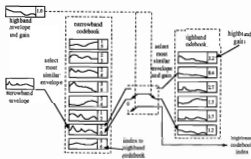


Figure 5. Block diagram example of a classified codebook mapping-based highband envelope and gain encoder using one highband bit per frame.

5. SUBJECTIVE ASSESSMENT

18 listeners (16 male and 2 female, between the ages of 20 and 35) were each presented with 70 randomized samples of speech prepared using various methods for determining the highband excitation and envelope. Codebooks of size 1024 were employed in wideband enhancement, while speech coded according to section 4 used a narrowband codebook of size 1024 and a highband codebook of size 8192 (i.e. three highband bits per frame). The resulting preliminary mean opinion scores (MOS) are shown in Table 1, where the 95% confidence interval is (0.15).

These results show that with only a few bits per frame, wideband spectral coding can be achieved at close to the quality of the original wideband speech. While the sinusoidal-based synthetic excitation performed reasonably well when combined with the original highband spectral envelopes, this excitation generally attracted lower scores, indicating that more attention still needs to be paid to perceptual artifacts arising from its implementation.

Table 1. MOS results

Condition	MOS
Original wideband speech	4.25
Wideband coded speech, 3 highband bits per frame, original highband excitation	3.99
Wideband enhanced speech, original highband excitation, synthetic highband envelope	3.65
Wideband enhanced speech, original highband envelope, synthetic highband excitation	3.31
Wideband coded speech, 3 highband bits per frame, synthetic highband excitation	2.83
Wideband enhanced speech, synthetic highband excitation and envelope	2.78
Original narrowband speech	2.74

6. CONCLUSION

A new scheme for wideband speech coding at a bit rate only slightly greater than that of narrowband coding has been presented. This scheme is based on wideband enhancement techniques, but improves upon these by transmitting a small amount of highband spectral envelope and gain information. Listening test results show that a significant quality improvement over narrowband speech can be achieved using this scheme.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Motorola Australia Pty Ltd, and wish to express their thanks to Dr M.M. Thomson for his advice and encouragement.

REFERENCES

- Carl, H., and Heute, U. (1994). "Bandwidth enhancement of narrowband speech signals", *Signal Processing VII, Th. and Appl.*, EUSIPCO, 2, 1178-1181.
- Chan, C-F., and Hui, W-K. (1996). "Wideband enhancement of narrowband coded speech using MBE re-synthesis", *Proc. Int. Conf. on Signal Processing, ICSP 1*, 667-670.
- Cheng, Y.M., O'Shaughnessy, D., and Mermelstein, P. (1994). "Statistical recovery of wideband speech from narrowband speech", *IEEE Transactions on Speech and Audio Processing*, 2(4) 544-548.
- Enbom, N. (1998). *Bandwidth Expansion of Speech*. Thesis, Royal Institute of Technology (KTH).
- Epps, J.R., and Holmes, W.H. (1998). "Speech enhancement using STC-BESSEJ bandwidth extension", *Proc. ICSLP (Sydney, Australia)*, 519-522.
- Epps, J.R., and Holmes, W.H. (1999). "A new technique for wideband enhancement of narrowband coded speech", *Proc. IEEE Workshop on Speech Coding (Porvoo, Finland)*, 174-176.
- Epps, J.R. (2000). *Wideband Enhancement of Narrowband Speech*, Ph.D. Thesis (submitted), The University of New South Wales.

Gersho, A. (1990). "Optimal nonlinear interpolative vector quantization", *IEEE Trans. Comm.*, 38(9), 1285-1287.

Jax, P., and Viry, P. (2000). "Wideband extension of telephone speech using a Hidden Markov Model", *Proc. IEEE Workshop on Speech Coding (Delavan, Wisconsin)*.

Makhouli, J., and Berouti, M. (1979). "High frequency regeneration in speech coding systems", *Proc. ICASSP (Washington D.C.)*, 428-431.

McAulay, R.J., and Quatieri, T.F. (1995). "Sinusoidal coding", in *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal (Eds), Elsevier, Amsterdam, Chapter 4, pp. 121-173.


McElroy, C., Murray, B., and Fagan, A.D. (1993). "Wideband speech coding in 7.2 kb/s", *Proc. ICASSP*, 2, 620-623.

Schnitzler, J. (1998). "A 13.0 kbit/s wideband speech codec based on SB-ACELP", *Proc. ICASSP*, 1, 157-160.

Seymour, C.W., and Robinson, A.J. (1997). "A low-bit-rate speech coder using adaptive line spectral frequency prediction", *Proc. EUROSPEECH (Rhodes, Greece)* 3, 1319-1322.

Yoshida, Y., and Abe, M. (1994). "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping", *Proc. ICSLP (Yokohama, Japan)*, pp. 1591-1594.





The Australian Speech Science and Technology Association Inc. (ASSTA) aims to advance the understanding of speech science and its application to speech technology in a manner appropriate for Australia. Accordingly, we seek to facilitate significant exchange between speech scientists and speech technologists; between ASSTA and the community in general; and between ASSTA and other national and international bodies with related aims. Annual membership fees (including GST): \$44 (ordinary members), \$33 (associate members), \$11 (student members). Full details can be found at: www.assta.org.

Financial members are offered:

- discount registration for the biennial SST conferences, plus any other Australian or overseas conferences co-sponsored by ASSTA;
- a free copy of SST proceedings;
- a free hard copy of the ASSTA research directory each time this is published;
- regular electronic bulletins;
- quarterly newsletters;
- free personal advertisements in the newsletter;
- free personal or research group entries in the ASSTA WWW directory service;[DB1]
- eligibility to apply for a range of financial assistance schemes - travel funds for overseas conferences, research event funding etc.

RAPID CHANNEL COMPENSATION FOR SPEAKER VERIFICATION IN THE NIST 2000 SPEAKER RECOGNITION EVALUATION

J. Pelecanos and S. Sridharan

Speech Research Lab, RCSAVT

School of Electrical and Electronic Systems Engineering

Queensland University of Technology

ABSTRACT: A technique is proposed for rapidly compensating for channel effects of telephone speech for speaker verification. The method is generic and can be applied to both the one and two speaker detection tasks without re-training the separate systems. The technique has the advantages that it can be performed in real time (except for the small initial buffering), it does not suffer from a relatively long settling time such as certain RASTA processing techniques, and in addition, it is computationally efficient to apply. Results of the application of this technique to the NIST 2000 Speaker Recognition Evaluation are reported.

1. INTRODUCTION

Speaker Verification is the process of accepting or rejecting the claimed identity of a speaker based on a sample of their voice. Applications of speaker verification include secure building access, credit card verification and over-the-phone security access. High performance speaker recognition has been achieved under controlled laboratory and office recording conditions (Liou and Mammone, 1995) and is suitable for practical implementation under these circumstances. Unfortunately, performance of these systems severely degrades under adverse environmental and mismatched conditions. High performance speaker verification performed over the telephone network is consequently a challenging task. In the recent NIST Speaker Recognition evaluation (NIST, 2000), the recognition performance reported for matched recording conditions is significantly better than mis-matched tests and the latter remains a formidable challenge. The NIST evaluation is an annual international event aimed at advancing the state-of-the-art technology in speaker recognition. A large portion of research has been directed at minimising the effects of varied channels and handsets. Of interest in this research, is the compensation of multiple channel sources with the aim of enhancing recognition performance. In addition, there is a goal of not retraining a speaker recognition system for different speaker detection scenarios. A constraint in this experiment requires the channel compensation technique to perform well under the one and two speaker detection tasks. In this way, once a speaker model is obtained, there is no need to re-evaluate it given a different testing scenario.

The two scenarios of interest are the one and two speaker detection tasks. The one speaker detection task is the most basic. It is the process of accepting or rejecting the claimed identity of a speaker from their voice signal when the voice signal contains the content of a single speaker. In contrast, with two speaker detection, the speech signal contains up to

two speakers, one of which may be the target speaker. In the NIST 2000 evaluation (NIST, 2000), the two-speaker test utterance is formed by the addition of the two channels of the speaker conversation into a single channel. Compensating for channel effects is now more difficult. This is due to there being two separate sources of speech, with each source being affected by a different channel.

We propose a computationally efficient method of performing channel compensation on the speech with one or more speakers present in the voice segment content. In addition, we compare the performance of this method across both the one and two speaker detection tasks with varied window lengths. These experiments utilised the speaker recognition system submitted by the authors for the NIST 2000 evaluation.

2. CHANNEL COMPENSATION AS APPLIED TO PARAMETERISATION

The traditional and effective method of channel compensation for a single channel source has been to subtract the mean of the corresponding cepstral coefficients determined over the entire speech segment. The problem with this approach when the inclusion of multiple speech sources through different channels is the case, is that this approach would average the channel effects rather than remove them. Ignoring this effect may be somewhat damaging to recognition performance.

Given linear channels (and ignoring handset transducer effects), the sampled output signal, $Y(t)$, can be considered as the summation of the two speech signals $S_1(t)$ and $S_2(t)$, convolved with their corresponding channel impulse responses $H_1(t)$ and $H_2(t)$.

$$Y(t) = H_1(t) * S_1(t) + H_2(t) * S_2(t) \quad (1)$$

Diagrammatically, the recording configuration is indicated in Figure 1.

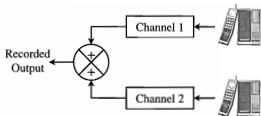


Figure 1. Configuration of the two speaker detection recordings.

In the one speaker detection configuration, the second speech signal source $S_2(t)$, and channel $H_2(t)$, are disregarded. Adding further signal source contributions can accommodate for the N-Speaker detection case.

To curtail the problem of multiple channel effects, there have been several methods proposed. These include subsegment length feature adjustment techniques such as general IIR (Infinite Impulse Response) RASTA processing (Hermansky and Morgan, 1994) and LDA-FIR (Linear Discriminate Analysis - Finite Impulse Response) Modulation Spectrum analysis (Van Vuuren, 1999). RASTA processing was introduced for speech enhancement purposes and improving speech recognition performance. This technique has since been applied to speaker verification, particularly for over-the-phone applications. This method has been found to have only a limited improvement over the standard segment length mean cepstral coefficient subtraction for one speaker speech segments. The other issue is the settling time of the IIR filter at the start of a speech segment. For short test speech segments (~3 seconds), such as some speech examples trialled in the NIST evaluation, the performance can significantly degrade in comparison with other channel estimate techniques. One of the issues with the IIR filter is how to initialise the output feedback component of the filter. An alternative to this approach is the use of an FIR filter. Here, the N coefficients (~300) of the filter are determined from an LDA of speakers examined across different conditions. This system performs comparably to the standard RASTA method. The drawback of this approach is the effort required for determining the filter using a data-driven approach on an external set of phonetically transcribed speech segments that are consistent with speech in the target application. Recent work (Van Vuuren, 1999) has determined the filter properties based on minimising the signal variation across handsets exclusive of the channel and not the handset deviations with varied telephone channels.

To avoid many of the inherent difficulties with that of the LDA-FIR and IIR RASTA techniques we propose another method to handle one and two speaker speech segments. This method lends itself to compensating for the presence of several speakers also. The rapid channel compensation technique in part, applies a box-car filter to the corresponding cepstral features running in time. Here, the output of the filter is subtracted from the corresponding raw cepstral features. The formal representation of this approach is indicated in z-

transform notation in equations (2) and (3), where $X(z)$ is the set of input feature observations and $Y(z)$ is the corresponding output.

$$Y(z) = X(z) - \frac{1}{2N+1} \sum_{i=N}^N X(z)z^{-i} \quad (2)$$

for a window length of $2N+1$.

$$Y(z) = X(z) - \frac{1}{2N} \sum_{i=N}^{N-1} X(z)z^{-i} \quad (3)$$

for a window length of $2N$.

Initially we selected a window size of (say) 300 speech frames at intervals of 10ms. Thus, once a score for the first 300 frames was accumulated to estimate the summation term, proceeding summation estimates could be determined quickly by a simple addition of the next feature coefficient and removal of the last frame of the window. A mean estimate is subtracted from the feature present in the middle of the current window.

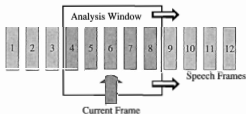


Figure 2. Diagram of the Filtering Approach for Channel Compensation.

The simplicity of this system also allows for the simple compensation of the features at the ends of the speech segment. This is useful for applications of limited length recording segments. To estimate the channel compensated features for the beginning of the speech segment, the nearest available windowing mean estimate is subtracted from the initial set of features. A similar approach is applied to determine the features at the end of the file.

This method indicated by Figure 2 and equations 2 and 3, is significantly faster to calculate than the FIR-LDA Filtering approach. In the LDA-FIR scheme, each compensated coefficient requires the weighted addition of the features spanning the window to be calculated. An alternative being the IIR RASTA method, has the difficulty of seeding the RASTA equation with an initial estimate of the output variable. Hence, a certain number of initial speech frames would have to be ignored to allow the filter to produce a stabilised estimate. As identified earlier, the box-car filter method is not limited to such an extent by this problem.

This style of compensation is suitable for varied channel sources over time such as for two and N speaker detection and speaker tracking. But for these instances, there remains the issue of selecting a suitable window length. A window size that is too short will not capture the channel specific

information, while a longer window length will increase the probability of having two speakers present within the window estimate period. Under this circumstance, the channel estimates of the two speaker signal source would become somewhat averaged. Thus, a suitable window length to balance these effects must be selected. The method of channel compensation proposed and the effects of window size on performance will be examined in our speaker verification system.

3. SPEAKER VERIFICATION SYSTEM OVERVIEW

Introduction

The general structure of the speaker verification module applied to the one speaker detection task is given in Figure 3. One of the differences with this system and the two speaker detection system is that there is no speaker score normalisation in the testing phase of the two speaker detection process. In addition, the distribution of the raw frame based log-likelihood ratio scores was analysed to determine the two speaker detection scores.

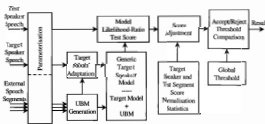


Figure 3. Block diagram of the Adapted-UBM One-Speaker Detection System.

Parameterisation

We used 24 parameters comprised of 12 MFCCs (using 20 filterbanks) with their corresponding delta coefficients. The speech frames were generated using 32ms of speech, offset at 10ms intervals. The signal was bandlimited from 300 to 3200 Hz. Channel compensation was applied to the baseline MFCCs before the delta coefficients were calculated. Silence removal was performed using an energy based histogram approach.

Target Speaker Modeling

We performed speaker modeling by use of the adapted Universal Background Model (UBM) method (Reynolds, 1997). This procedure adjusts the mixtures of a standard speech UBM model toward the distribution of the target speech. The model adaptation process requires the training of a high order GMM on a large quantity of speech. A GMM is a combination of $k = 1, 2, \dots, N$, single Gaussian components with dimensionality D , mixture weights p_k , means μ_k , and diagonal covariance matrices Σ_k . For a single speech feature vector observation, X , the probability density function for a speaker model λ , is described.

$$p(X | \lambda) = \sum_{k=1}^N p_k g(X; \mu_k, \Sigma_k) \quad (4)$$

with

$$g(X; \mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_k|}} \exp\left\{-\frac{1}{2}(X - \mu_k)^T \Sigma_k^{-1} (X - \mu_k)\right\} \quad (5)$$

For the verification system, there are two gender dependent models (male and female UBMs) using orthogonal mixture GMMs with 512 mixtures. Each UBM was trained on electret handset data from a large portion of the NIST 1999 Evaluation Target Speaker Set. After silence removal, only one in three parameterised frames were kept as training data. This was performed because adjacent frames are typically highly correlated, and keeping the extra data contributes little to the accuracy of the UBM but adds significantly to the training time. Target models were generated by adapting the corresponding gender-specific UBM to the target speaker using MAP adaptation. Both the UBM and the adapted model are stored for the testing phase.

In addition, validation speech was incorporated for performing Handset/Target Speaker Score Normalisation for each target speaker. The NIST 1999 data was partitioned such that validation speakers were not members of the speakers used to train the UBM. This speech data was trialled against the target models to derive the distributional statistics of the impostor speaker set for different handset types. This process called H-Norm, is performed for the carbon and electret handset types to improve performance across multiple handsets (Reynolds, 1997).

Testing Phase

Testing is performed for each frame of a test file, by finding the log-likelihood ratio (LLR) of a given target speaker model with its UBM (male or female depending on the target speaker). Given a speech feature vector X_t , a target speaker model λ_{TARGET} , and a UBM λ_{UBM} , the log-likelihood ratio may be determined.

$$A_t = \log p(X_t | \lambda_{\text{TARGET}}) - \log p(X_t | \lambda_{\text{UBM}}) \quad (6)$$

Only the top 5 scoring mixtures from the UBM were used for each frame, and the corresponding adapted 5 mixtures (McLaughlin et al, 1999) were used for all hypothesized target speaker tests. By taking advantage of the correspondence between the UBM mixtures and the adapted model mixtures, testing times can be dramatically improved.

The one speaker detection result was determined by averaging these LLR scores over the speech based segments and then performing H-Norm. The two speaker detection result was located by use of a bi-modal Gaussian mixture analysis of the log-likelihood-ratio scores and using the score of the highest scoring Gaussian mean (Myers, 2000). These systems have had proven performance in the NIST evaluations.

4. EVALUATION

Experiment Database

Of interest in this experiment is the performance of the fast channel compensation method and the effect of window size on the performance of one and two speaker detection. We aim to locate a suitable window size to suit both detection tasks. This experiment was examined according to the NIST 2000 speaker recognition specification (NIST, 2000). The database contained 457 male and 546 female target speakers, each with approximately two minutes of telephone speech. The one and two speaker detection tasks used these same target speakers to perform the test. Thus, by modeling each speaker in a universal fashion, the speaker models would not have to be retrained for each task.

One and Two Speaker Detection Results

Presented in Figures 4 and 5 are the one and two speaker detection results. Results are indicated in the form of a Detection Error Trade-off curve (DET). The better performing system has the lower Miss and False Alarm probabilities. For details concerning the DET representation see (Martin et al, 1997).

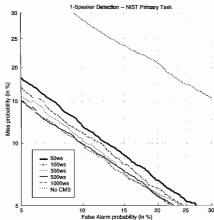


Figure 4. One Speaker Detection DET curve results.

The plot in Figure 4 indicates a generally improving trend of speaker recognition performance with increasing window length for channel compensation. As expected, the 1000ms (1000 frame window length) performed marginally better than the 500 frame compensation. This indicates that the longer the window length (to a certain limit) the better the channel/average vocal tract estimate. This demonstrates that whole utterance length cepstral mean subtraction is quite effective for one speaker detection. Figure 4 also contrasts the difference in performance between cepstral mean removed and the uncompensated speech features. It indicates that ignoring linear telephone network channel effects is detrimental to speaker verification performance.

For the two speaker detection task (Figure 5), an optimal performance was achieved for the 500 frame window length configuration and not the 1000 frame approach (as in the one speaker task). This shows that applying cepstral mean

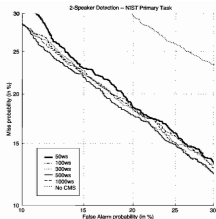


Figure 5. Two Speaker Detection DET curve results.

subtraction over long periods (or whole utterances) with multiple speakers and channels present will degrade multi-speaker detection performance.

5. CONCLUSIONS

It was determined that the running mean box-car filter cepstral removal approach for channel compensation was a successful approach. The optimal window length for both the one and two speaker detection tasks was 500 frames. This particular method of channel compensation is orders of magnitude faster to execute than FIR RASTA alternatives and more stable at the beginning of speech files than IIR based RASTA filter approaches. This method can also be adapted for a fast real-time implementation of speaker recognition applications.

ACKNOWLEDGEMENTS

This work was supported by a research contract from the Australian Defence Science and Technology Organisation.

REFERENCES

- Hermansky H and Morgan N. (1994) "RASTA Processing of Speech," *IEEE Trans. Speech and Audio Processing*, 2(4), 587-589.
- Liu H and Mammon R. (1995) "A Subword Neural Tree Network Approach to Text-Dependent Speaker Verification," *Proc. ICASSP*, pp 357-360.
- Martin A, Doddington G, Kamm T, Ordowski M and Przybocki M. (1997) "The DET Curve in Assessment of Detection Task Performance," *Proc. Eurospeech*, 4, 1895-1898.
- McLaughlin J, Reynolds D and Gleason T. (1999) "A Study of Computation Speed-Ups of the GMM-UBM Speaker Recognition System," *Proc. Eurospeech*, 3, 1215-1218.
- Myers S, Pelcasos J and Sridharan S. (2000) "Two Speaker Detection by Dual Gaussian Mixture Modelling," *Proc. Australian Internat. Conf. Speech Sci. and Tech.*, pp 300-305.
- NIST. (2000) *NIST's speech and Speaker Recognition Website*, <http://www.nist.gov/speech/>
- Reynolds D. (1997) "Comparison of Background Normalization Methods for Text-Independent Speaker Verification," *Proc. Eurospeech*, 2, 963-966.
- Van Veen S. (1999) *Speaker Verification in a Time-Feature Space*, PhD Thesis.

ADAPTIVE DYNAMIC RANGE OPTIMISATION FOR HEARING AIDS

Lois F.A. Martin^{1,2}, Peter J. Blamey³, Christopher J. James³, Karyn L. Galvin^{1,2}, & David Macfarlane^{1,2}

¹Bionic Ear Institute

²CRC for Cochlear Implant and Hearing Aid Innovation

³Department of Otolaryngology, University of Melbourne

ABSTRACT. ADRO (Adaptive Dynamic Range Optimisation) is a slowly-adapting digital signal processor that controls the output levels of a set of narrow frequency bands so that the levels fall within a specified dynamic range. ADRO is suitable for a variety of applications, including control of a hearing aid. In the case of a hearing aid, the output dynamic range is defined by the threshold of hearing (T) and a comfortable level (C) at each frequency for the individual listener. A set of rules is used to control the output levels, with each rule directly addressing a requirement for a functional hearing aid. For example, the audibility rule specifies that the output level should be greater than a fixed level between T and C at least 70% of the time. The discomfort rule specifies that the output level should be below C at least 90% of the time. In this study, open-set sentence perception scores for 15 listeners were compared for ADRO and a linear hearing aid fit. Speech was presented at three levels. ADRO improved scores by 1.9% at 75 dB SPL (NS), 15.9% at 65 dB SPL ($p = 0.014$) and 36% at 55 dB SPL ($p < 0.001$).

1. INTRODUCTION

The main problem resulting from hearing impairment in adults is poor audibility of sounds at normal intensities. This problem may be overcome to some extent by amplifying sounds with a hearing aid, however amplification can introduce further problems. The loudness of sounds often grows faster than normal in hearing-impaired ears (recruitment), so that loud sounds may become uncomfortable after they are amplified by the hearing aid. Often, hearing thresholds and maximum comfortable levels vary with frequency so that the gain of the hearing aid needs to change as a function of both frequency and intensity of the input signal to provide an output signal that is both audible and comfortable. Linear hearing aids attempt to meet the audibility criterion with a fixed gain and frequency response such that speech signals at a normal intensity are placed near the middle of the listener's range of hearing. The National Acoustics Laboratory (NAL) prescription is a widely used example (Byrne & Dillon, 1986; Byrne, Parkinson & Newall, 1990). Linear hearing aids usually incorporate a maximum power output limiter to meet the comfort criterion. Clearly, a linear aid with limiting can only provide a good approximation to the required gain as a function of frequency for a fairly narrow range of input levels. The NAL non-linear prescription (NAL-NL1, 1999) provides a more detailed description of the required gain function, together with recommendations for implementations using single- and multi-band compression hearing aids. Most alternative hearing aid prescriptions take a similar form: i.e. they specify the required gain as a function of the input frequency and intensity parameters (Skinner, 1988).

ADRO is designed to take a more direct approach by specifying target output levels as a function of frequency in such a way that the audibility and comfort criteria are met automatically. The gain of the hearing aid is adapted in order

to keep the output signal level within the optimum dynamic range. These target output levels are related to measured threshold and comfortable levels in a more straight-forward manner than the gain parameters specified in other types of processors.

The aim of this study was to validate the ADRO processing and fitting procedure for a range of hearing-impaired listeners. The hypothesis was that ADRO would produce higher speech perception scores than a standard fixed-gain hearing aid, especially at moderately low presentation levels where the additional gain provided by ADRO should improve audibility.

2. METHOD

Statistical description of the output signal

The first requirement for ADRO is to measure the distribution of output levels as a function of frequency and time. To achieve this goal, the ADRO processor uses a 128 point Discrete Fourier Transform (DFT) to split the sampled input signal into 64 frequency bins, F_c . A Hanning window is applied prior to the DFT. The complex input amplitude, I_c , of each frequency component is multiplied by a scalar gain factor, G_c , to obtain the output amplitude, O_c . ADRO uses estimates of the distribution of output levels in the form of percentiles. For example, the 90th percentile is the level which is exceeded 10% of the time, and the 50th percentile is the level that is exceeded 50% of the time. These percentiles are estimated by comparing the magnitude of the output amplitude, $|O_c|$, with the current value of the percentile estimate. If the magnitude is greater, the estimate is increased by a small amount, U dB. If the magnitude is smaller than the estimate, then the estimate is reduced by a small amount, D dB. If U and D are equal, then the estimate will tend to the 50th percentile because the number of upward steps will then be equal to the number of downward steps. Other percentiles may be estimated by changing the relative size of U and D . The percentile value is given by $100 U/(U+D)$. For example,

If U is 9 times larger than D , one upward step will be balanced by 9 downward steps, and the estimator will tend to the 90th percentile where the probability of a downward step is 9 times greater than the probability of an upward step. The rate at which the percentile estimates change is controlled by the absolute size of U and D , and the frequency with which the FFT windows are updated. Typically, the slew rate for the estimates in ADRO is about 20 dB per second.

The ADRO targets for a hearing aid user

The second requirement for ADRO is to measure, or define, the required output dynamic range for each of the frequency bands. For a hearing aid user, the limits of the useable dynamic range are the threshold of hearing and the maximum comfortable level, MCL, for each frequency. These parameters are measured using 1/3 octave bands of noise covering the frequency range of interest. These signals are generated by the ADRO processor itself, controlled by a PC program called AUDY. The target output levels for ADRO are derived from threshold and loudness estimates. Thresholds, T_n , are measured using a conventional adaptive detection procedure. Following the threshold measures, a 7-point loudness scale (Hawkins et al, 1987) is used to establish the dynamic range. The 7 categories are: very soft, soft, comfortable but slightly soft, comfortable, comfortable but slightly loud, loud but OK, uncomfortable loud. ADRO uses three target levels at each frequency: M_n , C_n , and A_n , which represent the maximum output level, a comfortable level, and a minimum audibility level at each frequency. The "loud but OK" level is used for M_n , the "comfortable" level is used for C_n , and the A_n level is either Ci-20 dB or T_n , whichever is greater.

The ADRO rules

ADRO uses a set of rules that are applied independently at each frequency: The comfort rule requires the 90th percentile to be below the C_n target level for every frequency. If the comfort rule is violated, the gain, G_n , at that frequency is reduced by a small amount. The audibility rule requires the 70th percentile to be above the A_n target for every frequency. The audibility rule is checked only if the comfort rule is satisfied. If the audibility rule is violated, the gain, G_n , is increased by a small amount. The sizes of the increments and decrements of gain are chosen so that the maximum rate of decrease is about 9 dB per second, and the maximum rate of increase is about 3 dB per second. In a more conventional automatic gain control, these parameters would be equivalent to very long attack and release times. The maximum gain rule requires the gain to be less than a fixed amount G_{max} . This rule limits the loudness of background noise and avoids feedback in quiet situations where the gain might otherwise become very high. A typical value of G_{max} , for profoundly deaf listeners would be about 60 dB. Finally, the maximum output rule requires the magnitude of the output level, $|O_n|$, to be less than M_n . If this rule is violated, the magnitude of O_n is reduced to be equal to M_n , leaving its phase unchanged. The final stages of processing are to apply an inverse DFT to the amplified output levels O_n , multiply the result with a Hanning window, and use the overlap/add method to generate an output signal. A final multiplier is incorporated to give the listener

control of the overall loudness. Usually, this volume control is set to attenuate the signal. This attenuation is most likely required to compensate for intensity and loudness summation across the frequency bands, each of which is capable of reaching a loud level on its own. In other words, the thresholds and comfortable levels of broad-band signals like speech are lower than the corresponding values for narrow-band noise.

The ADRO hearing aid

The ADRO hearing aid used in this study was a benchtop processor based on a Motorola DSP 56303 digital signal processor evaluation board, fitted with a microphone, preamplifier, output amplifier and an Oticon AN180, AN270, or AP1000 hearing aid receiver which was attached to an individually fitted hearing aid mould for each listener. The hearing aid receiver model was chosen according to the output power required for each listener. The sampling rate of the analog to digital converter was 9.6 kHz, giving a window length of 13.3 ms. Overlapping windows of data were analysed every 3.3 ms. The Motorola processor was interfaced to a personal computer running Windows 95 via a serial port so that the AUDY program could control the stimulus generation and parameter selection during the fitting procedure. The AUDY program could also be used to display a snapshot of the percentile estimates, output levels, and gains at about 1 second intervals. This display was useful to verify and explain the operation of the ADRO rules. It was also possible to implement a fixed-gain hearing aid by disabling the adaptive operation of the ADRO rules and setting the G_n values according to the NAL-RP prescription (Byrne et al, 1990). An upper limit to the $|O_n|$ values was implemented in a frequency-specific manner as for ADRO.

Participants and procedures

Fifteen adults with moderate to profound hearing loss (44 to 98 dB HL pure-tone-average hearing loss) took part in this study. All but two of the participants normally used a hearing aid. Results for the two participants who did not normally use hearing aids may be identified in Figures 1 to 3 by their hearing losses of 60 and 78 dB HL. The audiogram for each participant was measured using standard audiological procedures and equipment, and a NAL-RP prescription hearing aid was programmed. The ADRO hearing aid targets were determined for each individual using the loudness estimation procedure described above. The speech perception of each participant was tested with each of the 2 procedures at free-field intensity levels of 55, 65, and 75 dB SPL using open-set CUNY sentences (Boothroyd, Hanin, & Hnath, 1985). The CUNY sentences were recorded onto CD by a female Australian speaker and the RMS levels of individual sentences were equalised digitally prior to presentation. The list numbers used for different conditions were randomised and no participant was tested more than once with the same list. Each sentence list contains 102 words, and was scored according to the percentage of words correctly repeated. Prior to the speech perception testing, the volume setting of the NAL hearing aid was adjusted to match the loudness of speech at a normal conversational level for both aids.

3. RESULTS

The scores for the individual participants at the three presentation levels are shown in Figures 1 to 3. Three of the participants had insufficient time available to complete the testing and scores were not obtained at 75 dB SPL for two participants and at 55 dB SPL for 1 participant. A two-way ANOVA indicated that presentation level, hearing aid, and the interaction term were all significant with $p < 0.001$. F values were 111.02, 42.06, and 12.58 respectively. The mean scores for NAL and ADRO hearing aids at 75 dB SPL were 83.6% and 81.7%, respectively. The difference of 1.9% was not statistically significant (post hoc Tukey t-test, $t = 0.38$, $p = 0.99$). At 65 dB SPL, mean scores were 79.6% for ADRO and 63.7% for NAL. The mean difference of 15.9% was significant ($t = 3.42$, $p = 0.014$). At 55 dB SPL, the mean scores were 55.0% and 18.6% and the difference of 36.4% was highly significant ($t = 7.56$, $p < 0.001$).

The mean scores at different presentation levels were also compared using post hoc Tukey t-tests. The mean scores at 65 and 75 dB SPL for ADRO were not significantly different (difference = 6.1%, $t = 1.28$, $p = 0.80$), but the scores at 55 dB SPL were significantly lower than at 65 dB SPL (difference = 27.1%, $t = 5.71$, $p < 0.001$). For the NAL prescription, the mean score at 65 dB SPL was lower than at 75 dB SPL (difference = 20.2%, $t = 4.15$, $p = 0.001$), and the mean score at 55 dB SPL was lower than at 65 dB SPL (difference = 47.6%, $t = 10.02$, $p < 0.001$). These results indicate that ADRO maintains maximum intelligibility at lower intensities than the NAL prescription hearing aid.

4. DISCUSSION

The hypothesis was supported by the group results at 55 and 65 dB SPL. At 75 dB SPL, ADRO was no worse than the NAL prescription. These intensity levels correspond to speech at levels described as "casual", "raised", and "loud" (Keidser, 1995; Pearsons et al, 1977). It should be noted that the recordings were made at a "normal" level (60 dB SPL) and then adjusted to the presentation levels, rather than recording "casual", "raised" and "loud" speech which would have resulted in different spectral shapes for the three conditions. The results indicate that ADRO should provide a significant advantage in most common situations at normal conversational levels. At 75 dB SPL, most participants showed little difference between the NAL and ADRO scores which were both quite high (over 80%). These differences may have been restricted by a ceiling effect. The largest difference at 75 dB SPL occurred for a subject with a severe hearing loss (PTA = 82 dB HL) where there was no ceiling effect because both scores were lower. At 65 and 55 dB SPL, every participant scored at least a little higher with ADRO than with NAL. At 65 and 75 dB, the largest improvements were for participants with severe hearing losses over 75 dB HL. At 55 dB SPL, this trend was reversed and the largest improvements were for participants with moderate and severe hearing losses less than 75 dB HL.

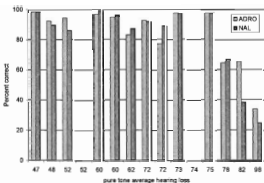


Figure 1. Comparison of open-set sentence perception scores at 75 dB SPL for 13 subjects using ADRO and a NAL linear hearing aid fit. Scores for individual subjects are ordered by increasing hearing loss in dB HL. The mean difference was 1.9%.

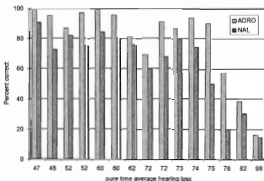


Figure 2. Comparison of open-set sentence perception scores at 65 dB SPL for 15 subjects using ADRO and a NAL linear hearing aid fit. Scores for individual subjects are ordered by increasing hearing loss in dB HL. The mean difference was 15.9%.

Further research is needed to evaluate the ADRO hearing aid with different materials and under different conditions. It remains to demonstrate that ADRO can protect listeners from the discomfort of loud sounds, both speech and environmental. An evaluation with more difficult materials may possibly indicate an advantage at 75 dB SPL if the present results are indeed limited by a ceiling effect. The present study does not indicate the relative performance of ADRO and NAL hearing aids in background noise. Background noise is a major problem for hearing aid users, and it is possible that ADRO may exacerbate this problem by amplifying the noise to louder levels. These additional issues have been addressed for the application of ADRO to the commercially available SPRINT cochlear implant speech processor (Blamey, James, & Martin, 1999). With SPRINT, a significant advantage was found for ADRO at high input levels for monosyllabic words but not for CUNY sentences, an advantage was found for

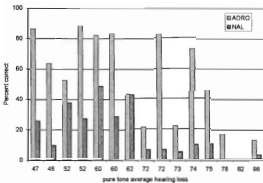


Figure 3. Comparison of open-set sentence perception scores at 55 dB SPL for 14 subjects using ADRO and a NAL linear hearing aid fit. Scores for individual subjects are ordered by increasing hearing loss in dB HL. The mean difference was 36.4%.

ADRO at moderate and low input levels for both monosyllabic words and CUNY sentences, and ADRO performed no worse than the standard processor in background noise. A questionnaire established that implant users preferred the ADRO processor over the standard processor in 53% of common situations, compared with no preference in 32% and a preference for the standard processor in 10% of situations. Indiscriminate generalisations should not be made from cochlear implants to hearing aids, but it seems probable that the results will also be good when ADRO has been implemented in a wearable hearing aid.

There are commercially available hearing aids that include various forms of automatic gain control or compression (reviewed by Dillon, 1996) that may perform as well as ADRO in quiet at different presentation levels. Further studies will be conducted to compare ADRO with a commercially available compression aid.

5. CONCLUSIONS

The concept of adaptive dynamic range optimisation (ADRO) potentially provides a straight-forward solution for some of the most pervasive problems faced by people with impaired hearing. This study demonstrated clearly that ADRO provides a good solution to the problem of poor audibility of speech over a broader range of input levels than a conventional fixed-gain hearing aid. It remains to be shown that the ADRO hearing aid provides a solution to the other two major problems of discomfort in loud noises and poor intelligibility of speech in background noise.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, the Bionic Ear Institute, and the National Health and Medical Research Council of Australia. The research was conducted under the auspices of the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (project 99/379H). The authors also

gratefully acknowledge the valuable contribution of the participants who volunteered their time for these experiments.

REFERENCES

- Blamey, P.J., James, C.J. & Martin, L.F.A. (1999) "Adaptive dynamic range optimisation: a pre-processing algorithm for cochlear implants." *European Custom Sound Outcomes Seminar, Sintra, Portugal, June 1999.*
- Boothroyd, A., Hain, I. & Heath, T. (1985) "A sentence test of speech perception: Reliability, set equivalence, and short term learning." *Speech & Hearing Science Report RC10* (City University New York).
- Byrne, D. & Dillon, H. (1986) "The National Acoustics Laboratory (NAL) new procedure for selecting the gain and frequency response of a hearing aid." *Ear & Hearing* 7, 257-265.
- Byrne, D., Parkinson, A. & Newall, P. (1990) "Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired." *Ear & Hearing* 11, 40-47.
- Dillon H. (1996) "Compression? Yes, But for low or high frequencies, for low or high intensities, and with what response times?" *Ear & Hearing* 17, 287-307
- Hawkins, D.M., Walden, B.E., Montgomery, A.A. & FURUKAWA, R.A. (1987) "Description and validation of an LDL procedure designed to select SSPL90." *Ear & Hearing* 8, 162-169.
- Keijser, G. (1995) "Long term spectra of a range of real-life noisy environments." *Aust. J. Audiol.* 17(1), 39-46.
- NAL-NLI (1999) *NAL Non-Linear User Manual, Version 1.1*, (National Acoustics Laboratories: Chatswood, NSW).
- Pearsons, K.S., Bennett, R.L. & Fidell, S. (1977) *Speech levels in various noise environments*, EPA report No 600/1-77-025. (Environmental Protection Agency: Washington DC).
- Skinner, M.W. (1988) *Hearing aid evaluation*, (Prentice Hall: Englewood Cliffs, NJ).



"Acoustics Workshop"

presented by
National Voice Centre,
The University of Sydney

Tuesday October 9, 2001 from 9am - 5 pm

Keynote speakers: Professors Johan Sundberg & Neville Fletcher
with: A/Professor Jennifer Oates, A/Professor Pamela Davis, Denzil
Cabrera, Jennifer Barnes, Debbie Phyland and others

A workshop for all interested in acoustics, voice or singing, speech pathology, linguistics and speech science. The course will be a scientific overview and practical demonstration of up-to-date methods for acoustic analysis of the speaking and singing voice. Issues in the acoustics of the human voice including singer's formant and related physiology and measurement of wind instrument performance will also be presented. Factors in voice recording including the effects of different recording environments and microphone placement will also be presented.

The format will be lectures and demonstrations in the morning and "hands-on" laboratory analysis in the afternoon or material supplied by the presenters and participants will work in small groups under the guidance of the presenters. Registrants should have had some experience in using acoustic analysis software.

Full Registration Fee	\$220
40% off early bird	\$170
U/T Students with ID	\$140
Optional dinner on 9/10/01 from 7.30 pm	\$ 40

Strictly limited to the first 40 paid registrations

Course notes and a Certificate of Attendance will be provided
National Voice Centre, The University of Sydney, NSW 2006

Payment by credit card, cheque or invoice

Credit card payment by fax: +61 2 9351 5351 or email to: voice@voca.usyd.edu.au
(early bird date is September 1 and closing date September 21)

The course will follow a 3.5 day workshop on "The Science of Voice and Singing", with Professor Johan Sundberg & Janice Chapman.
For more information call the Centre. Phone: +61 2 9351 5352

PROSPECTS FOR SPEECH TECHNOLOGY IN THE OCEANIA REGION

J. Bruce Millar

Computer Sciences Laboratory
Research School of Information Sciences and Engineering
Australian National University

ABSTRACT: The development of speech technology in the Oceania region is an issue for Australian speech scientists and technologists. In this paper we examine both the issues that govern the development of speech technology anywhere, the specific opportunities and inhibiting factors of the Oceania region, and the role that Australia, as the largest and most prosperous nation of the region, can have in the process. The necessary scientific resources required to establish both basic and more sophisticated speech technology are reviewed and mapped against the characteristics of the Oceania region. It is concluded that the most productive approach is likely to be one of creative partnership with the many island communities such that technology may be developed in a cost-effective and culturally sensitive manner.

1. INTRODUCTION

The development of new technology throughout the world is very uneven. Grand aspirations of universal benefit are typically voiced at its launch, then, as time progresses, the influence of other factors, such as individual greed within capitalism and individual lethargy within socialism, tend to erode the hope of benefit for those who are remote from the centre of power. By any reasonable measure a very large proportion of the residents of Oceania are geographically, economically and even linguistically remote from the power centres that drive speech technology. They suffer what we may call a three dimensional deficit of benefit.

Speech technology is developing at a rapid rate in highly developed regions of the world such as North America and Europe. It is providing access to information held by an increasingly wide range of public utilities and commercial companies via the nearly ubiquitous telephone handset. In this paper we will examine the potential for long-term deficit of such benefits that exists for the Oceania region and some measures that may be taken by the speech science and technology community to reduce the impact of such deficit on the residents of this region.

A full picture of the factors influencing the development of speech technology will include the status of current language resources in the region and the prospects for their further development and their application to speech technology. These prospects will depend on relevant linguistic characteristics of the region, the level of technological development in the region with specific focus on the development of telecommunications, and critically the relevance of access to available information to the communities of the region.

2. OVERVIEW OF THE REGION

The region is dominated in terms of land area, population, and economic prosperity by Australia (see Table 1). Many aspects of the region are characterised by the progressive move to independence from 19th century colonial powers. This pattern has resulted in a legacy of official languages made up of 81% English, 17% English-based pidgin, 1.5% French, and 0.6%

English/French. Populations range from 18.7 million in Australia to just 49 in the Pitcairn Islands. The economies range from the "western" economy of Australia with a gross domestic product per capita of 21,200 US dollars to the subsistence economies of Kiribati and Tuvalu whose gross domestic product per capita is just 800 US dollars. These data are derived from the USA Central Intelligence Agency (Central Intelligence Agency, 1999).

3. LINGUISTICS

The "official" languages of the Oceanic nations are strongly influenced by their colonial past which was dominated by Britain with smaller contributions from the USA and France. The official language will normally be the language of education and of formal administration. Its spoken form will be accented by the phonologies of the native languages of the speakers.

The "native" languages present a much more complex pattern arising from a history of tribal isolation and subsequent migration activity. Oceania has one of the lowest population to language ratios in the world. This makes the development of language resources linguistically complex and economically difficult. While there are certain anomalies, the Melanesian region extending off the north-eastern coast of Australia northerly to Papua New Guinea and easterly to Fiji, is characterised by many languages per island group (Pawley, 1995). In this region the number of speakers of a language can be of the order of 1000. In contrast, the regions of Micronesia, a northerly extension of Melanesia, and Polynesia, an easterly and southerly extension of Melanesia, are characterised by a single language per island group. Thus the indigenous languages of Polynesia and Micronesia are characterised by both a larger territorial range and population size than the those of Melanesia (Pawley, 1995). Throughout the region there are some 30 million people speaking approximately 1250 native languages.

The "contact" languages which represent the language of trade or other activities of forced contact between disparate language speakers often span much larger populations than "native" languages, and can rise in status to "national"

COUNTRY	Area sq.km	Pop'n 000s	Languages Spoken	GDP US\$	Economy
Pitcairn Islands	47	0	English + Pitcairnesse + Tahitian	n/a	Subsistence+Fishing
Tokelau	10	1	English + Tokelauan	1000	Copra+Crafts
Nue	280	2	English + Nuean	1200	Subsistence
Nauru	21	11	English + Nauruan	10000	Phosphate(exhausted)
Tuvalu	26	11	English + Tuvaluan	800	Subsistence
Wallis and Futuna	274	15	French + Wallisian	2000	Subsistence
Palau	458	18	English + Palauan + 4 others	8800	Subsistence+Fishing
Cook Islands	240	20	English + 4 other languages	4000	Supported by NZ
American Samoa	199	84	English + Samoan	2600	Fishing
Marshall Islands	181	65	English + Japanese + 2 others	1450	Copra+Fish+Crafts
Northern Marianas	477	69	English + Chamorro + Carolinian	9300	Tourist+Textiles
Kiribati	717	85	English + Gilbertese	800	Copra+Fish
Tonga	748	109	English + Tongan + 2 others	2100	Agric+Tourism
Micronesia	702	131	English + 4 other languages	1750	Farming & Fishing
Guam	541	152	English + Chamorro + Japanese	19,000	Military+Tourism
Vanuatu	14,760	189	English + French + 109 others	1300	Fishing,Finance,Tourism
New Caledonia	19,060	197	French + 38 other languages	11400	Nickel+Tourism
French Polynesia	4,167	242	French + Tahitian	10,800	Military+Tourism
Western Samoa	2,890	299	English + Samoan	2100	Agriculture
Solomon Islands	28,450	455	English + Pijin + 120 other languages	2600	Ag+Fish+Forestry
Fiji	18,270	812	English + Fijian + Hindi + 7 others	6700	Sugar+Tourism
New Zealand	282,680	3662	English + Maori	17000	Near Western
Papua New Guinea	462,840	4705	English + Tok Pisin + 714 others	2400	Agr+Minerals
Australia	7,686,850	18783	English + 234 other languages	21200	Western

Table 1: Areas, populations, languages and economies of the countries of the Oceania region

languages as with Tok Pisin in Papua New Guinea. Some of these languages have developed from being second languages used mainly for trading purposes (and living alongside native languages are labelled Pidgin languages) to being first languages for at least some of their speakers (and hence labelled Creole languages). Within the region there are six contact languages having speaker populations ranging between fifteen thousand and two million.

4. ECONOMICS AND CULTURE

When we view the geographical and linguistic characteristics of the region, the challenges for the development of speech technology are evident. One of the biggest of these challenges is to introduce this development in a culturally sensitive and economically viable manner. It is clear that speech technology developers, who already have a foothold in Australia, are looking to the bigger markets of Asia and beyond rather than to their Oceanic neighbours whose market is considered not viable. While this is a sound decision when viewed from the perspective of the shareholders in such speech technology companies, it suggests that a strong alternative view needs to be developed by those who are entrusted with public funding.

When considering the uptake of speech technology in a region it is imperative to examine the physical means, the financial cost and the motivation of the population of the region. Telecommunication infrastructure in the Oceanic region has developed slowly through four major network technologies. By the mid-1980s, the introduction of satellite telecommunication links of increasing sophistication and coverage, saw all nations with populations exceeding 30,000 hav-

ing some form of link (Ward, 1995). The 1990s have seen the advent of very high bandwidth fibre-optic cable but with typically longer runs which bypass many nations.

The incidence and rate of growth of fixed telephone lines in the region for the mid-1990s is found in data from the Telecommunication Development Bureau of the ITU (International Telecommunication Union, 1996). The incidence of telephone lines per 100 people divides the region into those countries with around 50 lines, those with around 20 lines, and those with 5 lines or less. The first group are those with essentially western economies, the second group are those where telephone use is strongly dominated by expatriate residents of colonial powers, and the third group represent a clearer picture of the status in independent developing nations of the region. It is in this third group that

the challenge exists.

Telecommunication costs in the region have been analysed by Ward (1995). He indicates some of the factors that influence the very varied and sometimes asymmetric cost structures encountered. These structures are governed by the density of telecommunication traffic (influenced by factors of colonial past), official language affiliation, major trading partners, and local trade and travel. Ward concludes pessimistically that the island states of Oceania are likely to be excluded from information economy based on the cost of connection. It seems likely however that the cost of telecommunications for speech technology could be minimised by the location of call-centres which then link to information sources via the internet. It is also clear that the forms of voice data transmission are developing, with voice over internet protocols likely to rival existing voice circuits and with more sophisticated ISDN connections. It seems likely that by concentration of "traffic" attracted by a relevant service, that the impact of telecommunication costs may be reduced. This more optimistic view is encouraged by reports from the region that indicate a very rapid uptake of internet access at main centres and that this is expected to spread to more remote islands and atolls where reliable telephone access exists (Early, 2000).

The question of what kinds of information would be most relevant for telephone users in the region does not as yet have a definitive answer, however scanning the major components of the national economies in table 1 will give an initial impression of application areas. It seems clear that information related to import and export of commodities, finance, and tourism is likely to feature strongly.

The "choice" of interface to such information, a graphical web-browser or voice, is worthy of consideration against the cultural and economic backdrop of the region. A web browser presupposes an investment in a personal computer and the skill to use it. While the use of the telephone may still be avoided on cultural grounds in some parts of the region, its acceptance must exceed that of a computer, and its relatively low cost and natural spoken language interface provides less impeding to its use.

5. INITIATIVES, OPPORTUNITIES AND INHIBITORS

We can now examine more closely the issues facing speech technology based on our review of the region. There are relatively few speech recognition or speaker verification systems in common use in the region. However there is strong speech-technology commercial activity in Australia and New Zealand at present. I am unaware of commercial installations of text-to-speech synthesis systems that are optimised to the local phonology despite a large research effort in Australia over many years. The major focus of speech research in Australia has been scientific and its has been driven to a large degree by academic rather than application goals. This has caused a focus on processes required to faithfully extract reliable information from speech data rather than on the data itself, its variability, its contamination, and where the limits of our knowledge demand that we adequately manage our ignorance (Millar, 1997).

The limiting factor in the development of speech recognition systems in Oceania is the need to acquire adequate data resources of local language patterns. Initial interest in the development of these resources has arisen in academic circles (e.g. Millar et al, 1990, 1994, 1997) where this development has been slow owing to labour intensive collection and annotation techniques. More recently a rapid expansion of speech data resources for telephone speech has occurred using the specific opportunities of commercial enterprises. One technique used has been to implement a rather simple speech recognition task using data models from a distant English speaking community, but to structure the task so that the perplexity of each step is limited in order to allow sufficient recognition accuracy to be obtained (Forsyth, 2000). If each recognised spoken token is also verified by a simple "yes/no" response, then the system can be used to collect reliable local data from which to develop local speech data resources. As these resources are expanded they can be used to build more accurate phone models which can support adequate accuracy with more difficult tasks. In this way the telephone speech applications can grow without the very heavy overheads of initial data model development. An alternative approach is to conduct an explicit data collection exercise and then integrate the data acquired with a distant model. Both kinds of process do, of course, produce data resources that are jealously guarded by their commercial developers as they have been gained by lengthy processes. Commercial enterprises do not wish to give competitors the opportunity to build sophisticated applications without the time delay inherent in their development.

The overall linguistic picture is complex and indicates that viable populations that speak the same language with a consistent phonology will be very few outside of Australia and New Zealand. However the region does have larger populations that speak accented forms of either English or French. One opportunity is for telephone speech technology to enter any population at a level of low perplexity recognition using external English or French models. This is initially tedious but has worked for applications in Australia. This development from low perplexity recognition is yet to be demonstrated in New Zealand. The penetration of such systems will clearly depend on the degree of accenting relative to the base models on which the externally developed systems are based. Studies that examine the "data increment" required to seed effective transformation of recognition performance, having a defined perplexity, from one accented form of English to another would shed light on the likelihood of success for this approach. Inhibitors to this approach may arise from a telecommunications cost perspective, and from any lack of relevance of available computer-based information.

Another initiative that could be examined is the analysis of the likelihood of telephone-based information access based on one of the "contact" languages such as "Tok Pisin" which has two million speakers. It seems clear that Tok Pisin has the semantic capacity to readily handle spoken information queries. Studies to examine the variance in the realisation of Tok Pisin could help to assess the likelihood of success. It is clearly within the bounds of cross-speaker intelligibility but may strain the viability of current speech recognition techniques. It is significant for this approach that Tok Pisin has been systematised into a grammar and dictionary so the necessary information on which to build components in addition to phoneme models required for speech recognition systems are available (Mihalic, 1971).

The major inhibitors are perhaps threefold: the cost of development of systems relative to the economies of the beneficiaries; the availability of relevant information at a "telecost distance" (Ward, 1995) that is viable for the use of such systems; and cultural distance from the high-tech western world.

6. CURRENT DEVELOPMENTS AND ASPIRATIONS

Developments are currently focused in Australia and New Zealand but aspirations cover most of the region. There is a strong awareness amongst many leaders in the region that they must access the information highway; of the world but the opportunities for the use of spoken language are not widely appreciated.

Overcoming the barriers of remote communities

The Community Teleservice Centre (CTSC) concept was developed for rural areas of Scandinavia in the 1980s and the derivative "Telecottage" concept in the UK and Ireland in the early 1990s, and in Australia during the 1990s (Qvortrup, 1993; 1994). Qvortrup points to three barriers to the effective introduction of information services to remote areas: The service barrier due to services related to urban rather than

remote rural needs; the cost barrier due to low usage with respect to capital investment, and the qualification barrier due to the lack of skills in navigating through complex systems towards valuable target information. The CTSC concept can lower the cost barrier by concentrating traffic but it needs cultural awareness and appropriate technology to overcome the service and qualification barriers. Speech technology could overcome both of these if appropriate language resources are available.

Language Resources

Speech Technology is built on a foundation of language resources. These resources for speaker communities in developed countries have become the focus of intense effort in recent years. They comprise several facets, notably, a pronouncing dictionary in which the orthographic form of the language is linked to its phonemic form, a language model in which the way that dictionary entries (words) are typically strung together is defined in stochastic terms, and a set of phone models in which the range of acoustic realisations of each phoneme in a substantial range of contexts are also represented statistically. The development of these resources for a new speaker community can be prohibitively expensive unless they can be adapted from existing models for related speaker communities and/or be developed using a partnership between low-cost labour and high-tech facilities.

Developing Partnerships

It is very clear that commercial interests in speech technology cannot see any satisfactory return on investment in the region (beyond Australia and New Zealand). There is however a history of governmental largesse from these two countries to the smaller and less developed parts of the region. Strong economic development, political stability, and distinctive cultural attractions in the region are all important to the largely western economies of Australasia.

Partnerships that can build personal and institutional linkages, create economic and cultural benefits in the region, and exhibit cost-effectiveness in bringing speech technology to the region do appear to be feasible. The essential partnership is between language technology developed in high-cost western economies and language knowledge held in low-cost regional economies. It appears that many of the very labour-intensive parts of speech technology development could be performed by suitably educated members of target language communities. Tools for the building of pronouncing dictionaries, capturing adequate volumes of text for language modelling, and adequate recordings of speech for phone modelling can be provided at little added cost by the international speech technology community. The efficient use of these tools in a culturally sensitive manner can be most effectively managed by linguistically trained nationals.

It should also be noted that a major driver for such partnership is also the cultural survival of the region. Language is at the heart of culture and one impact of such partnerships will be to elevate the self-image of communities through the recognition that their own language can be a bridge to the high technology that is driving the information age.

A mechanism for regional partnerships

In 1991 the Coordinating Committee for Speech Databases and Assessment (COCOSDA) was established to unite speech scientists and speech technologists in creating best practice for creating resources for the development and evaluation of speech technologies. In 2000 this international consortium has been restructured to focus on key spoken language resource technologies across six major world regions, of which Oceania is one. The immediate task is to develop a network of experts in speech technology, in the linguistics of the region, and in the cultures of the region to examine the prospects for regional partnership.

7. CONCLUSIONS

Several challenges for the introduction of speech technology in countries of Oceania have been presented. As access to information is an essential factor for human development, there is impetus for assessing these challenges and then recommending the implementation of the appropriate technology to deliver that information. The modern portable multimedia personal computer is well adapted to provide a personal information interface. The ubiquitous presence of such interfaces and presentation software that is adaptable to all levels of human-computer familiarity would present an ideal human-information network. As this ideal is currently unattainable on economic grounds we must look at the role of the simpler interface offered by the telephone. Its implementation costs are low by comparison and the demands that it places on even unfamiliar users are far less.

It is therefore the claim of this paper that initiatives to examine closely the opportunities for the use of the telephone as an information interface using speech technology should be advanced in the Oceania region. The region presents a unique challenge to speech technology on account of its linguistic, geographic, and cultural composition.

ACKNOWLEDGMENTS

The assessment presented in this paper is of necessity highly reliant on multi-disciplinary input. In preparing this paper I am highly indebted to the valuable insights of many colleagues who have all freely shared their expert insights with me. I specifically acknowledge the assistance of following people: Andrew Pawley, Malcom Ross and Gerard Ward (Research School of Pacific and Asian Studies at ANU), Robert Early (University of the South Pacific, Vanuatu), and Mark Forsyth (Voicenet Australia). All responsibility for any errors introduced into this paper rests with the author.

REFERENCES

- Central Intelligence Agency, 1999. *Central Intelligence Agency - The World Factbook*, <http://www.odci.gov/cia/publications/factbook>
- Early, R. 2000. Personal Communication on 12 April 2000.
- Forsyth, M., 2000. Personal Communication on 24 March 2000.
- International Telecommunication Union, 1996. *Main Telephone Lines*, Oceania, http://www.itu.int/tif/industryoverview/at_glance/deloc.htm
- Mihalic, F., 1971. *The Jackaranda dictionary and grammar of Melanesian Pidgin*, Brisbane: Jackaranda Press.

Millar, J.B., Dermody, P., Harrington, J.M., Vonwiller, J. 1990. "A national spoken language database: concept, design, and implementation," *Proc. Int. Conf. Spoken Language Processing (ICSLP-90)*, Kobe, Japan, pp.1281-1284.

Millar, J.B., Vonwiller, J.P., Harrington, J.M., Dermody, P.J. 1994. "The Australian National Database Of Spoken Language," *Proc. ICASSP-94*, Adelaide, 1, 97-100.

Millar, J.B., Harrington, J.M., Vonwiller, J.P. 1997. "Spoken Language Resources for Australian Speech Technology," *J. Elec. Electro. Eng. Australia*, 17, 13-23.

Millar, J.B. 1997. "Knowledge and Ignorance in Speech Processing," *Proc. Int. Conf. Speech Processing (ICSP'97)*, Seoul, Korea, 1, pp.21-27.

Pawley, A.K. 1995. "Language," in *The Pacific Islands: Environment and Society*, Bess, Honolulu, pp. 181-194.

Qvortrup, L. 1993. "Community Teleservice Centres and Rural Revival," *Proc. Telecottages '93*, Queensland, Australia, pp.69-82.

Qvortrup, L. 1994. "Community TeleService Centres: A means to social, cultural and economic development of rural communities and low-income urban settlements," *ITU World Telecommunication Development Conf. (WTDC)* Buenos Aires, 21 p.

Ward, R.G., 1995. "The Shape of Tele-Cost Works: the Pacific Islands Case," in A.D.Cliff, P.R.Gould, A.G.Hoare, and N.J.Thrift (eds.), *Diffusing Geography: Essays for Peter Haggett*, Blackwell, Oxford and Cambridge.



Noise Control-

your solution is here.

We have been designing, manufacturing and installing noise control equipment since 1970. We help you control noise in your plant from initial on-site evaluation to confirmation of performance on completion.

Our off the shelf and custom built solutions include: enclosures, control rooms, acoustic panel systems, silencers, acoustic louvres, doors, audiometric booths and so on.

Noise control is all we do. Call NOW for details.

Peace Engineering Pty. Ltd.
2-20 Marigold Street, Revesby, NSW 2212
PO Box 4160, Milperra, NSW 1891
Phone: (02) 9772 4857 Fax: (02) 9771 5444
www.peaceng.com.au



Quality
Engineered
Components

 **Peace**

NOISE & VIBRATION CONTROL

HIRE & SALES

NOISE & VIBRATION LOGGERS

The Environmental Noise Logger is a self-contained weather resistant sound level meter with statistical processing capabilities and battery backed memory.



- E N M - ENVIRONMENTAL NOISE MODEL NOISE PREDICTION SOFTWARE

A computer program developed especially for Government authorities, acoustic & environmental consultants, industrial companies and any other group involved with prediction of noise in the environment.

For further information and/or enquiries, contact:

 **RTA TECHNOLOGY PTY LTD**
Level 1, 418A Elizabeth St. Surry Hills NSW 2010
Telephone: (02) 8218 0570 Fax: (02) 8218 0671
Email: rtatech@rtagroup.com.au
Website: www.rtagroup.com.au

**Larson Davis
and OROS...
two sound reasons
to talk with
Davidson**

Davidson offers an extensive range of equipment from basic single channel sound level meters to multiple channel systems that are at the forefront of acoustic technology.



Products are supported by a specialised engineering expertise and a thorough understanding of the nature of sound.

Call Davidson now for technical information or visit our Website.

For measurements in acoustic and vibration applications plus modal analysis and vibration diagnosis. Multiple channel operation, PC integration. **OROS PC-Pack.**

M B & R J Davidson Pty Ltd
Head office, Victoria:
5-1 Lakeside Boulevard
Bonville Victoria 3194
Ph (03) 9580 4368
Fax (03) 9580 6499
E-mail: info@larsondavis.com.au
Web: www.larsondavis.com.au

New South Wales/ACT
Ph (02) 9748 0544

Queensland
Ph (07) 3853 5300

Western Australia/NT
Ph (08) 9301 0923

Tasmania
FreeCall 1300 816 687

South Australia
FreeCall 1300 816 687

Packed with features and capable of downloading and translating binary data to ASCII text.
Larson Davis Model 824 with S24 (11) software.



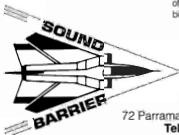
DAVIDSON
Measuring up to your needs
...since 1973

Welcome to the quiet world of **SOUND BARRIER SYSTEMS**

DOUBLE GLAZE and HEAR the difference

HERE IS WHY:

- ★ Up to 46 dB
- ★ Custom designed to suit all applications
- ★ Affordable and cost effective solutions
- ★ Existing external appearance of the building remains unaltered
- ★ Specially selected safety glass offers security
- ★ Extremely durable
- ★ Easy to clean and maintain
- ★ Powder coated aluminium frames in a variety of colours
- ★ suited to homes, schools, units, offices, hospitals and other buildings
- ★ Easy access to opening and closing
- ★ Simple and time effective application
- ★ Stay warmer in winter and cooler in summer
- ★ 20 years of experience
- ★ Award winning designs
- ★ Conditional 7 year warranty



SYSTEMS
ACN 002 3119 8155

SHOWROOM

72 Parramatta Road, Lewisham NSW 2049

Telephone: 02 9540 4333

Facsimile 02 9540 4355 Mobile 0412 478 275



A COMPARISON OF TWO ACOUSTIC METHODS FOR FORENSIC SPEAKER DISCRIMINATION

Phil Rose* and Frantz Clermont**

*Phonetics Laboratory, Linguistics Program, Australian National University

**School of Computer Science, University of New South Wales (ADFA)

ABSTRACT A pilot forensic-phonetic experiment is described which compares the performance of formant- and cepstrally-based analyses on forensically realistic speech: intonationally varying tokens of the word *hello* said by six demonstrably similar-sounding speakers in recording sessions separated by at least a year. The two approaches are compared with respect to F-ratios and overall discrimination performance utilising a novel band-selective cepstral analysis. It is shown that at the second diphthongal target in *hello* the cepstrum-based analysis outperforms the formant analysis by about 5%, compared to its 10% superiority for same-session data.

1. INTRODUCTION

A recurrent topic in Forensic Phonetics, where speaker verification under very much less than optimum conditions is a major concern, is its relationship to Automatic Speaker Recognition (ASR). Leading forensic phoneticians (e.g. Künzel 1995: 79) emphasise the difference in the real-world conditions between Automatic and Forensic speaker recognition, especially in the lack of control over operational conditions in forensic speaker identification, and point out that fully automated forensic speaker identification is not a possibility.

This should not imply, however, that some of the analytical techniques common in ASR are of no forensic use. Although forensic speaker identification must usually rely, *inter alia*, on comparison of individual formants (e.g. Nolan 1990, Labov & Harris 1990: 287ff.), it is generally assumed that in ASR cepstrally-based methods are superior. This is because the cepstrum tends to exhibit strong immunity to "noninformation-bearing variabilities" (Rabiner & Juang, 1993: 169) and, hence, greater sensitivity to distinctive features of speech spectra. Aside from actual performance, the cepstrum is more easily extracted than the F-pattern, with its inevitable problems of identification and tracking of the higher formants. Although the nature of the cepstrum *qua* smoothed spectral shape is far from *unanschaulich*, arguments against the use of the cepstrum in forensic phonetics centre on the abstract nature of its mathematical basis (van der Giet 1987: 125), and include its indirect relationship to auditory and articulatory phonetic features - the latter of considerable importance in forensics - and the difficulty of explaining it to the jury (Rose 1999b: 7). It is of both interest and importance, therefore, to examine the performance of that algorithmic mainstay of ASR - the cepstrum - on forensically realistic data. That is the aim of this paper. It has only recently become practical due to mathematical developments (Clermont & Mokhtari 1994), which make it possible to specify the upper and the lower bound of any frequency band directly in the computation of the cepstral distance.

The speech data we use are forensically realistic in four important ways. Firstly, they are from speakers that sound similar. This is an obvious requirement on any forensically realistic speaker discrimination experiment, since if two speech samples do not sound similar, *ceteris paribus*, it makes little sense to claim that they come from the same speaker. Secondly, the data are from different sessions. If the data were from the same session, the criminal would be known. Thirdly, the data is not controlled intonationally. This is because, even if the same word occurs in criminal and suspect samples, it is unlikely that it will occur in exactly the same prosodic environment in criminal and suspect material. Lastly, we use a word very common in telephone intercepts - *hello*. We therefore use the most controlled data that can be realistically expected, namely variation within a segment that occurs in the same position in intonationally varying repeats of the same word. The word *hello* is capable of taking naturally a wide range of contrasting intonational nuclei, thus providing a potentially greater range of within-speaker variations.

2. PROCEDURE

Subjects

Six demonstrably similar-sounding, adult-male native speakers of General to slightly Broad Australian English were recorded. Four of the speakers are closely related: JM, his two sons DM and EM, and his nephew MD. RS and PS are father and son. Similar-sounding means similar sounding to naive listeners, and presumably rests on similarities in auditory voice quality rather than *phonetic* quality (for this important distinction, see the collection of papers in Laver (1991)). The speakers had been chosen initially on the basis of anecdotally reported similarity (it was claimed for example that a father and son were commonly confused by their wife and mother over the telephone). The six speakers were shown in subsequent experiments reported in Rose and Duncan (1995) to indeed have voices similar enough to be confused in open identification and discrimination tasks even by closest family members. It is not surprising that perceptual discrimination

tests with naive unfamiliar listeners also showed the six voices to be highly confusable.

Recordings and Cepstral Processing

Use was made of two sets of recordings to furnish genuine long-term data for comparison. These were separated by a period of four years (DM) and one year (the others), and are referred to as R(ecording) 1 and R(ecording) 2. Details of the within- and between-speaker variation in the two sets of recordings can be found in Rose (1999a) for R1, and Rose (1999b) for R2. Two sets of data were obtained in the second recording, and data from the second set were used. In order to elicit a selection of realistically varying intonational patterns, speakers were asked to say the word *hello* as they imagined they might say it under six different situations. (1) answering the 'phone, (2) announcing their arrival home, (3) questioning if someone was there, (4) greeting a long-lost friend, (5) passing someone in the corridor, (6) reading it off the page. In the second recording session these were expanded to: (7) meeting the Prime Minister, (8) admiring someone's appearance, and (9) trying to attract someone's attention. Some speakers, especially EM, preferred utterances other than *hello* (e.g. *Hi, Hey, G'day*) for some situations, and so had less *hello* tokens than the others.

Table 1. Numbers of tokens recorded

	DM	JM	EM	PS	RS	MD
R1	17	6	3	4	7	6
R2	9	9	7	10	9	9

The *hellos* were recorded using professional equipment in the A.N.U. phonetics laboratory recording studio. The resulting analogue signals were then sampled at 10 kHz, and analysed (ILS API routine) by linear prediction (LP-order 14) of 20msec Hamming-windowed frames with 100% pre-emphasis and a frame advance of 6.4msec. The boundaries of the /l/, the offset of modal phonation in /ou/, and the onset of the first vowel were determined from inspection of the waveform produced by the ILS SGM command (yielding a quasi-spectrogram plot), in conjunction with conventional analog wide-band spectrograms. The following seven temporal landmarks were defined: the middle of the /l/; 25% intervals of the duration of the /ou/; and the middle of the first vowel if present. The ILS analysis frames corresponding to the landmarks were then identified, the centre-frequency of the first four formants identified, and transferred to a spreadsheet for statistical analysis. In addition, the set of 14 LP-derived cepstral coefficients corresponding to each landmark were retained for further processing.

Cepstral distances were calculated both for the entire Nyquist interval, and also for sub-bands of this interval. This was done in order to obtain cepstral analogues of the formant-based measures of variance and distance that are commonly sought in forensic speaker identification. To this end we used Clermont & Mokhtari's (1994) parametric formulation of the cepstral distance, which permits a posteriori specification of the upper and the lower bound of any frequency sub-band

between 0Hz and the Nyquist frequency. The sub-bands corresponded to the spectral regions straddling the frequency range of each of the four observed formants. The upper bound for each sub-band was set at the frequency of the highest mean-formant's centre-frequency observed plus one standard deviation, and the lower bound at the lowest mean minus one standard deviation. For example, the highest mean centre-frequency (499 Hz) for F1 in /l/ was produced by RS, with a standard deviation of 17 Hz; and the lowest mean centre-frequency (405 Hz) for F1 in /l/ was produced by PS, with a standard deviation of 30 Hz. The sub-band constrained to the F1-range was thus specified in terms of an upper bound of $499+17 = 516$ Hz and a lower bound of $405-30 = 375$ Hz.

3. RESULTS

Intonation

As intended, the different situations did elicit a forensically realistic variety of different intonational patterns. Thirteen different patterns occurred, which were formally classifiable according to their nuclear pitch into five types: *Fall, Rise, Downstep, Fall-Rise and Rise-fall* (Rose 1999b: 10). With the exception of JM, who produced proportionately more downsteps, the between-speaker intonational variety was largely comparable.

Auditory phonetic quality

Although the speakers were largely comparable in the suprasegmental aspects of their phonetic quality, they showed both between- and within-speaker segmental variation in the backness and rounding of the diphthongal offglide in /ou/. (Realisations of Australian /ou/ typically show a wide range in the backness of the diphthongal offglide). The /ou/ diphthongs in the data collected here have an offglide ranging between [y-] and [u-*u*+], and a fairly open central initial target [ɐ]. They are thus representative of a major part of the typical range. Two speakers (PS and RS) consistently had what sounded like a backer/rounder off-glide: [u-]; DM's offglide was consistently fronter: [u], and JM's offglide sounded slightly fronter and lower: [u_l]. The other two speakers showed within-speaker variation. Some of EM's /ou/ tokens sounded the same as DM's, and some sounded backer/more rounded, although not as much as PS and RS. MD was notable for his wide range of off-glide realisations, from [u+] through [u] to [y-]. Also noticeable were differences in the secondary articulation of /l/ (pharyngealised vs velarised), and incidental differences in the first vowel phoneme /a/ vs. /e/ vs. /æ/. An important point is that, as a result of these auditory linguistic differences, it was possible to discriminate rather easily some pairs of speakers who had similar voice quality but different phonetic quality.

Formant analysis

As might be expected from the similarity in their auditory voice quality, some pairs of speakers had very similar mean F-patterns. Within-session Euclidean distances were calculated for all between-speaker pairs for all four formants both combined, and individually for both recordings. Figure 1 shows the mean F-patterns of the two most similar speakers in R2 (PS, DM), according to overall Euclidean distance. The

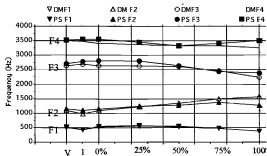


Figure 1. Mean F-patterns compared for PS and DM.

mean Euclidean distance for this pair over all four formants was 109 Hz, with individual formants, from F1 through F4, as follows: 15 Hz, 138 Hz, 118 Hz, 120 Hz. (In R1 the most similar pair was DM and MD, who were separated overall by 101 Hz, with individual formant differences of 39, 81, 160, and 85 Hz (Rose 1999a: 17)). It can be seen from figure 1 that PS and DM display a fairly high level of congruence in all formants except F3 at the onset of the diphthong, and F2, F3 and F4 at offset. Notably, the difference in F2 over the last two landmarks in /ou/ corresponds to an audible difference in the acuteness of the second diphthongal target.

In spite of the problems in formant identification alluded to above, it was generally easy to identify these six speakers' formants—for some even up to F5. There were two exceptions. In both recordings, JM appeared to have two close resonances in the area of F4, neither of which was unambiguously continuous. The higher of the two had to be identified as "true" F4 and the other as a singer's formant (Rose 1999a: 12-13). In RS's second recording, his F3 and F4 were not reliably extracted. These two speakers offer the possibility for cepstral analysis to demonstrate its superiority.

4. ANOVA COMPARISON

Preliminary to the discrimination, in order to find out where the points of greatest within- to between-speaker variation in *hello* lay, a single factor ANOVA was carried out for both the formant and cepstral data, on the data in both R1 and R2, at each of the sampling points. The resulting F-ratios are given in table 3. Very few comparable points exist between the cepstral and formant F-ratios (one of the reasons for this is because the F-ratios for the formants across both recordings are significantly correlated, whereas those for the cepstral analysis are not). However, both cepstral and formant analyses do agree in the status of the 75% landmark. This is the point at which the highest within- to between-speaker values occur both across recordings and across analyses. (For formants this is in terms of the sum of the F-ratios of the individual formants at a landmark; for the cepstra in terms of the highest whole-range value.) It is thus possible to say that the greatest between- to within-speaker long-term variation occurs at the same landmark (75%) in both the F-pattern and the C-pattern. Their discrimination performance was accordingly tested at the 75% landmark.

Table 2. F-ratios for cepstral and formant analysis.

	V	I	0	25	50	75	100
R1 cep							
F1-range	2	2	7	5	3	5	2
F2-range	7	5	4	3	5	16	7
F3-range	4	7	6	5	3	6	2
F4-range	4	6	6	6	4	4	2
Full-range	3	5	5	5	4	6	3
R2 cep							
F1-range	13	5	13	13	3	11	10
F2-range	10	14	12	9	9	13	10
F3-range	12	12	15	18	13	6	5
F4-range	5	4	5	7	15	15	7
Full-range	8	7	9	9	10	11	7
R1 form							
F1	4	3	8	9	3	3	8
F2	0	11	1	2	7	25	9
F3	4	6	7	5	4	6	2
F4	7	7	14	18	24	13	4
R2 form							
F1	5	2	7	9	2	7	9
F2	2	23	6	10	19	23	12
F3	10	15	13	11	13	11	10
F4	9	14	10	10	43	46	17

5. DISCRIMINATION ANALYSES

In forensic phonetic case-work, the emphasis is on discrimination between same-voice samples and different-voice samples. This differs somewhat from the conventional (identification) sense of discriminant analysis, which is concerned with assigning to a set of pre-established classes (here speakers) an unlabelled token observed in addition to those used to determine the classes (Woods, Fletcher & Hughes 1986: 266). Forensically, identification is the secondary result of a process of discrimination. If it is decided that two samples come from the same voice, the suspect is identified as the criminal. If not, no identification results. In this experiment, therefore, discrimination does not mean being able to identify individuals, but being able to say, given any pair of *hellos* from our data set, whether or not they come from the same speaker. In this experiment, we wanted to find out how much better a cepstral analysis can do this than a formant analysis.

Both cepstral and formant analyses in the preceding section showed that *hello* has the most individual-identifying information at the 75% landmark. Two tests were accordingly performed to compare the discriminant power of formant- and cepstrally-based analyses at the 75% landmark on same-speaker and different speaker pairs of *hellos*. The first test was carried out with the same-session data of the second recording. In this test, all possible within- and between-speaker pairs of *hellos* were tested. Thus, for example, DM's first *hello* token in his second recording was compared with all his other tokens in his second recording, and all other tokens from the second recording of all other speakers. In all, then, 210 within-speaker pairs of *hellos* were compared in the first test, and 1168 between-speaker pairs.

Although it quantifies the relative performance of the cepstral and formant analyses, this test is forensically unrealistic because it uses single-session data (Rose 1999b:1.2). Therefore a second, forensically more realistic, test was performed with the long-term different session data provided by recordings 1 and 2. In this test, the within-speaker comparison was, of course, across the two recording sessions. Thus, for example, all DM's hello tokens in his first recording were tested against all his hello tokens in his second recording, and against all the *hellos* of all other speakers in both recordings. The second test involved 376 within-speaker and 3688 between-speaker comparisons. The second test thus simulates a situation where a criminal and a suspect sample, separated by a long stretch of time, are being compared using one hello token in each sample. (In reality, of course, much more material in each sample would be compared, and usually the samples would be separated by a much shorter stretch of time.)

The tests are crude, and make use of nothing but unweighted distances between samples as thresholds. First, the mean between-speaker and within-speaker distances, and the mean standard deviation of the between-speaker and within-speaker standard deviations, were calculated for values at the 75% point. The discriminant threshold was then set at halfway between the between- and within-speaker mean values. Given the similar standard deviations observed with this procedure, this should ensure that values close to an EER should be obtained, assuming distributional normality. The EER was then found as the mean of the discriminant performances for the between- and within-speaker comparisons. Because the F-ratio values for F1 and F3 at 75% were not so high as for F2 and F4, performance was evaluated only for F2 and F4 in the formant analysis. We did not know what to expect for the cepstrum, so we evaluated the cepstral performance at all four formant ranges, as well as over the whole range.

6. RESULTS

Results are shown, as equal error percent correct performance, in table 3. Table 3 shows firstly that, as expected, performance decreases with the different session data. The best performance (79%) is clearly obtained by (whole-range) cepstral analysis for the same session data, but both analyses perform equally well, as far as best performances are concerned, for the different session data: the value for F4 (64%) is effectively the same as the 63% for the whole-range

cepstrum. It is, of course, highly unlikely that F4 will be available for use in real forensic case-work, barring comparison with rhotics, so perhaps it is more realistic to draw comparisons with F2. Here the results are clearer. The cepstral analysis is 10% better than the formant in the same session data (79% vs. 69%), and 5% better (63% vs. 58%) for the different session data. While F4 is not fully admissible on grounds of availability or measurement unreliability, cepstral analyses spanning the entire Nyquist interval are not thus hampered, and can therefore be justifiably exploited to implicate the higher-formant range. It can be noted, moreover, that the F2 range cepstrum performance (62%) is still 4% better than the formant analysis with F2.

The fairly good agreement observable between the performance for the individual formants and that for the cepstral formant ranges is presumably because the former are the primary determinants of the spectral shape. However, it is also a nice indication that the cepstral sub-band analysis works. It is important to note that the fact that no sub-band analysis outperformed the full-range analysis does not automatically indicate the superiority of the latter. This is because we deliberately constrained the sub-bands to correspond to formant ranges. In the different-session comparison, the F2 sub-band (62%) contains effectively as much discriminating information as the whole spectral range (63%). It is therefore entirely possible that better performance might occur with unconstrained sub-bands than with the full range. If this is the case, an unconstrained sub-band cepstral discrimination might have the potential to outperform a formant discrimination by more than the demonstrated 5%.

7. SUMMARY, CONCLUSION AND WAY AHEAD

This paper has shown that, in the very restricted task of comparing samples at a single landmark in hello, the cepstrum does discriminate forensically realistic data better than formants, by at least 5%. Moreover, the overall good discrimination performance of the cepstrum at some other landmarks in hello (not demonstrated in this paper) indicates that it is less sensitive to different landmarks than the formant analysis. In forensic science, performance must outweigh understandability for juries. In addition, the practicality of the cepstrum in avoiding measurement problems that are inherent to formant-frequency estimation is also a criterion in its favour. Thus we conclude that spectral shape parameters like the LP-cepstral coefficients do have a more important role to play in forensic speaker identification than has been demonstrated to date. Whether this role is as an adjunct or as an alternative to the formants remains to be seen.

Further research is required into the effects of different recording conditions (e.g. telephone); the pre-treatment of cepstral coefficients; of sample size; and the use of more sophisticated discrimination strategies, including weighting; the involvement of more than one landmark; and unconstrained sub-bands. It will also be interesting to see whether the cepstrum produces a more homogeneous set of results with respect to individual speakers and speaker pairs. With formants, different-session within-speaker

Table 3. Equal discrimination performance (%) of Cepstrum (C) and Formant (F) analyses at 75% of /ou/ in hello.

Formant/ Cepstral range	same session (R2)		different session (R1 & R2)	
	C	F	C	F
F1	59		56	
F2	70	69	62	58
F3	64		56	
F4	72	71	58	64
Full	79		63	

discrimination of RS is particularly bad, for example, and only offset by good performance with other speakers. Ideally, of course, all same-speaker hello pairs must be discriminable from different-speaker pairs. Ultimately, however, we will need to move away from the average probabilities of overall discrimination rates to the calculation of likelihood ratios for cepstral distances, so that the latter can be used within the appropriate Bayesian approach for forensic science (Champod & Meuwly 2000).

REFERENCES

- Champod, C. & Meuwly, D. (2000) "The inference of identity in forensic speaker recognition." *Speech Communication* 31, 193-203.
- Clermont, F. & Mokhtari, P. (1994) "Frequency-band specification in cepstral distance computation." In Roberto Togneri (ed.) *Proc. Fifth Australian Internat. Conf. on Speech Science and Technology*. Canberra: ASSTA, 354-359.
- van der Giet (1987) "Der Einsatz des Computers in der Sprechererkennung." In Künzel, H.J. *Sprechererkennung: Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag, 121-132.
- Künzel, H.J. (1995) "Field Procedures in forensic speaker recognition." In J.W. Lewis (ed.) *Studies in general and English phonetics. Essays in honour of J.D. O'Connor*. London: Routledge, 68-84.

Labov, William & Harris, Wendell A. (1990) "Addressing social issues through linguistic evidence." In John Gibbons (ed.) *Language and the Law*. New York: Longman, 265-305.

Laver, J. (1991) *The Gift of Speech*. Edinburgh: EUP.

Nolan, Francis (1990) "The limitations of auditory-phonetic speaker identification." In H.Kniffka (ed.) *Texte zur Theorie und Praxis forensischer Linguistik*. Tübingen: Niemeyer.

Rabiner, L. & Juang, B-H.J. (1993), *Fundamentals of Speech Recognition*. Englewood Cliffs, N.J. Prentice-Hall.

Rose, P. (1999a) "Differences and distinguishability in the acoustic characteristics of hello in voices of similar-sounding speakers: a forensic phonetic investigation." *Australian Review of Applied Linguistics* 22(1), 1-42.

Rose, P. (1999b) "Long- and short-term within-speaker differences in the formants of Australian hello." *Journal of the International Phonetic Association* 29(1), 1-31.

Rose, P., & Duncan, S. (1995). "Naive Auditory Identification and Discrimination of Similar Voices by Familiar Listeners." *Forensic Linguistics* 2(1), 1-17.

Woods, A. Fletcher, P. & Hughes, A. (1986) *Statistics in language studies*. Cambridge: CUP.



ARL Sales & Hire RION

Noise, Vibration & Weather Loggers Sound & Vibration Measuring Instruments

**New EL-316
Noise Logger
Type 1
Accuracy**



**Sound Level
Meters &
Vibration
Analysers
Environmental
Noise Loggers**

ACOUSTIC RESEARCH LABORATORIES

Proprietary Limited A.C.N. 050 100 804

Noise and Vibration Monitoring Instrumentation for Industry and the Environment



Reg Lab 14172
Acoustic & Vibration
Measurement

ARL Sydney: (02) 9484-0800 **ARL Melbourne:** (03) 9897-4711 **Australian Acoustical Services Perth:** (08) 9355 5699
Wavecom Adelaide: (08) 8331-8892 **Belcur Brisbane:** (07) 3820 2488

CATT - ACOUSTIC

Room acoustic prediction and desktop auralisation

CATT-Acoustic v7 is a seven-module Windows 95 & NT 4.0 application. It integrates prediction, source addition, auralisation, sequence processing, directivity, surface properties and post processing.

Prediction Module employs the unique Randomised Tail-corrected Cone-tracing (RTC) method as well as Image Support Module (ISM) and ray-tracing settings to create numerical results, plot-files and optionally data for the multiple source and post-processing modules. Geometry editing is performed in a customised editor linked to the main program or via the AutoCAD™ interface.

Surface Properties Module manages and controls surface properties. Named properties can also be defined directly in geometry files.

Multiple Source Addition

Module creates new echograms based on results from the prediction module. Source directivity, aim, eq and delay



can be varied without need for a full re-calculation. The module optionally creates data for multiple source auralisation.

Source Directivity Module imports data in the common measured 10° format, interpolates from horizontal and vertical polar measurements, or uses a unique DLL-interface, which can also perform array modelling.

Post-processing Module transforms octave-band echograms, created by the prediction module, via HRTFs and DSP procedures, into binaural room impulse responses. These are convolved with anechoically recorded material to produce the final 3D audio sound-stage. The module offers many post-processing options, transaural replay, multiple source auralisation, software convolution, head-phone equalisation, and an assortment of file format conversions, scaling and calibration utilities.

Plot-file Viewer Module displays, prints and exports graphics created in CATT. Lists of plot-files can be created for presentations, optionally with auto-playing WAV files.



Sequence Processing Module manages CATT tasks, allowing for batch processing of all stages, from prediction to binaural post-processing and convolution, unattended.

Lake Technology, the exclusive supplier of CATT-Acoustic in Australia and New Zealand also provides hardware based real-time 3D audio solutions.

Lake's proprietary zero-latency convolution algorithms and real-time simulation software are fully compatible with CATT-Acoustic. Live "head-tracked" 3D audio presentations are now possible utilising CATT-Acoustic and Lake's Huron Convolution Workstation or the CP4 Convolution Processor.

Contact Lake DSP or CATT-Acoustic (www.netg.se/~catt) for demo disks or Lake Technology demonstration CD-ROM.

Lake Technology Limited

Suite 502, 51-53 Mountain Street
Ultimo Sydney NSW 2007
Tel: + 612 9213 9000
Fax: +612 9211 0790
email: info@lake.com.au
web: www.lakedsp.com



AIRSERVICES AUSTRALIA

Environment Technical Specialist
MM1 (\$67,389 - \$ 83,893)

Branch/Section: Environment Services Branch/ Environment Monitoring Section
Location: Canberra

Primary Purpose: The role of this position is to provide technical expertise in the measurement, monitoring and reporting of environmental impacts, particularly noise from aircraft operations. It is also involved with: the operation of Airservices' Noise and Flight Path Monitoring System (NFPMS); the assessment of aircraft for compliance with statutory noise limits; and oversight of the program of calibration of technical equipment used for precision measurements.

Position in Context: While the Environment Services Branch is located in Canberra, its area of responsibilities covers aircraft operations Australia-wide. Reports, data and advice are frequently requested by and provided to the Department of Transport and Regional Services, the Minister, aircraft and airport operators and to the general public through airport environment committees. The work is linked to the requirements of international and national standards.

Qualifications and Experience: The applicant will require professional qualifications in engineering or science. Knowledge of acoustics, and extensive experience in the measurement and analysis of noise, are essential. A knowledge of the environmental effects of aircraft operations is desirable.

Selection Criteria for the position can be obtained by contacting Susan Hill on (02) 6268 4711 or email susan.hill@airservices.gov.au.

Applications should be forwarded by 8 June 2001 to:

Susan Hill
Airservices Australia/Environment Services Branch
PO Box 367
Canberra ACT 2601

Airservices Australia values social and cultural diversity and is committed to the principles of Equal Employment Opportunity and Occupational Health & Safety.

www.airservices.gov.au

AUDITORY AND F-PATTERN VARIATIONS IN AUSTRALIAN *OKAY*: A FORENSIC INVESTIGATION

Jennifer R. Elliott

Phonetics Laboratory

School of Language Studies

The Australian National University

ABSTRACT. An understanding of the acoustic properties, as well as the nature of within- and between-speaker variation, of words which occur with high frequency in natural discourse, is of great importance in forensic phonetic analyses. One word which occurs with relatively high frequency in natural discourse, including telephone conversations, which are often a source of data in forensic comparisons, is *okay*. This paper presents the initial findings of a study of auditory and F-pattern variations in *okay* in a natural telephone conversation spoken by six male speakers of general Australian English. Seven pre-defined sampling points are measured within each token to determine the most efficient sampling points and formants for distinguishing between-speaker variation from within-speaker-variation in *okay*. F-ratios at these seven sampling points are calculated as a mean of ratios of between- to within-speaker variation. The greatest F-ratio is shown to be for F₄ at voice onset of the second vowel. Forensic implications are discussed.

1. INTRODUCTION

When phoneticians are asked to compare samples of speech for forensic purposes they are faced with a specialised case of speaker verification which involves comparing a sample of speech which is known to be associated with a crime, with another sample of speech from a known person who is suspected of being involved in the crime. This forensic application of speech analysis is based on the assumption that there will be greater variation between speakers than within a speaker.

Nolan (1983: 6-14) notes that forensic speaker verification or identification tasks are inherently more complicated than other forms of speaker identification, where a sample of speech is compared against another predetermined sample for the purpose of authenticating or verifying a speaker is who he or she claims to be. Apart from the obvious difficulties inherent in comparing speech samples recorded at different times and usually under very different conditions, the speech samples used in forensic phonetics are invariably both uncontrolled and restricted in content, leaving a minimal amount of speech for analysis and comparison. The recording of the criminal, for example, may constitute only a few short words. It is desirable that the linguistic data from both samples used in a forensic comparison are, if possible, linguistically equivalent, and the best results are likely to be obtained when the same lexical items are compared. For this reason words which occur frequently in conversation are likely candidates for analysis and comparison.

One word that is used frequently as a tool of negotiation in conversational English is *okay*. This word functions both as a response such as agreement, acceptance or confirmation to preceding talk, and/or as a transitional device between two stages of a conversation (Merrit 1984; Condon 1986).

Furthermore, as Schegloff (1979, 1986) and Schegloff and Sacks (1984) have demonstrated, *okay* occurs frequently in both openings and closings of telephone conversations, which in turn are the most common source of recordings used in forensic comparisons. The question therefore arises: is *okay* an appropriate word to use in forensic analysis, and if so, how useful is it for distinguishing between speakers?

Research by Rose (1997, 1999), in which the within- and between-speaker differences in *hello* spoken by six speakers were examined, demonstrated that even similar sounding speakers "can be distinguished on the basis of significant differences in their acoustics" (Rose 1997: 35). Based on these findings, a similar hypothesis was proposed for the present study: that there will be greater variation in the acoustics of *okay* between speakers than within a speaker. If this hypothesis was confirmed then a secondary question would arise: which parts of the word *okay* provide the clearest evidence of between-speaker differences? The research was designed both to test the hypothesis and, if the null hypothesis was disproved, to seek an answer to this question. Although both auditory and acoustic analyses are indispensable in forensic analysis, one of the key measures of comparison of forensic phonetic acoustic analysis is the formant- (F-) pattern of short-term segments. This paper describes briefly the auditory variations in the phonetic realisations of ten tokens of *okay* from each of six different speakers of general Australian English (as described by Mitchell & Delbridge 1965, Burridge & Mulder 1996, for example), and reports the F-pattern variations of these same tokens when examined from an acoustic phonetic perspective. This study represents the first stage in a broader research project on the subject of auditory and acoustic within- and between-speaker variations in Australian *okay*.

2. EXPERIMENT DESIGN AND DATA COLLECTION

In keeping with the nature of data used in forensic phonetics, a premium was placed on the data being collected from natural conversation. A map task was devised to engage pairs of participants in a conversation requiring negotiation, potentially leading to the elicitation of several tokens of *okay* from each speaker. In order to encapsulate each conversation as a closed speech event, the task was carried out by telephone, thus providing a distinct beginning and end to each interaction. Recording a speaker engaged in a telephone conversation had two additional advantages. Firstly, it enabled a clean speech signal of a single speaker conversing with someone else to be recorded without the attendant confusion of overlapping talk from the other speaker (a common characteristic of natural conversation). Secondly, since there was no eye contact between the speakers, all communication had to be verbal, thus increasing the opportunity for negotiation, and hence the likelihood of eliciting numerous tokens of *okay*. The recordings used in acoustic analysis were made directly, and not through the telephone.

The study involved six native speakers of general Australian English working in pairs, as indicated in Table 1. All participants were aged between 16 and 20 years, and were from similar socio-economic backgrounds. In order to minimise the effect of convergence of linguistic styles between the participants (Giles & Coupland 1991: 60-93), each pair was also well acquainted. In addition, a number of the participants were from the same family (they were either brothers or cousins), and although they were not necessarily paired together, it was hoped that this would impose a slightly higher level of control over the possibility of confounding sociolinguistic variables.

Table 1. Pairs of participants

Caller	DL	EO	GO	MO	JE	PE
Recipient	JE	PE	PE	JE	MO	GO

The map task involved two similar, but not identical maps. The caller was required to guide their partner (the recipient of the telephone call) through a predetermined route marked on the map. The negotiation of the differences between the maps would provide the opportunity for the elicitation of tokens of *okay*. The caller was recorded directly in the recording studio of the Phonetics Laboratory at the Australian National University, using a SONY ECM-100 STEREO CASSETTE deck and a Sony ECM-909A microphone. From this recording the ten tokens of *okay* which could be most easily isolated from the surrounding talk, and which had the least excess noise, were extracted for acoustic analysis.

3. AUDITORY ANALYSIS

The Australian Oxford Dictionary (published in 1999) suggests that the Australian English pronunciation of *okay* /ou'keɪ/ has three phonemic segments, consisting of two diphthongs (V_1 and V_2), separated by a voiceless velar stop (C). Auditory analysis of each of the sixty tokens studied

Table 2. Occurrences of different phonetic realisations of Australian *okay* segments by each speaker

V_1 /ou/	DL	EO	GO	MO	JE	PE
ə	3	4	8	0	6	0
u	6	5	0	7	3	0
ɑ	1	1	2	1	1	0
o	0	0	0	0	0	9
ɛ	0	0	0	1	0	1
e ^ɹ	0	0	0	1	0	0
C /k/						
k ^h	10	10	6	10	9	4
k	0	0	4	0	1	1
g	0	0	0	0	0	4
x	0	0	0	0	0	1
V_2 /eɪ/						
eɪ	8	3	9	8	3	9
e ^ɹ	0	1	1	0	0	0
ɛ	0	6	0	0	7	1
æ	2	0	0	1	0	0
^h eɪ	0	0	0	1	0	0

showed considerable variation in the phonetic realisations of these segments, both within and between speakers. Phonetic realisations of each of the three segments from auditory analysis are set out in Table 2.

Table 2 shows that V_1 was realised as a diphthong only once out of the 60 tokens analysed. Interestingly, this particular token was also irregular in that V_2 was palatalised ([e^ɹk^heɪ]). Thus the generalisation can be made that V_1 of *okay* in conversational general Australian English is usually realised as a monophthong. Moreover, this monophthong was in the majority of cases, centralised to [ɔ] (a typical realisation of unstressed vowels) or centralised and lowered to [ʊ]. One token of the low back vowel [ɑ] was also elicited from each speaker except PE, whose V_1 was realised 90% of the time as the slightly raised central rounded vowel [o].

The /k/ was most commonly realised as an aspirated voiceless velar stop. For example this was the case 100% of the time for DL, EO and MO, and 90% of the time for JE. The stop was aspirated in six of GO's tokens, while the remaining four were unaspirated voiceless stops. PE again differed the most, with only four tokens being aspirated, while one was a voiceless unaspirated stop, four were realised as voiced stops, and in one token the consonant was fricated throughout, without an audible hold phase.

With V_2 , 43 of the 60 tokens were realised as diphthongs. In keeping with the findings of previous studies of Australian English (for example Harrington et al. 1997,) the first target for this vowel was consistently lowered, and was realised as [ɛ] rather than [e]. In two instances, the offglide was more central than high, but in one of these cases, this may have been due to anticipatory coarticulation (Laver 1994: 379) for a bilabial

approximant, /w/, which followed in the next word, however this requires further investigation. In a number of instances, V_2 was not realised as a diphthong at all, but was realised simply as an open-mid front [e]. Six instances of this were elicited from EO, seven from JE and one from PE. Extreme lowering of V_2 to [æ] was also occasionally heard, twice by DL and once by MO, and in each of these instances V_2 was also realised as a monophthong. The incidence of both /e/ and /æ/ in V_1 of *okay*, suggests that there is possibly a choice of phonemes for this syllable in Australian English (c.f. Rose's (1997, 1999) findings for V_1 of Australian *hello*).

While the suprasegmental structure will not be discussed in detail in this paper, it should be noted that there is also considerable variation in the realisation of stress. In all but one instance, the major stress fell on the second syllable: EO provided the only token where the stress fell on S1, and the general lenition and centralising of V1 noted above may well be accounted for in terms of stress.

4. ACOUSTIC ANALYSIS

Tokens were digitised at 16 kHz, and the F-pattern was analysed on a CSL 4300 by generating wideband spectrograms, and using the FFT power spectrum facility overlaid with the LPC filter response at selected sampling points. A filter order of 20 kHz was used, with hamming window and 100% preemphasis. The first four peaks were measured to extract an estimate of the centre frequencies of the F-pattern, based on the expected frequencies for each given phonetic segment.

The primary aims of the experiment were to determine whether or not it is practicable to use *okay* in forensic comparisons, and if so, which part of the word *okay* provides the best F-pattern for determining between-speaker differences. Since the tokens were to be used for comparing both within- and between-speaker variations, it was essential that the sampling points were also comparable across all tokens. To ensure the integrity of measurements between all the tokens, seven sampling points were chosen at which to measure the first four formants. The decision to use these particular sampling points was motivated by the goal of extracting as much acoustic information as possible which could highlight significant differences between speakers.

The seven sampling points, illustrated in figure 1, were identified as follows:

- S₁ 1. within the first three regular glottal pulses of V₁ (V₁ onset);
2. within the last three regular glottal pulses of V₁ (V₁ offglide);
- S₂ 3. at consonant release (C release);
4. at phonation onset follow the release phase (PO);
5. within the first three glottal pulses of V₂ (V₂ onset);
6. at the lowest point of F₂ within V₂ (V₂ mid); and
7. at the highest point of F₂ within V₂ (V₂ offglide).

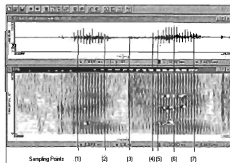


Figure 1. Wideband spectrogram showing sampling points of *okay* tokens

Table 3. F-ratios for each formant at each sampling point in order of magnitude.

Sampling Point	Formant	F-ratio	Confidence level
V2 onset	F4	32.367	.000
V2 offglide	F4	29.937	.000
V2 onset	F3	25.791	.000
V2 onset	F1	21.631	.000
PO	F3	19.439	.000
PO	F4	19.363	.000
V1 offglide	F3	18.102	.000
V2 mid	F4	16.581	.000
V2 mid	F1	14.419	.000
V2 offglide	F3	12.665	.000
V1 onset	F1	13.662	.000
V2 onset	F2	10.836	.000
V2 mid	F3	10.239	.000
V1 offglide	F2	9.694	.000
V1 onset	F3	9.635	.000
PO	F2	9.093	.000
C release	F3	8.872	.000
V1 onset	F4	8.036	.000
PO	F1	7.012	.000
C release	F1	5.124	.001
V2 mid	F2	4.193	.003
V2 offglide	F2	3.850	.005
V1 offglide	F4	3.656	.006
C release	F4	3.185	.014
V1 onset	F2	2.903	.022
C release	F2	2.802	.025
V2 offglide	F1	2.142	.074 (n.s.)
V1 offglide	F1	1.514	.201 (n.s.)

The estimated centre frequencies of the first four formants for each sampling point were collated for statistical analysis. One method which has been shown to be effective in determining the most efficient parameters for distinguishing between speakers is the analysis of variance, in which the ratio of variance of speaker means to the mean within-speaker variation is calculated (the F-ratio) (Pruzansky & Mathews

1964; Wolf 1972; Nolan 1983; Rose 1999, 1997). The greater the magnitude of the F-ratio, then correspondingly, the greater between- to within-speaker variation can be expected. A series of univariate ANOVAs was performed to calculate the F-ratio for each formant at each of the seven sampling points. The sampling points with the highest F-ratios were deemed to represent the most promising parameters for distinguishing between speakers. The results in order of magnitude of the F-ratio are set out in Table 3.

The results indicate that the most efficient sampling point for distinguishing between speakers in Australian *okay* is F_4 at V_1 onset, with an F-ratio of 32.367. This is followed closely by F_4 at the V_1 offglide (F=29.937), while the next most efficient sampling points are F_4 at V_2 onset (F=25.791) and F_1 , also at V_2 onset (F=21.631). The magnitude of these F-ratios is sufficiently high to suggest that these sampling points are acceptable for distinguishing between speakers, although higher F-ratios have been found to occur in a range of other parameters which have not been considered here. For example, Wolf (1972: 2048) found "individual fundamental frequency parameters had the highest F ratio of all the parameters investigated" in his study, with F-ratios for F_4 ranging from as high as 84.9 down to 30.9. In Wolf's study, the only formant measurements taken were F_1 and F_2 for vowels /æ/, /a/ and the schwa /ə/, and F-ratios for these ranged from 46.6 (for F_2 of /æ/) down to 15.5 for F_1 of /æ/. The highest F-ratio in the present study falls at around the median result of Wolf's study, while the four highest F-ratios noted above for the present study all occur within the top two-thirds of Wolf's values.

A further comparison could be made with Nolan's (1983) study in which F-ratios were calculated for 15 speakers for F_1 , F_2 and F_3 of the two English liquids, /l/ and /r/. Nolan found that F_1 provided the highest F-ratios (F=216.9 for /r/ and F=77.8 for /l/). Although, as Nolan (1983: 102) notes, the high value for /r/ may be due in part to "an artefact of the formant extraction process", these values are still considerably higher than the F-ratios obtained from Australian *okay*, which compare more closely with Nolan's lowest F-ratios, which were recorded for the two lower formants of /l/ (for F_1 , F=17.7, and for F_2 , F=21.6). Nevertheless, Nolan (1983: 115) concludes that "Spectral information from initial allophones of /l/ and /r/...yield moderate identification rates...[and] are worth incorporating in a speaker identification scheme making use of segmental information." The comparability of the top 25% of F-ratios found in Australian *okay* (set out in Table 3) suggests that the formants at these sampling points are also worthy of incorporation in a forensic analysis, particularly as this data was recorded from natural speech events, rather than having been obtained from read-out speech, as was the case for both the Wolf and Nolan studies. (Greater within-speaker variation would be expected from natural speech than from read out speech, thus lowering the F-ratios.)

Just over 50% of the F-ratios were below 10, indicating that these parameters are the least efficient formants and sampling points in Australia *okay* for distinguishing between speakers. Nevertheless, with the exception of the two lowest F-ratios (for F_1 of the offglides of each of V_1 and V_2) they were still

statistically significant, and could be used. It should also be noted that the highest F-ratio for each formant was always found at voice onset of V_2 , that is, within the first two or three glottal pulses of V_2 .

Further analysis of the data in this study using a Bonferroni post hoc test for the analysis of variance, showed that an average of 8 out of a possible 15 between-speaker distinctions were found in each of these top 25% formant X sampling points. The highest number of between-speaker distinctions occurred in F_4 at V_2 onset, where 9 statistically significant differences between speakers were found. The more conservative Scheffé post hoc test (which may be preferable to use in a forensic analysis) indicated that on average, 7.3 distinctions were made in the top 25% of F-ratios, with 8 out of 15 speakers showing a significant difference for F_4 at V_2 onset.

One point which should be made is that the integrity of using the higher formants (and particularly F_4) in the context of telephone recordings is highly questionable, due to the bandpass limitations which affect the acoustic properties of the transmitted signal (Rose & Simmonds 1996). When this is taken into account, the actual sampling points which may prove useful in forensic analyses, where data has been gathered from recordings of telephone conversations, is further reduced.

5. CONCLUSION

The analysis of F-pattern variations of *okay* in natural conversation has shown there is greater between-speaker variation than within-speaker variation in the F-pattern of *okay* in Australian English, making this frequently occurring word potentially useful in forensic comparisons. Given the questionable reliability of F_1 in speech samples recorded over the telephone, it would appear that the most efficient formants and sampling points for measuring between-speaker differences are likely to be F_4 and F_3 at voice onset of the second vowel, while F_4 at both PO and V_1 offglide should also be useful. Additional measurements for F_4 at V_1 onset and midway through V_2 , and for F_3 at V_2 offglide may also be valuable in distinguishing between speakers. F_2 has not shown itself to be a particularly efficient parameter at any sampling point in *okay*. In directly recorded data (as opposed to data collected via telephone), the most efficient sampling point for distinguishing between speakers is unquestionably at voice onset of V_2 , where a significantly high F-ratio is obtained for all of the first four formants.

No forensic analysis should rely on F-pattern alone for determining likelihood ratios. While auditory analysis is also clearly important, ongoing research on the potential value of using the frequently occurring word, *okay*, in forensic investigations will consider other acoustic parameters, including fundamental frequency and duration, and will attempt some form of quantification of coarticulatory effects, such as the extent of "velar pinching" in V_1 triggered by the following consonant. In addition a survey will be made of intonational and stress patterns of each token, and how these relate to their discourse function. Forensic phonetics would

also benefit from similar studies of other high frequency words, such as *yeah*, *so*, *well* and *y'know*, as well as other discourse markers such as *oh*, *ah* and *um*, and these could be the focus of future research.

NOTES

This paper was first presented at the Eighth Australian International Conference on Speech Science and Technology, Canberra 5-7 December 2000. I would like to thank Phil Rose for his readiness to provide guidance and advice while undertaking this project, and the two anonymous conference reviewers for their very constructive comments.

REFERENCES

- Burridge, K. & J. Mulder. (1996) *English in Australia and New Zealand. An Introduction to Its History, Structure, and Use*. Melbourne: Oxford University Press.
- Condon, S. (1986) "The Discourse Functions of OK." *Semiotica* 60, 73-101.
- Giles, H. & N. Coupland. (1991) *Language Contexts and Consequences*. Milton Keynes: Open University Press.
- Harrington J., F.Cox & Z. Evans. (1997) "An Acoustic Phonetic Study of Broad, General, and Cultivated Australian English Vowels." *Aust. J. Linguistics*, 17, 155-184.
- Laver, J. (1994) *Principles of Phonetics*. Cambridge: Cambridge University Press.
- Merrit, M. (1984) "On the Use of OK in Service Encounters." In J. Baugh & J. Scherzer (eds.), *Language in Use: Readings in Sociolinguistics*. New Jersey: Prentice-Hall Inc.

- Mitchell, A.G. and A. Delbridge. (1965) *The Pronunciation of English in Australia*. Sydney: Angus & Robertson.
- Nolan, Francis. (1983) *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Pruzansky, S. & M.V. Mathews. (1964) "Talker-Recognition Procedure Based on Analysis of Variance". *J. Acoust. Soc. Am.* 36(11), 2041-2047.
- Rose, Philip J. (1999) "Long- and short-term within-speaker differences in the formants of Australian hello." *J. Internat. Phonetic Assoc.*
- Rose, Philip J. (1997) "Differences and Distinguishability in the Acoustic Characteristics of Hello in Voices of Similar-Sounding Speakers - A Forensic Phonetic Investigation". *Austr. Rev. Appl. Linguistics* 22(1), 1-42.
- Rose, Philip J. & Alison Simmonds (1996) "F-pattern variability in Disguise and Over the Telephone - Comparisons for Forensic Speaker Identification" In Paul McCormack & Alison Russell (eds.) *Proc. 6th Aust. Internat. Conf. Speech Science and Technology, Aus. Speec'96 Sci. & Tech. Assoc.* 121-126.
- Schegloff, E.A. (1986) "The routine as achievement." *Human Studies*: 9, 111-152.
- Schegloff, E.A. (1979) "Identification and Recognition in Telephone Conversation Openings." In G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology*. New York: Irvington.
- Schegloff, E.A. & H. Sacks. (1984) "Opening Up Closings." In Baugh & J. Scherzer (eds.), *Language in Use: Readings in Sociolinguistics*. New Jersey: Prentice-Hall Inc.
- Wolf, Jared J. (1972) "Efficient Acoustic Parameters for Speaker Recognition". *J. Acoust. Soc. Am.*, 51(6), 2044-2056.

Noise Logger Hire

1 Type 1 or Type 2 noise loggers available for hire.

2 Expecting noise levels below 35dB(A)? Don't take a chance with a Type 2 logger, use an iM3 logger with Type 1 accuracy down to 25dB(A).

3 Full hire and placement service available, you tell us where and we do the rest - call for a free quote.

4 Free data review at the end of each job, including hardcopy of daily noise level graphs.

5 Central metropolitan location in Ryde

6 Store Leq,0.1s for over a week to capture aircraft flyovers, train passing events etc.

Sales

Infobyte

The Infobyte iM3 Type 1 precision noise monitor is a unique instrument designed specifically for the task of long term noise monitoring in hot climates. Compare these features ...

16MB of nonvolatile storage providing secure space for over a year of data.

95dB dynamic range allows the use of a single range setting minimising setup errors.

Peak Overload Detector to ensure Leq results are reliable as required by AS1259.2-1990.

15dB(A) microphone self noise allows accurate results down to 25dB(A).

+65°C internal temperature rating combined with a custom engineered environmental case allows operation in hot Australian conditions.

iM3 software allows virtually any Leq or Ln to be calculated from the binary survey data and produces report quality graphs at the click of a button.

MICROTECH GEFELL

The Microtech Gefell range of microphones and accessories are sold by Infobyte. These high quality German microphones are cost effective replacements on most name brand sound level meters and analysers. Contact us for a price and specification for your application.

Call Geoff Veale on 9807-8786 to make your booking or enquiry

Infobyte Pty Ltd 19 Curtis Street, Ryde, NSW 2112
Ph. +61-2-9807-8786 E-mail: gveale@f1.net.au
<http://www.f1.net.au/users/infobyte>

Maintaining AAS Archives

With the Australian Acoustical Society having had its official beginnings in 1964 in both New South Wales and Victoria and a National (not Federal) body, it is now old enough to have accumulated an interesting history. An account of some of the activities of the AAS and its members was published in the December 2000 issue of *Acoustics Australia*.

At least some of the material for these articles depended on the AAS possessing archival records. Unhappily, such archival material was not as readily available as we would have liked. For example, in the articles describing the beginnings of the AAS in NSW and Victoria, there was insufficient information available at the time of writing to give a full and precise account of the earliest meetings, particularly those held in Sydney in 1964 during August and September. The finding early this year in the Victoria Division archives (currently held by the General Secretary) of the minutes of the meeting held in 1964 in Sydney on September 23 yielded some further interesting information.

As a result of these continuing findings, which have enabled a more precise recording of these early AAS events, the general secretary, David Watkins, suggested that I should contribute a brief note to *Acoustics Australia* on the kinds of material that should be retained in AAS archives. After discussions at the first Victoria Division committee meeting for 2001 it was considered that the following items were necessary:

1. Membership records, such as those in AAS Directories;
2. Division committee meeting agenda, and minutes: with the minutes being confirmed as true and correct, and including place, date and time of meeting, due constitution of the meeting as opened and closed, names of those present, apologies, substantive summary of inwards and outwards correspondence, financial statement, sub-committee and other reports received, and a record of all decisions duly taken, with, if required, a précis of arguments for and against, and a summary of the main points raised and discussed;
3. Acoustically-substantive précis reports of technical and other general meetings;

4. Official AAS publications (eg, Newsletters, Bulletins, Acoustics Australia, etc);
5. Conference and Symposium Proceedings, with a record of numbers registered, social events, etc;
6. Division AGM agenda, minutes (confirmed) and audited financial statements;
7. National Council meeting agenda, minutes (confirmed) and financial statements;
8. National AGM agenda, minutes (confirmed) and audited financial statements; and
9. Official administrative documents, including constitution, code of ethics, administrative procedures, etc.

What then is needed is an established, recognised and ordered procedure by which the National Council and State Divisions regularly (e.g. annually) transfer their historical records (e.g. those more than two or three years old) into the relevant archive. At present, some re-organisation of the Victoria Division archives is being undertaken.

Louis Fovvy

Errors in Noise Modelling

I am presently working on a project as part of my studies investigating errors in noise modelling. I am interested in hearing about any experiences and observations associated with this topic. For example, if you have been using some noise modelling software, even if it is a program developed by yourself, I am interested in hearing about any problems you may have had or any observations you may have made that are relevant to my research.

I am calling for this information as there is limited literature exploring this topic and I wish to draw on the experience and knowledge of others working in the area of noise modelling. Naturally, I would acknowledge your contributions in my project. So if you take some time to draft up a "dot point" list or something more detailed and e-mail it to me it would be gratefully appreciated. As I have to meet the project's deadlines, I look forward to receiving your comments by 16 May 2001. I would love to hear about any references you consider to be relevant to this project as well.

Namiko.Ranasinghe@cnviro.n.wa.gov.au,
tel: (08) 9222 7141 or (08) 9389 9334

Traffic noise prediction validation.

A number of studies have recently been undertaken in Australia into the accuracy of various traffic noise prediction models, including studies funded by VicRoads, the NSW Roads & Traffic Authority and Queensland Main Roads. It is clear that further work is likely and Austroads is currently developing guidelines to assist researchers in the field. Austroads is the association of Australian and New Zealand road transport and traffic authorities. One of Austroads' aims is to develop and promote national practices among the member organisations and national and international bodies.

Marshall Day Acoustics Pty Ltd has been commissioned by Austroads to prepare guidelines for the assessment of validation studies of traffic noise prediction models. Key objectives for the project are: to provide recommendations, on a technical basis, for a preferred traffic noise model for use in Australia and New Zealand to develop a preferred methodology for validation studies of road traffic noise models to prepare guidelines for undertaking, reporting and assessing traffic noise validation studies.

The first two tasks have now been completed, and feedback on the draft recommendations is now being solicited from the Austroads member organisations, members of the Australian Acoustical Society and members of the Australian Association of Acoustical Consultants, as well as selected individuals.

Copies of the draft recommendations are available from
neilhuybregt@marshallday.com.au.

New Members

NSW

Graduate: Jeffrey Parnell

QLD

Member: Terry Anderson,
Mark Batstone,
Jackson Yu

Future Meetings

Acoustics 2001

This conference, organised by the Australian Acoustical Society (AAS), is being held from 21 to 23 November in Canberra, the seat of Federal Parliament. It is therefore appropriate to take noise and vibration policy as a theme for the conference. Recently there have been many changes and revisions of policies and this conference provides an opportunity for discussion of the various issues along with other aspects of acoustics.

The keynote speaker at the opening will provide an overview of the noise and vibration policies and provide a personal view of the way forward. For each of the sessions addressing various aspects of noise and vibration policy the session leader will be invited to summarise the current situation. Contributed papers related to the theme and to other topics on sound and vibration are invited. Each paper will be allocated 15-20 minutes. Papers related to the themes will be allocated in the sessions indicated in the preliminary program. Papers on other areas of acoustics will be in the parallel sessions.

All sessions, the technical exhibition and the social functions will all be held at RYDGES CANBERRA. All registrants are encouraged to stay in the conference hotel and a special room rate has been negotiated. This conference will combine contributed papers, technical presentations, awards and a range of social activities included in the delegate registration fee such as welcome buffet, conference dinner and farewell lunch.

Further information: Acoustics 2001, Aust Defence Force Academy, Canberra, ACT 2600, tel:02 6268 8241 (0402 240009), fax:02 6268 8276 m.burgess@adfa.edu.au and www.users.bigpond.com/Acoustics

ICA

The 17th International Congress On Acoustics (ICA) will be held in Rome, Italy, 2-7 September 2001. The congress will be held at the Engineering Departments in San Pietro in Vincoli, next to the Colosseum, in the centre of Rome. The ICA is the only congress devoted to all aspects of acoustics, where any acoustician should find him/herself at home. The Congress has also been, since the very first one in its history, a moment where - not only people - but organizations, institutions and groups, do meet.

Further information from <http://www.ica2001.it/> or Secretariat, ICA 2001, Dipartimento di Energetica, University of Rome "La Sapienza", Via A. Scarpa, 14, 00161 Rome, Italy fax: +39 06 4976 6932, ica2001@uniroma1.it

ISMA 2001

The Interuniversity Center of Acoustics and Musical Research (CIARM) and the Catgut Acoustical Society (CAS) are pleased to present a joint International Symposium on Musical Acoustics (ISMA). This will be a satellite symposium of the 17th ICA and will be held September 10-14 in Perugia (Italy): the beautiful chief town of the Umbria region. In keeping with previous conferences in this series, ISMA 2001 will bring together international leaders in the musical acoustics field.

Further information from <http://www.cini.vc.cnr.it/ISMA2001> or Musical Acoustics Laboratory, Fondazione Scuola di San Giorgio - CNR, Isola di San Giorgio Maggiore, I-30124, Venezia, Italy, Fax: +39 041 5208135, isma2001@cini.vc.cnr.it

Internoise 2001

Internoise 2001, the 30th International Congress on Noise Control Engineering to be sponsored by I-INCE, the International Institute of Noise Control Engineering, will be held in The Hague, The Netherlands (or Holland), on 2001 August 27-30. The theme of Internoise 2001 will be Costs & Benefits of Noise Control and a great number of papers will be presented.

Further information from <http://www.internoise2001.tudelft.nl>, or Congress Secretariat, P.O. Box 1067, NL-2600 BB Delft, The Netherlands, fax +31 15 2625403, secretary@internoise2001.tudelft.nl

ICSV8

The Eighth International Congress on Acoustics and Vibration sponsored by IIAV, the International Institute of Acoustics and Vibration, will be held in the Hong Kong Special Administrative Region, China, from 2 to 6 July 2001.

Further information from <http://www.iiav.org> or ICSV8, Dept Mechanical Engineering, Hong Kong Polytechnic University, Hungghom, Hong Kong, China, fax: +852 2365 4703, mmicv8@polyu.edu.hk

Acoustics Workshop

This workshop will be presented by National Voice Centre, The University of Sydney, Tuesday October 9, 2001 from 9am - 5pm. The main speakers will be Professors Johan Sundberg & Neville Fletcher with presentations by A/Professor Jennifer Oates, A/Professor Pamela Davis, Denis Cabrera, Jennifer Barnes, Debbie Phylland.

It will be a workshop for all interested in acoustics, voice or singing, speech pathology, linguistics and speech science. The course will be a scientific overview and practical demonstration of up-to-date methods for acoustic analysis of the speaking and singing voice. Issues in the acoustics of the human voice including singer's formant and related physiology and measurement of wind instrument performance will also be presented. Factors in voice recording including the effects of different recording environments and microphone placement will also be presented.

The format will be lectures and demonstrations in the morning and "hands-on" acoustic analysis in the afternoon of material supplied by the presenters and participants will work in small groups under the guidance of the presenters. Registrants should have had some experience in using acoustic analysis software. The course will follow a 3.5 day workshop on "The Science of Voice and Singing", with Professor Johan Sundberg and Janice Chapman

Further information from National Voice Centre, The University of Sydney, NSW 2006 tel 02 9351 5352, fax 02 9351 5351, voice@chs.usyd.edu.au

Acoustics and music: theory and applications

This conference will be held on Koukouaries, Skiathos Island, Greece from September 26-30, 2001. It is sponsored by the World Scientific and Engineering Society (WSES) — Sector of Acoustics and Music, Sector of Oceanic Engineering and Technical Committee of Signal Processing. The range of topics to be covered at the conference are extensive and range through all the areas of acoustics.

The venue for the conference is a beautiful, wooded Greek island which was the birthplace of many famous Artists of Greece, like Alexander Papadiamantis and Alexander Moraitidis. Skiathos is known as the Island of Poets. During the Conference Official Dinner, a small festival / concert will take place and conference registrants are invited to register their interest in participation in the concert.

The call for papers has been distributed and the conference details are available from <http://www.worldses.org/wses/conferences/skiathos/amta/>

Responsive systems for active vibration control

This course sponsored by the NATO Advanced Study Institute will be held in Brussels September 10-19, 2001. Over the 10 days a comprehensive coverage of active vibration control in its multidisciplinary aspects will be provided. The course is intended for PhD holders or advanced PhD students. Some background is assumed in structural dynamics, acoustics, control and signal processing.

The course is free of charge for citizens from countries belonging to NATO, Euro-Atlantic Partnership Council (former USSR) and Mediterranean Dialogue. A nominal registration fee of 200 Euros applies to all other participants.

Further information: ASI-NATO@ulb.ac.be, <http://www.ulb.ac.be/scmero>

Meeting Reports

NSW Division

On Wednesday 22 November 2000, the NSW division of the AAS were treated to a talk by Assoc Prof Fergus Fricke from the University of Sydney. Fergus talked about the research that has been carried out at the University since the first architectural science department was set up in 1954 and in particular his work since joining the Department in 1974. The evening lecture was held at NAL and was followed by an informal gathering with drinks and nibbles.

One of the first PhD's awarded in the department, during the early 1970's, was for research into aircraft noise around Sydney Airport by Carolyn Mather. This research became the basis of Australian Standard AS 2021 Acoustics - Aircraft noise intrusion building siting and construction. Current research work being carried out at the department includes:

Neural network analysis

Attenuation of sound through ventilation openings by active and passive attenuators;
Harmonic form in resonating sound art;
Just Noticeable Differences (JND's) in frequency and duration of sounds may be useful as a measure of acoustic quality;
Sound diffusion - did Lord Rayleigh get it right?;

Preferred small room characteristics for music listening and performance;

Characteristic of speech delivered at different intensities;

Reducing feedback in audio systems; and
Developing a spider (not funnel webs!) free loudspeaker.

Fergus did spend some time on neural network analysis. Traditionally, designers have used either scientific methods or their experience and knowledge of previous solutions to solve acoustical problems. Scientific methods have generally been limited to modelling, such as finite element methods or ray tracing, (which have their limitations), or simplified equations that have limited accuracy. Scientific methods generally become too complicated to use when the number of defining variables exceeds six. The designer then has to rely to a large extent on experience. Approximately a dozen geometric variables have been identified in concert halls around the world as affecting the quality of the acoustics.

Fergus used a simplified example of four parameters in the design of concert halls to explain how a neural system works. The length, height, width and volume of the hall are individual weighted and then summed to give an Acoustical Quality Index (AQI). The system is subjected to feedback to minimise the error. Most analysis systems assume linear relationships, which are rarely true in practice. Fergus explained that with two stages of weightings the neural network system can handle non-linear relationships and the result moves much closer to the peak of minimum error. Neural networks have also been used with some success in the prediction of sound transmission loss of cavity wall constructions using inputs such as surface density, cavity width, cavity insulation thickness and flow resistivity etc. In concluding his talk, Fergus identified areas of acoustics where work is desperately needed. They include:

Acoustic provisions of the Building Code of Australia - some areas of the BCA, such as the inter-tenancy wall construction, are woefully inadequate, while others such as the STC requirements for waste pipes are just simply wrong.

Effectiveness of Sydney Aircraft Noise Insulation Project (SANIP);

Research into adaptation to the noise environment - if we are able to acclimatise to a noisy environment, why don't we get used to our clock radios and continue sleeping?

Additionally, Fergus identified some items on his personal acoustic "wish list". This includes an acoustic "diode" to control the direction of sound propagation. He also gave an impassioned plea for the acoustic community to lobby the government in general, and the Minister of Education in particular, to protect the threatened remaining pockets of acoustic research in our teaching

facilities. Without these departments a whole generation of acoustic skills could be lost! With that in mind Fergus was given a well deserved round of applause and the conversation continued over much needed food and drink.

Ken Scannell and Matthew Harrison

Victoria Division

The final Victoria Division meeting for 2000, at which 21 AAS members and friends were present, was held on Nov 29 at the Malvern Valley Golf and Function Centre. This took the form of a dinner meeting, with Ms Lisbet Bruel, niece of Per Bruel (formerly of Bruel & Kjaer), as guest speaker. She spoke in general rather than technical terms about his more recent activities, mostly since his leaving B&K. At the conclusion, Charles Don on behalf of all present thanked her for a most interesting talk.

The first technical meeting for 2001, at which 23 AAS members and friends were present, took the form of a site visit held on Feb 28 at the CSIRO centre at Highett, followed by a grilled-chops-and-steak and sausage-sizzle barbecue.

John Davy, CSIRO Acoustic Services Manager, described that section of the CSIRO responsible for acoustical work, referred to research work currently in hand and listed the various items of test equipment available for the relevant tests.

This acoustical research is carried out within the CSIRO Built Environment Sector, in the Thermal and Fluids Engineering Department (or core capability) of the Building, Construction and Engineering Division, under a supervisory staff of four at Highett, Melbourne, and two at North Ryde, Sydney. Major current projects include: active noise attenuation for windows (two projects — one using piezo-electric film on single glazing, the other a loudspeaker within the cavity of double-glazed windows), sound insulation systems using concrete panels, quietening kitchen range hoods (within a model kitchen), and a workshop on acoustical building regulations. Other current work includes: research into bubble acoustics (to detect the size and numbers of bubbles) the acoustical detection of leaks in water pipes, ceiling fan safety, the theoretical prediction of sound insulation, and quietening pulse combustion.

Laboratory equipment at North Ryde includes a two-microphone impedance tube, apparatus for measuring the resistance to airflow, and a two-reverberation room installation with 3.6 x 3.0 m wall opening, ceiling frame for mounting ceiling tiles, fork lift and mobile crane access, rotating microphone booms, and tapping machine and stand.

The laboratory at Hightt includes a four-reverberation room installation with 3.68 x 3.22 m wall opening, 3.68 x 3.22 m floor/ceiling opening and ceiling frames. One of the upper rooms contains 150 and 125 mm thick concrete slabs for impact tests, a tapping machine for impact insulation tests and a tapping machine stand for wall impact tests. An anechoic room with an 80 m³ space, for testing above 55 Hz, can be changed to a hemi-anechoic room by inserting a sound-reflecting floor.

These facilities and associated equipment enable a range of tests of the performance of building materials and sound output of sources. In addition, work outside the laboratory can be undertaken, such as before and after site measurements for the Sydney Aircraft Noise Insulation Project.

Louis Fowry

NOISE CON 2000 CD

The successful NoiseCon 2000 conference was held in California, Dec 3-5 2000 in conjunction with the 140th meeting of the Acoustical Society of America. This was the 17th in a series of national conferences on noise control engineering that began in the USA in 1973.

The CD-ROM containing all the proceedings for the 2000 conference also includes the proceedings of NoiseCon 96, 97, 98 and the 1998 Sound Quality Symposium. There was no NoiseCon in 1999. This comprehensive CD ROM is available from Bookmasters International USA, fax 1 419 281 6883, order@bookmaster.com for US\$75.

Standards

Revised Version of AS 2107

The recent release of AS/NZ 2107:2000 supersedes the 1987 version of the standard on 'Recommended design sound levels and reverberation times for building interiors'. This standard is intended for use in assessing the acoustic performance of buildings and building interiors. The revised version of this very useful document incorporates some changes in the recommended levels as well as rationalisation of the types of spaces. Further information from Standards Australia www.standards.com.au

Quality Guidelines for Building

Australia's building and construction industry has received a boost with the launch of the a new set of global construction guidelines that detail how to meet international quality standards. The latest International Standard

for Quality Management places an emphasis on identifying and managing business processes with a focus on consistent performance and continual improvement. According to the Chief Executive of Standards Australia, Ross Wraight, the new guide has been tailored for the construction industry and provides easy to understand information on the new edition of ISO 9001:2000. The guide entitled HB90.3 'The Construction Industry - Guide to ISO 9001:2000' sets out requirements for quality management systems which enable suppliers and project leaders to demonstrate their capability to design, supply for and construct projects.

The Australian Standard, April 2001

BCA On Line

A new on line version of the Building Code of Australia (BCA) had been developed by Standards Australia and is a subscription service which enable users to access the regulatory code. Features of BCA on line include being fully searchable for the relevant clause, word to topics; optimised on-screen viewing with documents laid out in an easy-to-use format; hyperlinks between the BCA volumes as well as links to referenced Standards and a print option. Check out the free demonstration from <http://www.standards.com.au/>

The Australian Standard, April 2001

Participation in Committee

The Australian Acoustical Society is represented on a number of Standards Australia Committees working in fields of acoustics related to the committees work. It is important that the Society continues to be represented on Standards Committees by members and that they provide feed-back on the activities of the Committees.

A vacancy has arisen in the AV/3/3 Committee - Audiology. Any member who would like to nominate for this committee is invited to provide a brief description of their experience in this particular field and return it by May 31st 2001 to the General Secretary, Australian Acoustical Society, PO Box 4004, East Burwood, Victoria 3151, watkinsd@melbpc.org.au

FASTS

The Australian Acoustical Society, AAS, has been a member of The Federation of Scientific and Technological Societies, FASTS almost since the inception of the Federation in 1985. Over the years Council of AAS has regularly considered the value of

continued membership and is seeking advice and feedback from the membership to assist with its deliberations.

This timing is appropriate in view of the recently well advertised policies from the Government which are intended to increase the funding for science and technology in Australia. Also the AAS has been within the Physical Science Board Grouping and it is now proposed that the AAS should be placed with the Australian Society of Biomaterials and the Clean Air Society of Australia and NZ in a newly-formed board entitled "Technology". This means that the AAS may have the opportunity for a greater participation in the activities of the FASTS Board but it also means that efforts need to be made to make use of this opportunity.

Information about the various activities of FASTS can be found from www.FASTS.org and all members of the AAS are invited to provide comments on the value of the continued membership of FASTS and advise on any issues which may be appropriate for presentation to FASTS for action. Please send all comments to Marion Burgess, m.burgess@adfa.edu.au or Acoustics and Vibration Unit, ADFA, Canberra ACT 2600.

News

NOHSC Move

The National Occupational Health and Safety Commission. (NOHSC) will occupy its new office in Canberra on 1 May 2001. The new office will be at 25 Constitution Ave and full contact details will be available from www.nohsc.gov.au.

Awards for Challis

Louis Challis and Associates, consulting acoustical and environmental engineers in Sydney has received two Highly Commended Engineering Excellence Awards from the Institution of Engineers Australia, Sydney Division. The separate categories were 'Welfare, Health and Safety' and 'Innovation and Invention'. The project was the national Rail Corporation Locomotive Test Cell at Spotswood Victoria. It is unusual as it is the first facility of its type to be designed and successfully constructed in Australia. The test cell is designed to operate at the end of the day or night and not disturb the residential neighbours living on both sides of the cell.

The Alliance

MB & KJ Davidson Pty. Ltd. has announced the exciting news of OROS, m+p international inc., and M&M Corporation join The Modal Shop, Larson Davis, GRAS, Brüel Bertrand & Johnson Acoustics, and Brüel Acoustics in the formation of a new and potent force known as The Alliance.

Following the successes of "The Acoustic Alliance" over the past year in providing measurement solutions with "Total Customer Satisfaction", the scope of The Alliance is now vastly expanded with very powerful vibration measurement and analysis techniques available from the newest team affiliates.

"By sharing the expert knowledge found in each group" said Maurice Mergey, President of M&M Corporation, "we are developing application solutions that will give our extensive customer base a choice of high quality systems providing them with the innovative and intuitive tools they need to improve and develop their products". Thomas Lagö, President of Larson Davis Inc. added, "The Acoustic Alliance philosophy of "Empowering our Customers' Success" has enabled us to focus on the real issues of developing application solutions that our customers need in order to increase their own performance, both personally and organisationally". Key elements are full integration of member group products, ease of contact and the most complete range of solutions, resulting in what is truly a "One-Stop-Shop".

For further information: M.B. & K.J. DAVIDSON PTY. LTD, 1-3 Lakewood Blvd, Braeside Vic 3195, Tel 03 9580 4366 Fax 03 9580 6499, www.davidson.com.au or to info@thealliance.com

Sound Check

Brüel & Kjaer announces that an alliance has been formed with LISTEN Inc., based in Boston, USA. Under the agreement, Brüel & Kjaer will market the SoundCheck Electroacoustic Test System throughout the world.

LISTEN Inc. was formed by Steve Temme who has developed this highly versatile, software-based system for the production line testing of loudspeakers, microphones, hearing aids, telephones and other acoustic transducers. No special hardware is necessary as the system operates using a standard professional sound card installed in a normal PC. SoundCheck is easily programmed and is delivered with a range of options that automate testing.

SoundCheck has been optimised for fast production testing, and performs very rapid

frequency response and distortion tests, typically in less than 2 to 4 seconds. The system evaluates frequency response characteristics of the transducer using swept sine or noise-based tests. Harmonic distortion and special distortion parameters such as Rub & Buzz (an evaluation tool for poor speaker/enclosure installations) can also be measured. PASS/FAIL criteria are easily set or changed and the PASS/FAIL status is shown clearly after each test run, showing which test parameters are within the defined tolerances. The system also provides extensive tools for statistical evaluation of the production output. NASA and the U.S. Navy have chosen SoundCheck for the evaluation of their communications systems. A number of "high-end" audio companies and two of the world's major manufacturers of cellular phones use SoundCheck, for a variety of applications.

For further information: Brüel & Kjaer Australia, Syd 02 9450 2066, Melb 03 9370 7666, Bris 07 3252 5700, Perth 08 9381 2705, bk@spectris.com.au, www.bksv.com

New Agent for LMS

LMS International is a company that specialises in making laboratory and mobile testing systems, multidisciplinary virtual prototyping software such as Sysnoise and Raynoise, and providing engineering services in vibration, acoustics and durability to various industries. LMS has appointed AVESA Pty Ltd as its Australian agent effective from January, 2001. Based in Melbourne, Victoria, AVESA will be responsible for sales, technical support, product training, as well as consulting services to existing and new customers Australia wide.

Further information: AVESA Pty Ltd, tel 03 9584 6185, fax 03 9515 3495 and admin@avesa.com.au.

Vitech Changes

Vitech provides instrumentation and condition monitoring systems and has recently changed its name to Vitech Asia Pacific to reflect its new responsibilities throughout the region. It has also opened a new Melbourne office in Glen Waverley and a new web site at www.vitech-apac.com. From this site you can get further information on the range of products including the iLearn interactive CD based series of products designed to deliver step by step tuition and hands on experience in condition monitoring and vibration analysis

New Products

ARL Noise Logger

Acoustic Research Laboratories are pleased to announce the release of the Mark 2 version of their EL-315 (Type 2) and EL-316 (Type 1) environmental noise logger. Since the release of the Mark 1 version in 1999, users of the new ARL logger have been making suggestions about additional features they would like to see. One of the advantages of Australian production is that many of these features can - and now have been - incorporated into the EL-315/6 Mark 2.

These features include:-

- Overload indication now activated on statistical intervals and Leq's on LCD and host software.
- Leq's now broken up into more manageable file sizes for use with Microsoft Excel or similar.
- Trigger function now available that allows Leq's and/or tape recorder to activate at preset level.
- Screw in, fixed post microphone now a standard feature.

Further information or free demonstration, contact Acoustic Research Laboratories on 02 9484 0800, your local branch of ARL or www.acousticresearch.com.au

OROS 4-channel Analyzer

OR24 represents the latest product in OROS' expanding range of PC-based noise and vibration analyzers. OR24 extends the existing OR25 range of portable, robust, multichannel analyzers (2 to 16 channels) by adding a 4-channel instrument at half the size. OR24 is a complementary solution to the OR25 PC-Pack II for engineers seeking a no-compromise, 4 channel, and super-portable instrument. The main characteristics of OR24 are: 4 input channels with signal conditioning for microphones and accelerometers; DSP powered real-time analysis up to 20kHz; fast PC-card connection to any laptop computer operating under Windows environment (95/98/NT/2000); comprehensive analysis capabilities; FFT, 1/n octave, order tracking, recorder; and only 2kg in an A5 footprint and extremely robust. Further information from M.B. & K.J. DAVIDSON PTY. LTD, 1-3 Lakewood Blvd, Braeside Vic 3195, Tel (03) 9580 4366 Fax (03) 9580 6499, www.davidson.com.au

BRÜEL & KJÆR FFT software for 2260

With FFT Software BZ7208 installed, Brüel & Kjær's widely-used 2260 Investigator(tm) is transformed into a single-channel FFT analyzer. It is suitable for measuring continuous and transient signals (for both sound and vibration) in environmental and industrial applications. A flexible internal trigger is provided, as well as an external trigger for transients. To evaluate the total content in noise, the software can identify tones and calculate their audibility.

For vibration measurements 2260 Investigator uses a DeltaTron Adaptor ZG0423 that accepts DeltaTron accelerometers and, via Charge Converter Type 2647, also accommodates charge accelerometers. Most importantly, all the functions you need in the field (e.g., frequency span, zoom and cursors) are easily activated from 2260 Investigator's front panel.

The combination of sound and vibration capability draws on Brüel & Kjær's long experience in FFT for sound and vibration applications.

Laser Doppler Vibrometer

Brüel & Kjær has introduced a new non-contact vibration transducer for use in applications where it is impossible or undesirable to mount a conventional vibration transducer onto a vibrating object. The introduction of this exciting new product follows the recent signing of an exclusive sales distribution agreement between Brüel & Kjær and the UK based company Omnetron. Under this agreement, all Omnetron products will be sold via the worldwide Brüel & Kjær sales distribution network.

Based on a Michelson interferometer, the new Brüel & Kjær Laser Doppler Vibrometer, Type 8329, offers an alternative to microphones or accelerometers in applications where, for example, extreme temperatures may preclude the use of conventional sensors. Other applications include measurement of vibration on lightweight, small, delicate and soft objects where an accelerometer would cause mass loading effects. Further applications include impact measurements, relative vibration measurements (e.g., on board ships, aircraft and cars), railway track and track bed vibration monitoring.

Capable of measuring vibration in any direction, Type 8329 features a velocity range up to 425mm/s, a frequency range from <0.1Hz to 25kHz and a dynamic range of 73.5dB over full bandwidth. Measurements are possible from as close as 0.4m and up to 25m away, usually without the need for any surface treatment or retro-reflective tape. For

measurements at distances greater than 25m, retro-reflective tape can be used.

Type 8329 is extremely quick and simple to set up. A built-in LED bar-graph confirms that the laser is adequately focused. A second LED bar-graph indicates the approximate measured velocity level. It connects via a simple BNC cable with any Brüel & Kjær sound and vibration analysis system such as PULSE. Battery or mains-operated, the portable and compact Type 8329 offers integrated optics and electronics and features a Class II laser for safe operation. Despite the precision nature of the optical and electronic components, the unit is sufficiently robust for normal laboratory and field use.

Sound Intensity Calibrator

Brüel & Kjær introduces the Sound Intensity Calibrator Type 4297 which enables instruments which measure sound intensity to be accurately calibrated. This new calibrator is used for on-site Sound Pressure Calibration (at 251.2Hz, Type 1 IEC 60942) and Pressure-Residual Intensity Index Verification.

The important, unique feature is that there is no longer any need to dismantle the sound intensity probe. The calibrator is optimised for use with Brüel & Kjær's 2260E Investigator sound intensity system for phase enhancement, but it can also be used with sound intensity analysis systems based on PULSE. Type 4297 is a complete sound intensity calibrator in one compact, portable unit with built-in sine and broadband sound sources. Type 4297 fulfils IEC 61043 and is intended for use with Brüel & Kjær Sound Intensity Probes Types 3583, 3584, 3595 and 3599 (or earlier Types 3545 or 3548) with Sound Intensity Microphone Pair Type 4197 (or earlier Type 4181).

For calibration of sound pressure sensitivity, the microphones are both positioned in the calibration chamber and are therefore exposed to exactly the same sound pressure (amplitude and phase). The broadband sound source is provided for measurement of the pressure-residual intensity index spectrum and this is also used to assess the accuracy of sound intensity measurements. A calibration chart is supplied which states the levels that should be detected during calibration. A barometer is not needed because an accurate feedback system holds the sound pressure level at a constant value. Type 4297.

Miniature Triaxial Accelerometer.

Brüel & Kjær has launched a new miniature triaxial ISOTRON accelerometer, ENDEVCO Model 65. High sensitivity and high resolution distinguish Model 65 triaxial

accelerometer from comparable products. Housed in a small welded titanium cube measuring 10 x 10 x 10mm, model 65 weighs just 5 grams and is ideal for structural analysis applications.

Shockproof and overload protected, Model 65 delivers excellent frequency response for both amplitude and phase to provide users with a triaxial accelerometer that excels in structural and component testing, drop tests and general laboratory vibration work. Its small size enables test engineers or technicians to measure the accelerations of three orthogonal axes of vibration simultaneously on lightweight structures.

Interface to Model 65 is via a single Microtech 4-pin connector with three BNC connectors provided at the instrumentation end. Model 65 is supplied complete with temporary retro-wax adhesive and a 3m cable. For further information: Brüel & Kjær Australia, Syd 02 9450 2066, Melb 03 9370 7666, Bris 07 3252 5700, Perth 08 9381 2705, bj@spectris.com.au, www.bkscv.com

CSR SoundScreen

Bradford insulation has introduced a new product into the market to address the growing problem of sound transfer between rooms and floors in homes - Rockwool SoundScreen. This is an acoustically optimised insulation specially developed for use in the internal walls and between floors to reduce noise transfer, allowing you to create quiet zones where your client wants them. For example; a wall consisting of 10 mm Gyprock either side of 70 mm timber studs with Rockwool SoundScreen has an Rw of 41. A further increase to Rw of 44 can be achieved by substituting the heavier 0mm Gyprock Soundchek.

For further information: Bradford Insulation tel 1800 023 380, www.csr.com.au/bradford

People

Neil Gross has recently been appointed Managing Director of Wilkinson Murray. This change has been a reflection of the work undertaken by Neil over recent years, mainly concerned with the day to day management of the consultancy business. Contact details: tel: 02 9437 4611 or neil@w.mpl.com.au

Ian Jones has been appointed as the Business Development Director for Vipac's Building Technology Team, servicing the Australian and overseas market. This newly created position is designed to assist architects, developers and builders in the many aspects of Vipac operations. Contact details: 03 9647 9700 or melbourne@vipac.com.au.

Book Review

Science of Percussion Instruments

Thomas D Rossing.

World Scientific Publ Co, hardcover ISBN 981-02-4158-5, 648 Whitehorse Rd, Mitcham 3132, Australia, tel 03 9219 777, fax 03 9210 7788, Price including GST A\$52.91.

The family of instruments played by the percussionists in a symphony orchestra is large, diverse and interesting. Despite the name of the section, even their manner of playing is diverse: there are instruments that are shaken (e.g. wind machine), scraped (guiro), rubbed (glass armonica) and even blown (whistles) and bowed (musical saw).

The attraction of this book is that it shows that their science of percussion instruments is diverse and interesting as well. Here are some of my favourite examples: the non-linear behaviour in cymbals, tam-tams and gongs that gives rise to period multiplication and chaotic behaviour, and to the interesting time variation in the spectral envelopes of these

instruments. Or the subtle tuning of the partials of bells and tuned drums, which combine to produce notes whose pitches often do not correspond to the frequency of the lowest partial. Or the complexities of tuning and locating the different pitched areas in the pan of Caribbean steel drums. Or the ancient two-tone bells of China, whose pitch depends upon the point of striking.

Tom Rossing is the author of "The Science of Sound" and co-author (with Neville Fletcher) of "The Physics of Musical Instruments", as well as a large number of research papers in musical acoustics. His laboratory specialises in hologram interferometry, and he has used this and other techniques to study many types of percussion instruments. He has also won awards for science education. He is superbly qualified to write about this subject.

He has written a book that can be understood by a musician who is keen to understand this family, but who has little technical background. The use of equations is kept to an absolute minimum. Nevertheless, most of the important physics is retained, so the physicist or engineer will not be disappointed. The references at the end of each chapter include research papers and more technical books but also refer to the literature of music.

Some of the chapters contain brief sections explaining some relevant general principles in physics (vibrations of strings, bars and membranes). There is also a brief introduction to sound and hearing. These sections are called interludes and inserted when needed, so as to avoid having an extended theoretical section to begin the book. I should have preferred to see more of these, especially in the later sections: why not an interlude on non-linear behaviour, for example? But length and readership must be considered, too. The book is copiously illustrated with drawings and photographs, which have been kept small in size to keep the length and price of the book quite reasonable.

For a reader interested in music, this is a good read. It is certainly not beyond a non-technical reader with a little patience, but it is perhaps more fun for those with a background in physics or engineering.

Joe Wolfe.

Joe researches music acoustics in the School of Physics, UNSW. He has occasionally played orchestral percussion.



Australian Hearing
National Acoustic Laboratories

ACOUSTIC & NOISE SPECIALISTS Superb Anechoic and Reverberant Test Facilities Servicing:

- Transmission, Sound Power and Absorption testing
- General Acoustic Testing
- Comprehensive Analysis of Sound and Vibration
- Measurement and Control of Occupational Noise
- Electro-Acoustic Calibration • Vibrational Analysis

Experts in Noise Management and other Services - Including:

- Measurement and Control of Occupational Noise
- Reference and Monitoring Audiometry
- Residential and Environmental Noise
- Education and Training • Acoustic Research

126 Greville Street, Chatswood, NSW 2067
Phone: (02) 9412 6800

National Acoustic Laboratories is a Division of
Australian Hearing Services
a Commonwealth Government Authority



NOISE CONTROL
AUSTRALIA PTY. LTD.

ABN 11 076 615 639

Committed to Excellence in
Design, Manufacture & Installation of
Acoustic Enclosures, Acoustic Doors,
Acoustic Louvres & Attenuators

SUPPLIERS OF EQUIPMENT FOR:

PROJECT: King Street Wharf
CLIENT: Allstaff A/C

70 TENNYSON ROAD
MORTLAKE NSW 2137

Tel: 9743 2413
Fax: 9743 2959

A Sustaining Member of the Australian Acoustical Society

ACOUSTICS 2001

Noise and Vibration Policy - The Way Forward?



**Australian Acoustical Society Annual Conference
21 to 23 November 2001 CANBERRA, ACT**

The annual conference, organised by the Australian Acoustical Society (AAS), is being held in Canberra, the seat of Federal Parliament. It is therefore appropriate to take noise and vibration policy as a theme for the conference. Recently there have been many changes and revisions of policies and this conference provides an opportunity for discussion of the various issues along with other aspects of acoustics.

This conference will combine contributed papers, technical presentations, awards and a range of social activities included in the delegate registration fee such as welcome buffet, conference dinner and farewell lunch.

CONFERENCE SESSIONS

A key representative from Government has been invited as the Opening Speaker at the conference welcome buffet on Wednesday evening. The sessions will be held on Thursday from 0800 to 1800 followed by the Dinner and on Friday from 0800 to 1400.

The conference will include **special sessions** on aspects of noise and vibration policy and the session leader has invited presentations on relevant areas:

Occupational noise	Environmental noise	Noise in buildings
Transportation noise	Vibration control	Airport/Aircraft noise

Contributed Papers related to the theme and to other topics on sound and vibration are invited. Peer review is available for all papers.

Technical Exhibition will be open from breakfast each day for the duration of the Conference for manufacturers and suppliers to show their latest products. Those interested in exhibiting should contact conference organisers for details.

Technical Tours Wed afternoon: a 'behind the scenes' tour at Screen Sound Australia, the National Screen and Sound Archive, focussing on the techniques used for audio preservation. Friday afternoon: a special tour of Parliament House focussing on the diverse acoustic requirements.

Social Activities throughout the conference will all be held at Rydges. The welcome buffet, conference dinner and farewell lunch as well as breakfast, lunch and tea breaks each day are included in the registration fee. They will give plenty of opportunities for the delegates to discuss issues that have been raised in the sessions as well as to renew friendships and establish contacts for the future.

Standard Registration \$450. Discounts: Early registration \$50, Student registration \$100, AAS Member \$50. These rates include the proceedings, welcome buffet, conference dinner and farewell lunch plus breakfast, lunch and tea breaks.

Accompanying Members program includes the welcome buffet, conference dinner and farewell lunch. The extended breakfast on Thursday will allow for planning of the afternoon tour. Accompanying member registration \$150

Critical Dates Paper Submission:	Critical Dates Registration:
Abstract 21 July 2001	Early by 21 August 2001
Paper 21 September 2001	Standard by 21 October 2001

INFORMATION and REGISTRATION DETAILS

Acoustics and Vibration Unit, ADFA, CANBERRA 2600
Tel: 02 6268 8241 Mob: (0)2 240009 Fax: 02 6268 8276
m.burgess@adfa.edu.au www.acoustics.asn.au

ANNOUNCING!

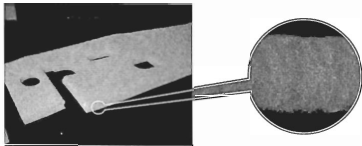
DECI-TEX 3D

inc engineered materials

What is it?

A revolutionary new acoustic blanket of fibres aligned in the vertical direction using the patented "Struto" process with a wide variety of surface membranes optimised for sound absorption.

- Up to Nine Times the Sound Absorption per unit weight
- A Breakthrough in Price / Performance ratios for Sound Absorption and Decoupling applications.
- Very high bond strength between substrate, facings and structure.
- Outstanding mechanical resilience.
- Able to be tuned for specific surface density, thickness, flow resistance and absorption spectra.



Deci-Tex 3D cut part, edge magnified to show structure

To find out more about Deci-Tex and typical applications, please email sales@inccorp.com.au or call us on (03) 9543 2800.

Achieve the ultimate with Brüel & Kjær service

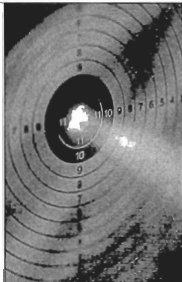
**Brüel & Kjær offers faster and better
service than any other lab in Australia
– at very competitive prices!**

For more information on how to freeze your expenses and save
a fortune on repairs and calibration costs...



Reg Lab No 1301

...call Brüel & Kjær's
Service Centre today on
(02) 9450 2066



SERVICE AND CALIBRATION

HEAD OFFICE, SERVICE AND CALIBRATION CENTRE:
24 Tepko Road • PO Box 177 • Terrey Hills • NSW 2084
Telephone (02) 9450 2066 • Facsimile (02) 9450 2379
e-mail: bk.service@spectrls.com.au • www.bk.com.au

Brüel & Kjær 

Diary...

2001

January 14-17, PATRAS

EURONOISE 2001

<http://euronoise2001.upatras.gr/> or LFME, Laboratory of Fluid Mechanics and Aerodynamics, University of Patras, P.O. Box 1406, 26500 Patras, Greece, fax: +30 61 996344 euronoise2001@upatras.gr

* February 7-9, MELBOURNE

7th Annual Conference

Assoc. of Occup Health & Safety Educators

Colin Findlay, OH&S Group, RMIT Applied Chemistry, PO Box 71, Bundoora Vic 3083

April 6-8, CAMBRIDGE

Noise Pollution and Health

www.ucl.ac.uk/noiseandhealth

June 4-8, CHICAGO

141st Meeting of Acoustical Society of America.

<http://asa.aip.org>, ASA, 500 Sunnyside Blvd, Woodbury, NY 11797-2999, USA, fax: +1 516 576 2377.

July 2-6, HONG KONG.

8th ICSV

<http://www.iav.org/>, mticiv8@ip.sjyu.edu.hk Dr K M Li, Dept Mechanical Engineering, Hong Kong Polytechnic University, Hung Hom S, Kowloon, Hong Kong, fax: +852 2365 4703

Aug 28 - 30, THE HAGUE

INTER-NOISE 2001

<http://www.internoise2001.tudelft.nl/> or Congress Secretariat, P.O. Box 1067, NL-2600 BB Delft, The Netherlands, fax: +31 15 263403, secretary@internoise2001.tudelft.nl

Aug 30 - Sept 1, RENNES

Social life and communication: an element of understanding in the evolution of language
Huguette Schaeck-Grillou, UMR 6552, Université de Rennes 1, Campus Beaulieu, 35042 Rennes cedex, France, Huguette.Schaeck@univ-rennes1.fr

September 2-7, ROME

17th ICA

<http://www.ica2001.it/> or A. Alippi, 17th ICA Secretariat 4, Dipartimento di Fisica, Università di Roma "La Sapienza", Via A. Scarpa 14, 00161 Roma, Italy, fax: +39 6 4424 0183,

September 10-14, PERUGIA

ISMA 2001

CIARM & Catgut Acoust Soc
<http://www.cim.vc.cnr.it/ISMA2001>, Musical Acoustics Laboratory, Fondazione Scuderie di San Giorgio - CNR, Isola di San Giorgio Maggiore, I-30124, Venezia, Italy, fax: +39 041 5208135, isma2001@cim.vc.cnr.it

* October 9, Sydney

Acoustics Workshop

National Noise Centre, University of Sydney, NSW 2006
tel Q 9351 5352, fax 02 9351 535,
wn@ccchs.usyd.edu.au

October 7 - 10, ATLANTA

2001 IEEE Int Ultrasonics Symp joint plus World Cong on Ultrasonics.

<http://www.itee-usfc.org/2001>,
fax: +1 217 244 0105

* November 21-23, CANBERRA

Acoustics 2001 AAS Annual Conference

www.acoustics.asn.au, Acoustics 2001, Aust Defence Force Academy, Canberra, ACT 2600, avanti@acfa.edu.au

03 - 07 December, FT. LAUDERDALE

142nd Meeting of the Acoustical Society of America.
<http://asa.aip.org>, ASA, 500 Sunnyside Blvd, Woodbury, NY 11797-2999, USA, fax: +1 516 576 2377

2002

19-21 August, MICHIGAN

Internoise 2002

<http://www.internoise2002.org/> or Congress Secretariat, Dept Mech Eng, Ohio State Univ, West IX B Ave Columbus OH 43210-1107 USA, peersen.1@osu.edu

19-23 August, MOSCOW

16th International Symposium on Nonlinear Acoustics (ISNA16).

O Rudenko, Physics Department, Moscow State University, 119899 Moscow, Russia,
isna@ica3000.phys.msu.su

16 - 21 September, SEVILLA

Forum Acustico 2002 (Joint EAA-SEA-ASJ Symposium) <http://www.cica.es/aliens/forum2002>, fax: +34 91 411 76 51

30 Nov-8 Dec, MEXICO

1st joint meeting of ASA, Iberian Fed. Acoustics, Mexican Int Acoust Soc
<http://www.igp.org/igp/can-con.html>

WWW Listing

The ICA meetings Calendar is available on <http://gold.asn.nrc.ca/ims/ica/calendar.html>

Australian Acoustical Society NEW INTERNET ADDRESS

The AAS has now dispensed with that cumbersome long www address which has served well for many years now.

David Watkins, the AAS Secretary, deserves many thanks for establishing, updating and maintaining the page - no small task.

This responsibility is now to be taken over by Terry McMinn in WA.

The new streamlined domain name is

www.acoustics.asn.au

The 'asn' stands for association and will become a common part of the domain name for organisations and societies.

This page provides many details on the activities of the Society including contact details for office bearers, membership applications, conference information, contents of the journal Acoustics Australia, useful links and lots more. Check it out and pass any suggestions onto Terry McMinn.



VIBRATION ISOLATION

Matrix Industries Pty. Ltd. patented wall ties provide structural support while reducing transmission of structure borne vibrations. Resilient floating systems are available for all masonry and cast in place walls and lightweight concrete floors.

Matrix Industries' wall ties are reducing noise in studios and theatres throughout Australia.

Enquiries are Sales:

MATRIX INDUSTRIES PTY LTD

144 OXLEY ISLAND ROAD, OXLEY ISLAND NSW 2430

PH: (02) 6553 2577

FAX: (02) 6553 2585

AUSTRALIAN ACOUSTICAL SOCIETY ENQUIRIES

NATIONAL MATTERS

- * Notification of change of address
- * Payment of annual subscription
- * Proceedings of annual conferences

General Secretary

AAS - Professional Centre of Australia
Private Bag 1, Darlinghurst 2010
Tel/Fax (03) 9587 9400
email: watkinsd@melbpc.org.au
www.acoustics.asn.au

SOCIETY SUBSCRIPTION RATES

For 2000/2001 Financial Year:

Fellow and Member	\$103.40
Associate and Subscriber	\$82.50
Retired	\$34.10
Student	\$23.10
Including GST	

DIVISIONAL MATTERS

Enquiries regarding membership and sustaining membership should be directed to the appropriate State Division Secretary

AAS - NSW Division

Professional Centre of Australia
Private Bag 1,
DARLINGHURST 2010
Sec: Mr D Eager
Tel (02) 9514 2687
Fax (02) 9514 2665
david.eager@uts.edu.au

AAS - Queensland Division

PO Box 760
Spring Hill Qld 4004
Sec: Rebecca Ireland
Tel: (07) 3367 3131
Fax: (07) 3217 0660
rireland@kamstsimpson.com.au

AAS - SA Division

C/-Department of Mech Eng
University of Adelaide
SOUTH AUSTRALIA 5005
Sec: Colin Kastell
tel: (08) 8303 3556
Fax: (08) 8303 4367
ckastell@mecheng.
adelaide.edu.au

AAS - Victoria Division

PO Box 417
Collins St. West
PO MELBOURNE 8007
Sec: Elizabeth Lindqvist
Tel (03) 9925 2144
Fax (03) 9925 5290
elind@rmit.edu.au

AAS-W A Division

PO Box 1090
WEST PERTH 6872
Sec: Mr J Macpherson
Tel (08) 9222 7119
Fax (08) 9222 7157
john_macpherson@eniron.gov.au

ACOUSTICS AUSTRALIA INFORMATION

GENERAL BUSINESS

Advertising Subscriptions

Mrs Leigh Wallbank
PO Box 579, CRONULLA 2230
Tel (02) 9528 4362
Fax (02) 9523 9637
wallbank@zipworld.com.au

ARTICLES & REPORTS NEWS, BOOK REVIEWS NEW PRODUCTS

The Editor
Acoustics Australia
Acoustics & Vibration Unit, ADFA
CAMPBERRA ACT 2600
Tel (02) 6268 8241
Fax (02) 6268 8276
email: acoust-aust@adfa.edu.au

PRINTING, ARTWORK

Scott Williams
16 Cronulla Plaza
CRONULLA 2230
Tel (02) 9523 5954 Fax (02) 9523 9637
email: print@cronullaprint.com.au

SUBSCRIPTION RATES

	Aust	Overseas
1 year	A\$ 57.20	A\$ 64
2 year	A\$ 96.80	A\$112
3 year	A\$137.20	A\$161

Australian rates include GST.
Overseas subscriptions go by airmail
Discounted for new subscriptions
20% Discount for extra copies
Agents rates are discounted.

ADVERTISING RATES

	B&W	Non-members	Sus Mem
1/1 Page		\$605.00	\$550.00
1/2 Page		396.00	357.50
1/3 Page		302.50	275.00
1/4 Page		258.50	242.00

Spot colour: \$110 per colour
Prepared insert: \$275 (additional postage may apply)
Column rate: \$19.80 per cm (1/3 p 5.5cm width)
All rates include GST

Discounted rates for 3 consecutive ads in advance

Special rates available for 4-colour printing

All enquiries to: Mrs Leigh Wallbank

Tel (02) 9528 4362 Fax (02) 9523 9637
wallbank@zipworld.com.au

ACOUSTICS AUSTRALIA ADVERTISER INDEX - VOL 29 No 1

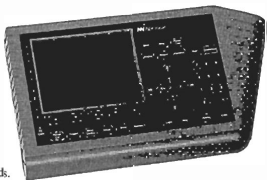
Acoustic Research Labs	35	Harcourt Australia	Insert	Nutek Australia	4
Airservices Australia	36	INC	50	Peace	29
Alliance Constructions	4	Infobyte	41	Rintoul	Insert
ASSTA	16	Kingdom	Inside front cover	RTA Technology	29
Australian Hearing	48	Lake Technology	36	Sound Control	Insert
Bruel & Kjaer	50, back cover	Matrix	51	Sound Barrier Systems	30
Davidson	30	Multi-Science UK	Insert	Wilkinson Murray	Insert
ETMC	Inside back cover	Noise Control Australia	48		

DATA ACQUISITION SYSTEMS

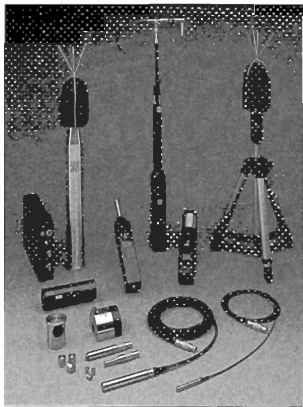
NORSONICS

Environmental Sound Analyser Nor-121

- SPL All these values measured in parallel with all time constants (F, S and I) and L_{\max} with A-, C- and Flat-weighting.
- L_{\min}
- L_{eq}
- L_E Except for L_{peak} these functions may also
- L_{peak} be measured in octave and 1/3 octave bands.



Measurement results and raw data may be stored on HD or standard PC card. The instrument may be operated in a stand-alone mode or be integrated into a comprehensive area noise monitoring system controlled over its computer interface.



G.R.A.S.

Sound and Vibration

- Condenser microphones
- Outdoor microphone systems
- Intensity probes and calibrators
- Probe and Array Microphones
- New series of Hydrophones and associated preamplifiers

ETMC Technologies

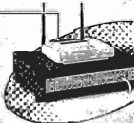
619 Darling Street ROZELLE NSW 2039

Tel: (02) 9555 1225

Fax: (02) 9810 4022

Web: www.ozpages.com/etmc

ONE SYSTEM - MANY SOLUTIONS - CHECK OUR PULSE..



With LAN wireless connection, measurement with PULSE is now even more flexible.

The PULSE™ platform is your advanced analyzer solution to sound and vibration measurement.

PULSE is modular. Its base software acts as a workspace to which any application from the ever-growing PULSE family line can be added - whatever your need. PULSE has the solution. Version 5.2 runs on Windows 2000® and has a new real-time, Order Tracking solution!

Solutions:

- General Noise and Vibration
- Product Noise Measurement
- Sound Quality
- Noise Source Identification
- Acoustic Material Testing
- Structural Dynamics
- Machine Analysis

PULSE is scalable. When you need more channels, just add them to your existing hardware, or scale up by adding a new front-end. The PULSE Analysis Engine gives you easily upgradeable, real-time, multi-analysis performance.

...And you can cut test time with PULSE's automatic transducer detection (TEDS) feature.

Want to know more?

Check out www.bksv.com.au, or contact your local sales representative.

OVER 55 YEARS OF EXPERIENCE

SYDNEY
Tel: (02) 9450 2066
Fax: (02) 9450 2379

MELBOURNE
Tel: (03) 9370 7666
Fax: (03) 9370 0332

PERTH
Tel: (08) 9381 2705
Fax: (08) 9381 3588

BRISBANE
Tel: (07) 3252 5700
Fax: (07) 3257 1370

Brüel & Kjær 

e-mail: bk@spectris.com.au