Australian Speech Science and Technology Association (Incorporated)
G.P.O. Box 143, Canberra

ASSTA Newsletter

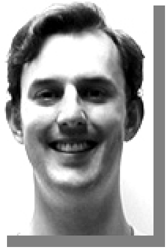
July 2002

Editorial

Welcome to the July 2002 issue of the ASSTA Newsletter.

In this issue you will find a thesis abstract from Jason Lukasiak, an announcement of the Christian Benoît award winner and a feature article from one of our members, Andrew Hunt, on standards for speech technology. I am happy to publish feature articles such as these, so please keep them rolling in! I have not received any 5th column submissions, but remember that the section is available for your views on ASSTA.

The ASSTA executive has suggested that it would be informative to include profiles of members' research groups in the ASSTA newsletter. We are keen to publish a short exposé on one such group in each newsletter. Would you like to volunteer for the next issue? Please



contact me if you are interested in taking part in this initiative.

Finally, the deadline for SST papers is drawing very close - see the final call for papers on page 12. I look forward to seeing you all there in December.

Michael Tyler

Contents

ASSTA Information

ASSTA Executive

President: Denis Burnham
d.burnham@uws.edu.au
Fax: (02) 9772 6736

Secretary: Steve Cassidy
steve.cassidy@mq.edu.au

Treasurer: Shunichi Ishihara
Shunichi.Ishihara@anu.edu.au

Member: David Grayden
D.Grayden@medoto.unimelb.edu.au
Fax: (03) 9283 7518

Member: Johanna Barry
J.Barry@medoto.unimelb.edu.au

ASSTA Newsletter Team

Editor: Michael Tyler
m.tyler@uws.edu.au
Fax: (02) 9772 6736

Sub-ed.(conferences): Marija Tabain
marija@maccs.mq.edu.au

Sub-ed.(reviews/abstracts): Johanna Barry
J.Barry@medoto.unimelb.edu.au

Contributions

Copy should be sent to the editor via Email as plain text or as an attachment in Rich Text Format (RTF).

Send notices or reports on conferences to Marija Tabain. Thesis abstracts and book reviews should be sent to Johanna Barry.

The address for hard copy is:

Michael Tyler
MARCS, University of Western Sydney
Bankstown Campus, Building 24
Locked Bag 1797, South Penrith D.C.
South Penrith, NSW, 1797.

Contributors to this issue: Denis Burnham, Andrew Hunt, Jason Lukasiak, and Marija Tabain. Thanks to Karen Mattock and Mel Gallagher for proofreading.

ASSTA recognises the assistance of the *MARCS Auditory Laboratories* and the *University of Western Sydney* in the production and distribution of this newsletter.

Please visit ASSTA on the Web:

http://www.assta.org

President's Report

Denis Burnham

Speech Scientists and Technologists of Australia, welcome again to the ASSTA Newsletter.

In this newsletter you will find an announcement about the second Prix Benoît, an award from the Christian Benoît Association, based at the Institut de la Communication Parlée in Grenoble, France, but with a scientific committee spanning continents, and areas of expertise in auditory-visual speech processing. There were 11 applications for the award this year, and the winner is both an Australian, and an ASSTA member, Ms Johanna Barry. Read more on Page 6 about the award and Johanna's project which will be supported by the award. Congratulations Johanna!!

You will also find information about the upcoming SST conference at the University of Melbourne from Tuesday December 3 to Thursday December 5, with a tutorial day on Monday December 2. The four invited speakers at the conference will each address one of the four themes of the conference: Associate Professor Kate Burridge (La Trobe University) will speak on how and why languages change; Professor Pim Levelt (Max Planck Institute for Psycholinguistics) will talk about human processing of language; Professor Max Coltheart (Macquarie University) will discuss professional applications for speech science, and; Dr Richard Cox (President of the IEEE Signal Processing Society) will address the theme of machine processing of spoken language. In addition, of course, there will be numerous papers from multifarious facets of speech including a good mix of technology and science, human and machine processing. Contributions to SST may be in the form of an abstract or a full 6-page paper, and the latter fulfils the DETYA requirements for funding under the Institutional Grants Scheme (E1: Conference Publications – full written paper – refereed proceedings). This is the premier speech science and technology conference in Australia, it attracts registrants from Australia and across the world, and it is the ideal setting for meeting fellow speechies and discovering who's doing what, where, and with

(Continued on page 5)

Ph.D. Thesis Abstract

Reviews compiled and edited by Johanna Barry

Ph.D. Thesis: *Techniques for low-rate scalable compression of speech signals*

Author: Jason Lukasiak, University of Wollongong, School of Electrical, Computer and Telecommunications Engineering

With the digitization of most communication channels and an ever-increasing demand for mobile communication services, the amount of traffic generated by coded speech signals continues to grow rapidly. To accommodate this increased traffic load in the finite bandwidth available for speech communication, it is necessary to develop speech compression algorithms that can dynamically scale to traffic and user demands. These scalable compression algorithms should be capable of dynamically altering the bit rate required for transmission, whilst smoothly and gradually varying the synthesized speech subjective quality with the changes in bit rate. To further increase the throughput of the communication channel, the scalable algorithm should operate in the lower range of bit rates currently used for speech compression (i.e. 2 to 8kbps).

This thesis proposes a number of scalable speech coding techniques that lead to the development of a single coding algorithm that is capable of scalable operation. Firstly, the characteristics of existing speech compression algorithms that limit scalable operation between bit rates of 2 and 8kbps are identified. The major limiting characteristics are identified as: 1) the existence of a distinct barrier at 4kbps below which parametric coders dominate and above which waveform coders dominate; and, 2) the large delay requirements for current low rate coding algorithms.

A method that exploits the simultaneous masking property of the human ear in a linear predictive filter is proposed. The proposed method modifies the linear predictive filter to remove more of the perceptually important information from the input signal than a standard linear predictive filter. This characteristic is shown to improve the subjective speech quality of low-rate linear prediction based speech coders.

To enable the pitch cycle redundancies of the speech signal to be exploited in the coding algorithm, without introducing excessive algorithmic delay, a novel low delay method for segmenting the speech into non-overlapped pitch length sub-frames is proposed. This method requires only a single frame of

speech and locates the pitch pulses by selecting the pulse locations in a closed loop function. The proposed segmentation is shown to produce a much more accurate pitch track in transient sections of the speech signal than the pitch track produced by traditional autocorrelation-based pitch detectors.

A number of low-delay decomposition techniques are proposed which decompose the speech into perceptually different components and allow scalable reconstruction of the speech signal. The preferred technique performs the decomposition in a closed loop function allowing quantization errors to be accounted for in the decomposition process.

The proposed scalable techniques are combined to produce a scalable algorithm that operates at a range of bit rates from 2 to 8kbps. The synthesized speech quality produced by the scalable algorithm varies smoothly as the operating bit rate is varied. A key feature of the proposed algorithm is the ability to migrate from a time-asynchronous parametric coder at low rates, to a time-synchronous waveform coder at higher bit rates. The coder also requires only a single frame of algorithmic delay (30ms) for operation.

Results indicate that the scalable coder produces subjective speech quality that is comparable with that achieved for fixed rate standardized coders at each of the tested bit rates.

(Pres report cont'd from page 3)
whom.

The new look ASSTA website (http://www.assta.org/) has been undergoing updates of late. We hope to have membership information back on-line soon, both for individuals and labs, which can be updated by members on-line. This will be a valuable resource - a good way of both finding out what others are doing, and letting others know what you are doing. In concert with this coming on-line, we will also have feature articles on particular labs around the countryside in subsequent issues of this newsletter.

Finally, since the last newsletter we have had applications for ASSTA research events, and for ASSTA PhD study awards. These are currently being evaluated, and you will know the results of these deliberations soon.

Christian Benoît Award

Christian Benoît was a CNRS researcher (Centre National de la Recherche Scientifique) based at ICP (Institut de la Communication Parlée) in Grenoble, France. He was recognised worldwide for his work in the area of speech synthesis and speech recognition and believed strongly that laboratory research should be used to benefit mankind, and most particularly the handicapped in society. After his untimely death in 1998, the Association Christian Benoît was set up in his memory, and one of the initiatives of the association was the "Prix Christian Benoît" which is offered biennially. The award is open to final year PhD students working in speech communication research worldwide, and a key aim of the award is to promote research in speech communication, by supporting the development of a multimedia project that will demonstrate the benefits of research in speech communication as well as promoting the career and research of the prizewinner. The inaugural prize was awarded in 2000 to Mr Tony Ezzat of the US. This year's winner, chosen from 11 entries from doctoral students from across the world, is Ms Johanna Barry, a student member of ASSTA.

Ms Johanna Barry is a PhD student from the University of Melbourne based in the Departments of Otolaryngology and Linguistics and Applied Linguistics. For her doctoral dissertation, Ms Barry is studying the development of spoken language in Cantonese-speaking children who have received a cochlear implant. Part of her research has involved investigating how well children are able to perceive and use information provided through the cochlear implant to produce Cantonese lexical tones. Ms Barry's research suggests that habilitation is an important factor in aiding children to develop tone in their spoken language.

"…This year's winner, chosen from 11 entries from doctoral students from across the world, is Ms Johanna Barry, a student member of ASSTA"

As part of her research, Johanna analysed and compared the tone productions of a large group of cochlear implant users with tone productions by normally-hearing adults and children. To enable her to do this in the time available, Dr Peter Blamey, also of the University of Melbourne, wrote a program called

(Continued on page 8)

Standards for Speech Technology

Author: Dr. Andrew Hunt

Director ASR Products, SpeechWorks International
andrew.hunt@speechworks.com

In recent years a range of efforts to standardize key speech technologies has emerged as core drivers for the speech industry. A broad spectrum of speech technology providers and consumers has participated in the development and promotion of these standards. There is also work on open source implementations of many of these standards, for example, at the CMU speech group (http://fife.speech.cs.cmu.edu/hephaestus.html). This article summarizes some of the key developments and provides links for additional information. While standards are far from being research activities familiar to ASSTA membership I hope this article gives an interesting angle on how speech research is transforming into a mainstream technology.

Core World Wide Web Consortium (W3C) Speech Standards: Speech recognition and speech synthesis are the core technologies that underlie most deployed speech applications. The W3C's Voice Browser Working Group (VBWG) (http://www.w3.org/voice/) recently published the Speech Recognition Grammar Specification (SRGS), that specifies standards for the representation of context-free grammars in both XML and Augmented BNF forms. The VBWG has also published a draft Semantic Interpretation specification for a tag attribution language that can be embedded in grammars to convert spoken input to a semantic form. The Speech Synthesis Markup Language (SSML) defines a standard set of XML elements to annotate text input to speech synthesis. The elements indicate document structure (paragraphs, sentences), provide controls of pronunciation, language and speaking voice, and provide prosodic controls (e.g., pitch, speaking rate, pausing, emphasis).

> *"…I hope this article gives an interesting angle on how speech research is transforming into a mainstream technology."*

Application Standards: Building upon the core speech specifications are two application-level languages, VoiceXML and SALT (Speech Application Language Tags), which add capabilities for implementing interactive spoken dialog. VoiceXML is the more established application standard, having been initially published 3 years ago, and now having widespread product support and ongoing development by the VBWG. It provides specialized dialog constructs for control-

(Continued on page 8)

(Benoît prize cont'd from page 6)
"Pitch". With the aid of this program, Ms Barry was able to identify some key parameters for describing tone, which could be presented both visually and numerically.

Johanna's proposal for the Prix Benoît, entitled "Tone visualisation for the habilitation for profoundly hearing impaired speakers of Cantonese", involves automating the "Pitch" analysis procedure, in conjunction with a pitch extraction routine. This will be used to develop a speech habilitation tool which can be used by clinicians to assist prelingually deafened children and post-lingually deafened adult speakers to visualise their tone productions. It is hoped that such a tool will help these populations of Cantonese speakers better understand the tone forms that they currently produce so that they can more effectively integrate speech input from the implant into their spoken language.

> *"[Johanna will] develop a speech habilitation tool which can be used by clinicians to assist prelingually deafened children and post-lingually deafened adults to visualise their tone productions"*

It is with great pride and exuberant gusto that we at ASSTA congratulate Johanna Barry on her award, and wish her every success in bringing this project to fruition.

(Feature article cont'd from page 7)
ling ASR and TTS that support telephony dialog interactions, and also basic telephony (e.g., transfer and hang-up) and platform controls. SALT (http://www.saltforum.org/) aims to meet the development needs for multimodal and mobile devices in addition to telephony applications. By defining a lightweight set of XML tags, SALT can be added to existing web languages, such as HTML, and take advantage of the execution model, tools and so on.

Other Standards: Distributed speech recognition, particularly important for mobile devices, is supported by the ETSI Aurora standard built upon the MFCC feature processing used in most speech recognizers. A new IETF activity to define protocols for control of ASR, TTS and Speaker Verification (CATS) promises to standardize the interface to network-based speech services. Additional standards are being considered for, among others, pronunciation lexicons, call control, and interoperation of voice browsers.

Through collaboration on standards, the industry aims to promote interoperability of products while maintaining robust competition amongst implementations of those standards. Standards organizations encourage an active public review of the specifications, so follow the links in this article if your research can inform these standardization efforts.

Conference Log

Conference log compiled by Marija Tabain

2002

August 24 – September 1

19th International Conference on Computational Linguistics-COLING2002

Location: Taipei, Taiwan.

Information: http://www.coling2002.sinica.edu.tw/

September 2 – 6

Linguistics and Phonetics 2002 (LP2002)

Location : Urayasu, Japan.

Information : midori34@meikai.ac.jp

September 9-12

Fifth International Conference on Text, Speech and Dialogue

Location: Brno, Czech Republic

Information: http://www.fi.muni.cz/tsd2002/

September 11 – 13

IEEE Workshop on Speech Synthesis

Location: Santa Monica, CA, USA

Information: http://www.research.att.com/conf/ttsworkshop/

September 14 – 16

3rd Biennial ICVPB: International Conference on Voice Physiology and Biomechanics

Location: Denver, USA.

Information: http://www.nwu.edu/csd/ICVPB/

September 17 - 22

7th International Conference on Spoken Language Processing: ICSLP 2002

Location: Denver, Colorado, USA

Information: http://www.icslp2002.org/

October 14 - 16

IEEE 4th International Conference on

Multimodal Interfaces

Location: Pittsburgh, USA.

Information: http://www.is.cs.cmu.edu/icmi/

November 1 – 3

9th International Phonology Meeting: Structure and Melody

Location: Vienna, Austria

Information: http://www.univie.ac.at/linguistics/conferences/phon02/

November 30 - December 6

144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics, 9th Mexican Congress on Acoustics

Location: Cancun, Mexico

Information: http://asa.aip.org/cancun.html

December 1 – 5

9th Australian International Conference on Speech Science and Technology – SST 2002

Location : Melbourne, Australia

Information : http://www.conferences.unimelb.edu.au/SST/

2003

January 9-11

Old World Conference in Phonology I (Segmental Phonology)

Location: Leiden, Netherlands.

Information: http://www.let.leidenuniv.nl/ulcl/events/ocp1/

January 21 - 24

8th International Symposium on Social Communication.

Location: Santiago de Cuba, Cuba.

Information: leonel@lingapli.ciges.inf.cu

March 27-29

International Colloquium on Prosodic Interfaces

Location: Nantes, France.

Information: ip2003@humana.univ-nantes.fr

April 7-9

Eighth Western Pacific Acoustics Conference

Location: Melbourne, Australia.

Information: http://www.wespac8.com/

6TH INTERNATIONAL SEMINAR ON SPEECH PRODUCTION

Macquarie Centre for Cognitive Science (MACCS) and the Speech Hearing and Language Research Centre (SHLRC) at Macquarie University are pleased to host the sixth in the series of Speech Production Seminars, which will be held in Sydney, Australia from December 8th to 10th, 2003. We look forward to continuing the tradition established at the previous meetings in Kloster Seeon, Grenoble, Leeds, Old Saybrook and Autrans of bringing together leading experts for the presentation and discussion of current research into all aspects of speech production.

Topics

The meeting will maintain the customary focus of the Speech Production Seminars on the analysis and modelling of all aspects of articulatory processes. The following list of themes is not meant to be exhaustive: Articulatory-acoustic relations, Models of motor control, Articulatory synthesis, Audiovisual synthesis, Acoustic to articulatory inversion, Aerodynamic models, Connected speech processes, Cerebral organization of speech, Coarticulation, Disorders of speech motor control, Biomechanical modelling, Instrumental techniques.

Registration and Deadlines

To be placed on the mailing list for further information please visit http://www.maccs.mq.edu.au/mailman/listinfo/speechprodconf. Call for papers will be announced in December 2002, and the deadline for submission of abstracts will be March 31, 2003. The conference website can be viewed at http://www.maccs.mq.edu.au/events/2003/issp2003.

Keynote Speakers Include:

Prof Andy Butcher, School of Medicine, Flinders Uni
Prof Frank Guenther, Department of Cognitive and Neural Systems, Boston Uni
Prof Pat Keating, Phonetics Lab, UCLA

Venue

The conference will be held at The International College of Tourism and Hotel Management, which is housed within St Patrick's College, one of the most spectacular buildings in Sydney. Located at Manly, this former monastery offers magnificent views of the Pacific Ocean and Manly beach and is only a short walk into the centre of Manly. From here Sydney city can easily be reached by ferry.

An ARC Funded Special Research Centre

Macquarie University • Sydney • Australia

SST-2002 Melbourne

9TH AUSTRALIAN SPEECH SCIENCE AND TECHNOLOGY CONFERENCE
TUESDAY 3 DECEMBER TO THURSDAY 5 DECEMBER 2002
TUTORIAL DAY - MONDAY 2 DECEMBER 2002
MELBOURNE AUSTRALIA

FINAL CALL FOR PAPERS

Submissions will not be accepted for review later than 16th August 2002.

For more information on the 'Call for Papers' please visit the conference website Presenters page:

http://www.conferences.unimelb.edu.au/SST/Presenters.htm

For updates on the conference please keep watching the conference website:

http://www.conferences.unimelb.edu.au/SST/

This is a conference too good to miss out on, so please pass the word on to your friends and work colleagues!